The High-Redshift Quasar Luminosity Function from Multi-Epoch Imaging Surveys

Yusra AlSayyad

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

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Andrew J. Connolly, Chair

Ian D. McGreer

Željko Ivezić

Program Authorized to Offer Degree:
Astronomy

University of Washington

Abstract

The High-Redshift Quasar Luminosity Function from Multi-Epoch Imaging Surveys

Yusra AlSayyad

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

Upcoming time-domain imaging surveys such as the LSST will detect over a million high-redshift $z > 4$ quasars, making complete spectroscopic followup unfeasible. Statistical estimates such as luminosity functions and clustering measurements will require purely photometric methods for classifying objects, estimating redshifts and estimating selection functions. We validate these methods and constrain the optical, type I quasar luminosity function (QLF) at $3.75 < z < 4.5$ for $-27.5 < M_{1450} < -23.5$. This thesis presents the first variability-selected QLF measurement at high redshift ($z > 3.75$) and constraint on the characteristic luminosity ($M^*_{1450} = -26.7$) from a single, uniformly-selected survey at $z \sim 4$.

We used the Sloan Digital Sky Survey (SDSS) repeated imaging of the $275\,\mathrm{deg}^2$ equatorial region of the sky ($-50 <$ R.A. $< +60; -1.26 <$ Dec. $< +1.26$) known as Stripe 82 to select a statistical sample of $z \sim 4$ quasars. We extracted 40 million lightcurves from the imaging using forced photometry on all u, g, r, i, z epochs at the positions of sources detected on a deep i-band co-add. We developed a classification method based on photometric information alone (colors and variability metrics derived from these new multi-band lightcurves), which we validated with a spectroscopically complete $55\,\mathrm{deg}^2$ sub-region augmented with 102 new spectroscopic observations of quasars at $z > 3.4$ with $i < 22.5$. We demonstrate that selection functions for ensemble classifiers can be estimated by building generative models of empirical distributions of quasars previously selected with a diverse set of selection criteria.

The $z \sim 4$ QLF contributes to our understanding of supermassive black hole growth and cosmic reionization of both H I and He II which likely began at $z \sim 4$ as a result of hard UV emissivity from quasars. The resulting QLF measurement is consistent with the previous lower number densities reported from deep, narrow-field surveys (COSMOS); it is not consistent with higher number densities reported from the NDWFS-DLS and CANDELS GOODS-S fields. In the context of recent $2.2 < z < 5$ estimates from SDSS-III BOSS and Stripe 82, it indicates that the QLF evolves in both luminosity and density: the characteristic luminosity increases and density decreases log-linearly from $z = 2.2 \rightarrow 5$. It predicts that quasars contribute 100% of the ionizing flux needed to balance hydrogen recombinations at $z \sim 4$. Given that at $z \sim 6$ hydrogen recombines faster and quasars are less abundant, this result suggests that quasars contributed a negligible fraction to H I reionization. Finally, we find that the UV ionizing flux from quasars at $z \sim 4$ is sufficient to ionize He II by $z \sim 3$, consistent with independent observations of the He II Lyα forest and temperature of the intergalactic medium.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Quasars are thought to be powered by gas accretion onto supermassive black holes (SMBHs) at the centers of galaxies. Today, we see that SMBHs are ubiquitous. They inhabit all nearby galaxies (Kormendy & Richstone 1995), and their masses correlate with properties of their host galaxies (e.g., Magorrian et al. 1998; Ferrarese & Merritt 2000). As more evidence is gathered, it appears that SMBHs coevolve with their host galaxies and that quasars represent a short-lived stage in galaxy evolution that occurs only when gas is available near the black hole. It is during the quasar phase that SMBHs grow most rapidly and quasar feedback processes impact star formation in the host galaxies. Despite extensive theoretical research, a full theory of the coevolution of galaxies and quasars has yet to emerge (see the reviews by Alexander & Hickox 2011; Kormendy & Ho 2013), and key theoretical questions remain: How do SMBHs form? How and when do they grow?

To address these questions, phenomenological models have been developed (e.g., Conroy & White 2013; Hickox et al. 2014; Veale et al. 2014; Caplar et al. 2015) that predict quasar populations given realistic relations: the distribution of SMBHs that power quasars, galaxies that host the SMBHs, dark matter halos that host the galaxies, accretion rates and quasar lifetimes. These models are constrained by a key observable: the number density of quasars per unit luminosity per comoving unit volume, known as the quasar luminosity function (QLF). The QLF's shape is well represented by a double power law (Boyle et al. 1988), where the luminosity at which the power-law slope changes, called the break or the "characteristic luminosity," separates the bright end of the QLF from the faint end. The optical QLF has been measured for bright quasars from $z \lesssim 6$ in wide-area surveys, such as the Sloan Digital Sky Survey (SDSS) (Fan et al. 2001; Richards et al. 2006, 2008; Shen & Kelly 2012; Ross

et al. 2013; Jiang et al. 2009; Willott et al. 2010; Palanque-Delabrouille et al. 2012), and for faint quasars at $z \lesssim 5$ (McGreer et al. 2013; Wittman et al. 2002; Jiang et al. 2006). However, at redshift 4, the faint-end luminosity function measurements (Ikeda et al. 2011; Masters et al. 2012; Glikman et al. 2011; Giallongo et al. 2015) are discrepant (by a factor of ≈ 4). These measurements are poorly constrained because they are limited to small areas (1–3 $\deg^2$) and thus are subject to low number counts and cosmic variance.

The previous, discrepant measurements of the QLF at $z \sim 4$ imply very different cosmic histories, because the evolution of the shape of the QLF impacts our understanding of how SMBHs grow (§1.1), and the faint-end slopes affect our understanding of cosmic reionization (§1.2). The steeper faint-end slopes measured by Glikman et al. (2011) and Giallongo et al. (2015) imply that quasars alone could have reionized the universe by $z \sim 6$ (Madau & Haardt 2015); the flatter slopes measured by Ikeda et al. (2011); Masters et al. (2012) would imply that quasars contributed a negligible amount of UV photons to reionization.

Resolving these divergent measurements, and their implications for cosmic history, calls for more data. Specifically, it requires a survey deep enough to reach beyond the break in the QLF, but at the same time wide enough to find a statistically significant sample of these relatively rare objects. The resolution of this discrepancy can be achieved by using existing multi-band imaging such as the SDSS data on the repeatedly scanned, 275 $\deg^2$ equatorial region of the sky ($-50 <$ R.A. $< +60; -1.26 <$ Dec. $< +1.26$) known as Stripe 82. At any given location on the stripe, there are ≈ 80 observations taken over ten years, each with a depth of $r \sim 22.5$. By combining information from the multiple epochs, SDSS data on Stripe 82 reaches two magnitudes deeper than the main survey. These Stripe 82 observations combine medium depth and medium sky area (Figure 1.1)(relative to the breadth of the SDSS survey and the small-area deep fields surveyed by COSMOS and the NOAO Deep Wide-Field Survey and the Deep Lens Survey (NDWFS+DLS; Wittman et al. (2002)). Stripe 82 thus offers an ideal combination for searching for rare faint high-redshift quasars.

Stripe 82 also provides 10-year variability information, which allows high-z quasars to be accurately selected from the dense background of galaxies and stars. Quasars exhibit unique

variability signatures; they vary stochastically and on longer timescales ($\sim 10^{2-3}$ days) than do variable stars. The best variability statistics thus utilize timescale information to discriminate quasars from variable stars. For example, a simple χ^2 on a time-binned lightcurve would test the hypothesis that the source was variable on the timescale set by the bin cadence. To probe the timescale more directly, Kelly et al. (2009) showed that quasar lightcurves are well described by a continuous-time autoregressive process of order 1 (CAR(1)), also referred to as a damped random walk (DRW) or an Ornstein-Uhlenbeck process. Modeling quasar lightcurves as CAR(1) processes provides estimates on the timescale of variability, and previous studies with Stripe 82 data have demonstrated the power of CAR(1) parameters to select quasars (MacLeod et al. 2011; Butler & Bloom 2011; Choi et al. 2014).

Independently, the added depth and variability information available from Stripe 82 have already proven fruitful in measuring the QLF at lower redshift $z < 4$ (e.g., Ross et al. 2013; Palanque-Delabrouille et al. 2012) and at $z \sim 5$ (McGreer et al. 2013). The discriminating power of this added depth and variability information is amplified when used in combination for selection.

We present a measurement of the QLF at $z \sim 4$ selected using color and variability information derived from the optical (probing the rest-frame ultraviolet) Stripe 82 imaging. By selecting quasars on Stripe 82, we are able to reach sufficient depth ($i < 22.5$) to probe the faint end of the luminosity function over a wide enough range ($-27.5 < M_{1450} < -23.5$) to cover the break luminosity, constraining its overall shape in one uniformly selected, homogenous sample. By basing selection simultaneously on color and variability, using powerful variability metrics such as CAR(1) parameters χ^2 of annually averaged lightcurves, we achieve selection accuracy sufficient to circumvent the need for full spectroscopic coverage. This uniform sample of $z \sim 4$ quasars represents the first variability-based QLF at high redshift and provides a key link in our understanding of reionization and the growth of SMBHs at high redshift.

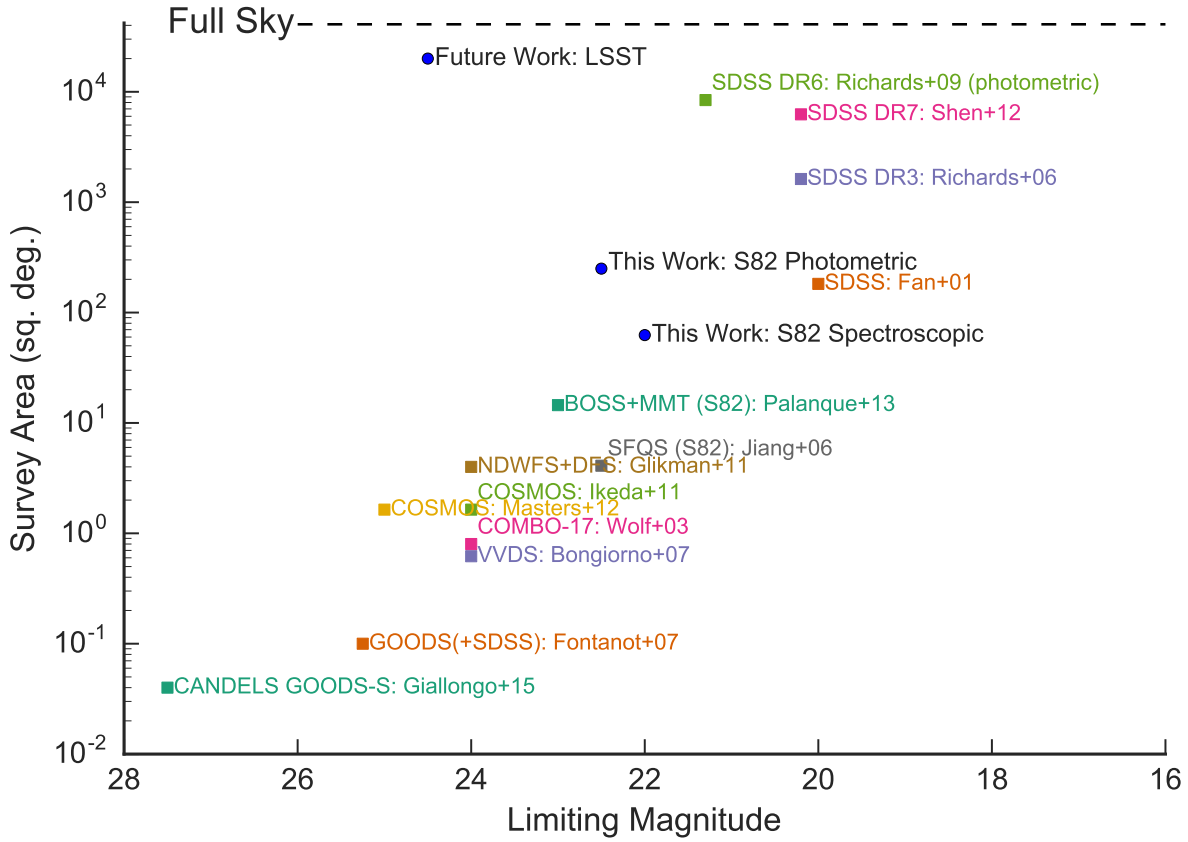

Figure 1.1 This $z \sim 4$ QLF study was done in the context of other studies that covered $z = 4$ or $z = 3.75$. This work consists of two $z \sim 4$ QLF measurements on Stripe 82. The spectroscopic measurement covers an effective survey area of $55\,\mathrm{deg}^2$ to an effective depth of $i \sim 22$, with 150 spectroscopically confirmed quasars. The photometrically selected sample covers an effective survey area of $216\,\mathrm{deg}^2$ with 593 quasars (spectroscopic redshifts were used for 346 objects).

1.1 Implications of QLF Evolution for Growth of SMBHs

Evidence is growing that SMBHs co-evolve with the galaxies that host them. First, the masses of SMBHs correlate with the characteristics of their host galaxies (Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Gültekin et al. 2009). Second, the cosmic star formation rate in galaxies appears to evolve in lockstep with the luminous quasar population, which is a manifestation of SMBH accretion. Like galaxies exhibiting active star formation, quasars become more abundant with cosmic time, peak in number density at $z \sim 2.5$, and then become less abundant to present day. The evolutionary trends on either side of $z \sim 2.5$ are attributed to different reasons. Before $z \sim 2.5$, black holes have not had enough time to build up; after $z \sim 2.5$, they do not have enough gas to power accretion.

During any stage when SMBHs are accreting gas, they are described as "active" galactic nuclei (AGNs). The majority of studies on the joint cosmological growth of galaxies and their SMBHs focus on typical AGNs with luminosities below the break in the LF. Quasars are defined as bright AGNs (historically: $M_B < -23$, and more recently: $L_{bol} > 10^{45}\mathrm{erg\ s^{-1}}$), but the distinction is only of historical importance, as no discontinuity of properties is observed across luminosities. While low-luminosity AGNs—the majority of AGNs in the universe—play a role in galaxy evolution, they are unlikely to be responsible for the bulk of the accretion onto, and thus the growth of, SMBHs. Synthesis models of the SMBH population (Merloni & Heinz 2008; Shankar et al. 2009) show that the bulk of the growth of black holes $M_{\mathrm{BH}} > 10^8 M_\odot$ occurs when AGNs are powerful quasars at high redshift, accreting at a high Eddington ratio: the ratio of the actual accretion rate to the maximum theoretical rate at which a luminous body can spherically accrete gas.[1] Additionally, quasars are sufficiently bright to be observable across a wide range of redshifts; the earliest quasar observed is at $z = 7.1$ (Mortlock et al. 2011).

Just as the overall number density of quasars evolves across time, so does the *shape* of

[1]Quasars do not necessarily accrete gas spherical symmetrically, but the Eddington limit and Eddington ratio are physically meaningful and scale-free quantities. Quasars can have an Eddington ratio of greater than one, and these are described as "super-Eddington."

the QLF. Over a wide redshift range, the QLF shape is well fit by a double power law,

$$\Phi(L) = \frac{\phi^*}{(L/L^*)^\alpha + (L/L^*)^\beta}, \tag{1.1}$$

parameterized by the faint-end slope α, the bright-end slope β, the "break," or characteristic luminosity, L^* that separates the bright and faint ends, and the normalization, defined as the density at the characteristic luminosity, $\phi^* = \Phi(L^*)$. The history of SMBH growth can be constrained by the observed evolution of the shape of the QLF in terms of its two slopes, break, and normalization.

Evolution of the QLF is usually modeled in terms of *luminosity* evolution, *density* evolution, or some combination thereof. The QLF is a function of both luminosity and redshift, and is often plotted as one-dimensional slices through either luminosity or redshift. In a luminosity function plot with L on the x-axis and $\Phi(L)$ on the y-axis, pure luminosity evolution (PLE) would manifest as a horizontal shift in the LF plot with redshift. In PLE, some fraction of the high-redshift galaxy population are active as quasars, and remain so while they dim and brighten; the rest of the galaxy population never becomes active. In this model, the apparent lack of local luminous quasars is caused by the dimming of each quasar over time due to decreasing accretion rates. Because SMBHs are ubiquitous in the local universe, this seems unrealistic. Observations of the QLF at low redshift can be described by PLE, but at higher redshift the PLE model fails to fit the observations Croom et al. (2009).

At the other extreme is pure density evolution (PDE), which would manifest as a vertical shift in the LF as a function of redshift. In PDE, quasars do not have long lifetimes; they have short lifetimes, referred to as their "duty cycles." At high redshift, PDE predicts substantially fewer faint high-redshift AGNs than PLE (Wolf et al. 2003). At low redshift, the local absence of luminous quasars is because the fraction of SMBHs that are active as quasars drops as $z \to 0$. PDE does not fit the observed QLF at any redshift. It over-predicts the number of luminous quasars in the local universe, as the QLF shows a more dramatic decrease of the brightest quasars than of the fainter quasars from $z \sim 2.5 \to 0$. In other words, the evolution of the QLF appears to be a combination of luminosity and density

evolution (LEDE) or luminosity-dependent density evolution (LDDE). Indeed, before the quasar epoch, $z \sim 2.5 \rightarrow 4$, LEDE has been shown to better describe the optical QLF (Ross et al. 2013).

These evolutionary models may be tested by measuring how the luminosity and normalization of the QLF break evolved. It is difficult to model the evolution of the luminosity function if only the bright or faint ends are observed, because each end is a single power law. PLE and PDE models are completely degenerate for a single power law. Only by observing the break in the QLF can we constrain models for the evolution of the quasar population at high redshift.

1.2 QLF Implications for Cosmic Reionization

The Epoch of Reionization is the period in the history of the universe when the predominantly neutral intergalactic medium was ionized by the first luminous sources. These sources may have been low-luminosity galaxies, population III (POPIII) stars, quasars, or some combination thereof. Observations show that reionization was complete by $z \sim 6$–10, but we still do not know definitively which sources initiated the reionization of intergalactic hydrogen.

One path forward involves estimating the number of UV-ionizing photons from each of these sources by integrating the sources' observed luminosity functions and comparing this estimate to the number of photons required to ionize the intergalactic medium. However, low-luminosity galaxies at $z \gtrsim 6$ are difficult to observe because they are so faint, and POPIII stars have not yet been observed. Quasars, on the other hand, are bright enough to be observed at high redshift. Indeed, the high-redshift QLF provides a key constraint on the fraction of the ionizing background that came from quasars.

While evidence has increasingly shown that low-luminosity galaxies were the major contributors to hydrogen reionization, and that quasars contributed a negligible fraction, each piece of observational evidence thus far has had a caveat or counter-observation, leaving open the possibility that quasars did play an important part in cosmic reionization. Better constraining the QLF may help rule out these caveats.

For example, one argument that galaxies were the major contributors to hydrogen reionization is based on evidence that helium was reionized later than hydrogen, at $z \sim 3$ (see recent review from) McQuinn 2015). Because quasars have a harder spectrum than massive stars, quasars would have reionized both helium and hydrogen simultaneously.[2] Since the energy required to ionize hydrogen (13.6 eV) is much lower than that required to remove one helium electron (24.6 eV) or two helium electrons (54.4 eV), reionizing hydrogen at $z \sim 6$ without fully ionizing He II would have required a soft UV spectrum. However, the possibility remains that after reionization, the UV background spectrum became softer, keeping hydrogen ionized while allowing He II to recombine. This scenario is possible only if the ionizing output of stars increased fast enough to become the dominant contribution to the UV background $z < 6$. While it is possible, this scenario remains unlikely, because the luminosity function of quasars increases at least as fast as those of galaxies $z \sim 5 \rightarrow 2.5$. It is unclear how quasars could have contributed to reionization but not to the UV background after reionization.

In another argument for low-luminosity galaxies as the primary reionization sources, the bright quasar population declines at higher redshifts $3 < z < 6$, but the star formation rate evolves less rapidly. This implies that galaxies become increasingly important at high redshifts, and it is thus unlikely that there were sufficient numbers of bright quasars to start reionization. The caveat to this argument is that, like galaxies, quasars have a power law faint-end luminosity function; faint quasars greatly outnumber bright quasars. The faint end of the QLF has been poorly constrained at high redshift, which leaves open the possibility that there may have been enough *low*-luminosity quasars to begin the ionization process. Indeed, Glikman et al. (2011) and Giallongo et al. (2015) have reported sufficient number densities of faint quasars at $z \gtrsim 4$ to reionize the universe by $z \sim 6$ (Madau & Haardt 2015), without the need for contributions for other sources.

By observing the break in the QLF at $z \sim 4$, we improve constraints on the evolution of

[2]POPIII stars also are theorized to have hard spectra, and this same line of reasoning is used to argue that POPIII stars were unlikely contributors to reionization.

the faint-end slope. This provides an improved constraint on the ionizing UV background, specifically the contribution of quasars to the reionization of H I and He II. While the QLF at $z \sim 4$ provides indirect constraints on the sources of H I reionization at $z \sim 6 - -8$, it provides very direct constraints on He II reionization. Measurements from the He II Lyα forest (McQuinn 2009; Shull et al. 2010; Syphers et al. 2011; Worseck et al. 2011) and the QLF show that helium reionization began at redshifts $z \gtrsim 3.5$ or $\gtrsim 4$. Ionization of He II measurably heats the IGM over $3 < z < 4$ (Becker et al. 2011), which can be modeled as a result of He II ionizing flux (Sanderbeck et al. 2015; D'Aloisio et al. 2016) from quasars. We can tie the observed causes of He II reionization (quasars) with the observed effects (e.g. IGM heating, Lyα forest) to form a complete picture of this important phase transition in the history of the IGM. By observing the QLF at $z \sim 4$, we can directly measure the buildup of the far-UV background from quasars that likely drove He II reionization at precisely the epoch at which it happened.

Why $z \sim 4$ quasars? While the primary motivation for this study of $z \sim 4$ was the direct impact of quasars on He II reionization, which likely occurred at this epoch, and lack of consensus from previous measurements, there are several observational motivations for studying this redshift. First, color selection of quasars at $z \sim 4$ is relatively efficient, which allows us to use color selection (an established method of quasar selection) to complement the novel variability-selection methods we propose here. The fact that color selection is efficient at this redshift is important, as it means that we expect sufficiently pure and complete variability- and color-selected samples. Second, spectroscopic catalogs of brighter ($r < 21.5$) quasars from the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013) provide ~ 200 spectroscopic redshifts (which we augment with 102 long-slit spectra taken at the MMT, Gemini North, and Apache Point (APO) observatories) for calculation of the QLF. These spectra provide a training set that can be used to develop and test the variability-sselection methods that will be key for building statistical samples of quasars in future time-domain surveys, which will be too large for complete spectroscopic follow-up. Finally, $z \sim 4$ lies at the boundary between high- and low-redshift studies. By

virtue of being lower redshift, $z \sim 4$ has the highest number counts in high-redshift epochs, thus providing a large enough training sample for the development of purely photometric methods of measuring the QLF with LSST.

In this work, we refer to $z > 3.5$ as high redshift. In terms of cosmic history, quasars' number density rises until a peak at $z \sim 2.5$ and then declines until today. At $z \sim 4$, which corresponds to the epoch when the universe was 1.5 Gyr old, the number density of quasars was still growing. We expect different mechanisms to control the density and luminosities of quasars before and after the peak quasar epoch. $z \sim 4$ is closer to in time and shares more with the high-redshift epoch of reionization than with the epoch of peak star-formation and quasar activity.

In the observational sense, techniques for observing $z > 3.5$ quasars are the same as those for observing higher-redshift quasars. At $z > 3.5$, quasar colors become well separated from the colors of stars and galaxies, and the Lyman break first moves into optical bands, producing "dropouts." Whereas selection at high-redshift is more efficient, the number densities are too low to make use of multi-object spectroscopic surveys.[3] In terms of quasar surveys, the broad "all-redshift" surveys use multi-object fiber spectroscopy (e.g., Ross et al. 2013; Shen & Kelly 2012) go only to $z < 4$, with the center of the highest bin being $z = 3.75$. At $z \gtrsim 4$, targeted faint-end studies have pushed to higher redshifts—McGreer et al. (2013) at $z \sim 5$ and Jiang et al. (2009) and Willott et al. (2010) at $z \sim 6$—because high redshift studies offer better constraints on current topics such as hydrogen reionization. But because $z \sim 4$ is at the boundary of high and low redshift, this epoch has been under-studied.

1.3 Approach

To better constrain the evolution of quasar activity and the contribution of quasars to H I and He II reionization, we present a measurement of the $3.75 < z < 4.5$ optical, type I QLF down to $i = 22.5$ with 275 deg^2 of the repeated imaging of the SDSS Stripe 82. This dataset

[3] An SDSS plate would only cover tens of candidates with its thousand fibers.

covers an area that is two orders of magnitude greater than those offered by previous surveys in this redshift range and depth. It provides the dynamic range in luminosity to resolve the "break" in the QLF with one uniformly selected sample, providing better constraints on the bright-end slope, the faint-end slope, and the characteristic luminosity of quasars at $z \sim 4$.

The Stripe 82 dataset provides lightcurves with 50–100 epochs and a 10-year baseline: sufficient information to identify quasars by their variability characteristics. At $z \sim 4$, quasar optical colors are already distinct from those of stars and galaxies; we demonstrate that adding variability criteria means that purely photometric selection of quasars from large datasets using classifiers is now viable, and is sufficiently dependable to circumvent the need for spectroscopic confirmation on all candidates. Equipped with new forced-PSF flux lightcurves for all 3σ detections on an i-band co-add, we select candidates based on both colors and variability metrics. We used colors and lightcurves derived from forced photometry, a method we introduce in Chapter 2. These derived colors and lightcurves are less biased near the detection limits and allow for accurate selection of quasars down to the single-epoch 5σ detection threshold. From these lightcurves, we derived a variety of variability metrics, including annual averaged χ^2 parameters and CAR(1) parameters.

We augment a training set with 102 new spectroscopically confirmed redshifts, yielding a $55\,\mathrm{deg}^2$ subregion that is spectroscopically complete to $i \sim 22$. Because we use new methods to compute a pure photometric, variability-based luminosity function, we use this spectroscopic sample to validate the purely photometric selection methods used on the full processed and unmasked $216\,\mathrm{deg}^2$ of Stripe 82. The geometry of the spectroscopic survey area in relation to the full photometric survey area is shown in figure 1.2. We classify objects using a Random Forest ensemble classifier (Chapter 4) and quantify the selection function based on this classifier.

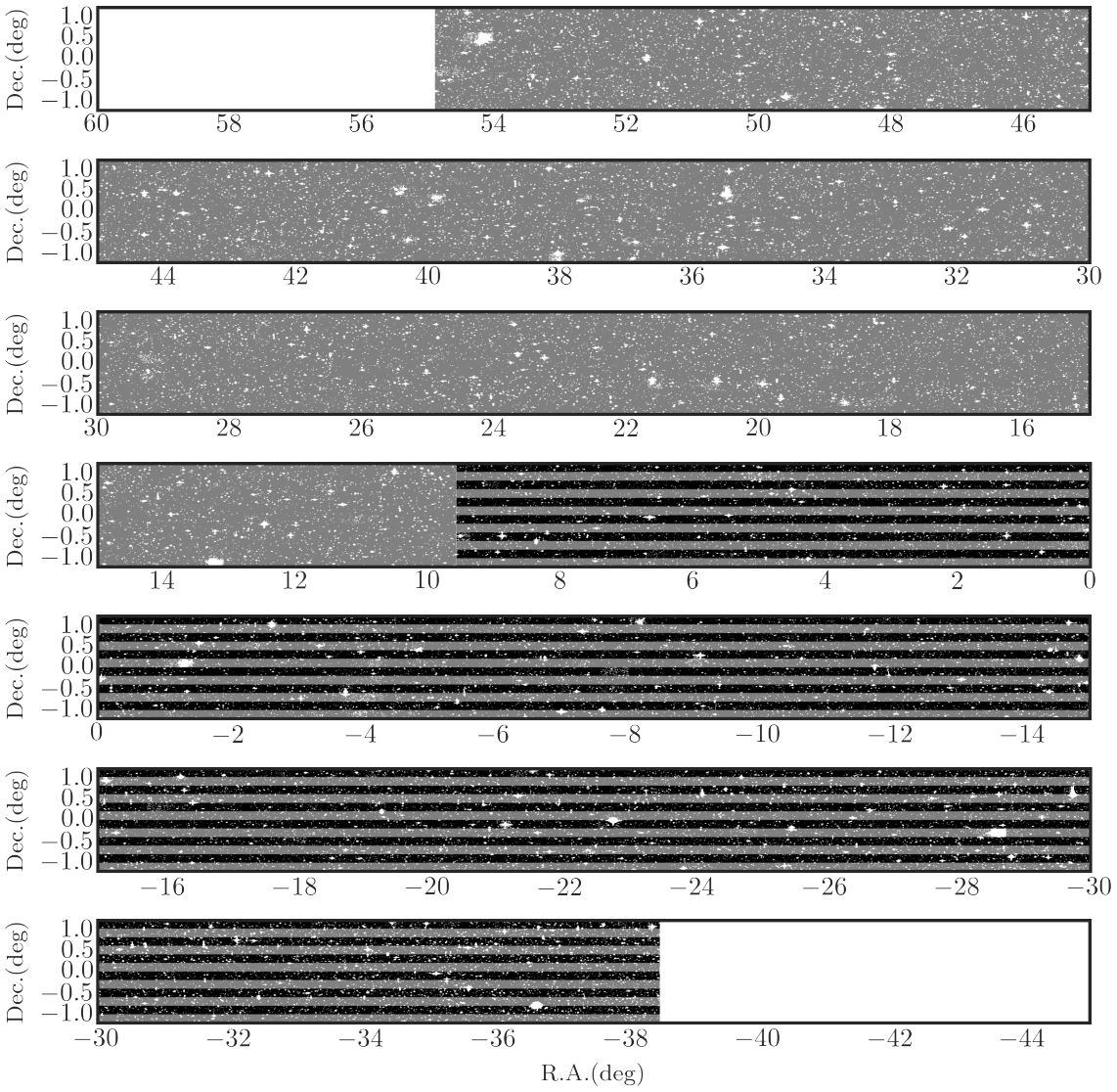

Figure 1.2 Survey area and geometry. Photometric Sample survey area (gray and black) and Spectroscopic Sample survey area (black only). Masked regions inaccessible to the survey are shown in white.

1.4 Outline

Using lightcurves extracted from the Stripe 82 imaging, we classify $3.5 < z < 4.5$ quasars trained on a sample of previously known and newly labeled lightcurves. We use this sample to measure a spectroscopic luminosity function at $z \sim 4$, which validates the luminosity function computed with only photometric information: colors, morphologies and variability metrics. In Chapter 2 we describe the process of extracting new lightcurves and colors from the Stripe 82 imaging. In Chapter 3 we present 102 spectra obtained specifically for this study. We describe the compilation of a spectroscopically confirmed training set from previously identified and newly identified quasars in Chapter 4, along with the classification methods used and the variability characteristics of $z \sim 4$ quasars. In Chapter 5 we present the spectroscopic QLF and the purely photometric QLF, along with a discussion of implications for reionization and the evolution of supermassive black holes. In Chapter 6 we close with a summary of this work and a discussion of future directions.

Chapter 2

FORCED PHOTOMETRY LIGHTCURVE CATALOG ON STRIPE 82

2.1 Introduction

During the fall months for nearly a decade, the Sloan Digital Sky Survey (SDSS) repeatedly scanned a 235 $\deg^2$ equatorial region of the sky ($-50 < R.A. < +59$; $-1.25 < Dec. < +1.25$) known as Stripe 82. At any given position on the stripe, there are between 60 and 100 observations, each with a depth of $r \sim 22.5$. The unique drift-scan camera (Gunn et al. 1998) images the sky in each of 5 bands—u, g, r, i, and z—nearly simultaneously. The survey ran between 1998 and 2007. Because of its 10-year baseline, Stripe 82 is an excellent dataset for studying the variability of quasars; quasars are known to vary on long timescales.

During the first five years, the cadence was only one to five observations per year. However, starting in 2005 with the SDSS-II supernova project (Frieman et al. 2008), observations were repeated more frequently, up to ~ 20 times per year, without waiting for photometric conditions. The dataset thus includes data from these later years that was collected on nights with poor conditions such as clouds and poor seeing.

Because it includes variable conditions, Stripe 82 is also an excellent dataset with which to test the capabilities of the LSST Software Stack. These observing conditions—the variable backgrounds, transparency, and seeing—can effectively "stress test" the Stack's algorithms, such as co-addition and PSF estimation. As part of the Summer 2013 Data Challenge, we—the LSST Data Management Team—reprocessed all Stripe 82 imaging to test the accuracy and scalability of the prototype data management software (Axelrod et al. 2010; Kantor et al. 2015). This reprocessing yielded the necessary data for this QLF study: *morphology, colors*, and *variability*. The reprocessing included the following steps:

1. Co-adding the imaging in all 5 bands

2. Detecting sources on the co-adds down to a 3σ detection threshold

3. Measuring the PSF flux at the positions of the i-band co-add detections on the u- ,g- ,r,- i,- and z- single-epoch images

These new measurements—lightcurves, colors, and morphologies—proved ideal for use in selecting fainter $z \sim 4$ quasars, the focus of this study. In this chapter, we describe the extraction of these lightcurves from this imaging set, and the use of these measured properties as features with which to select a uniform sample of $z \sim 4$ quasars. First, we discuss methods for multi-epoch photometry (§2.2). Next, we describe the Stripe 82 observations (§2.3); we also describe the specific algorithms used to co-add these images and create the forced photometry lightcurve catalogs (§2.4). In section §2.5, we present the results of our quality assessment and characterize the catalogs and co-adds. We conclude in §2.6.

2.2 Background: Optimal Multi-Epoch Photometry

How do we aggregate information from heterogenous multi-epoch images with the goal of extracting information with the highest possible signal-to-noise ratio? Traditionally, there are two ways that information has been combined in multi-epoch surveys. In the first way, sources are detected and measured on each single-epoch exposure, the measurements are matched between epochs, and the weighted mean of the measurements is taken. This method inherently discards information. Take, for example, a periodic star with a magnitude near the single-epoch detection threshold (Figure 2.1). When the star brightens, it is above the detection threshold, and when it dims, it is below the detection threshold—too faint to be detected. If measurements are taken only when the source is above the detection limit, the statistics will be biased high. Any variability metrics could also be biased, because they

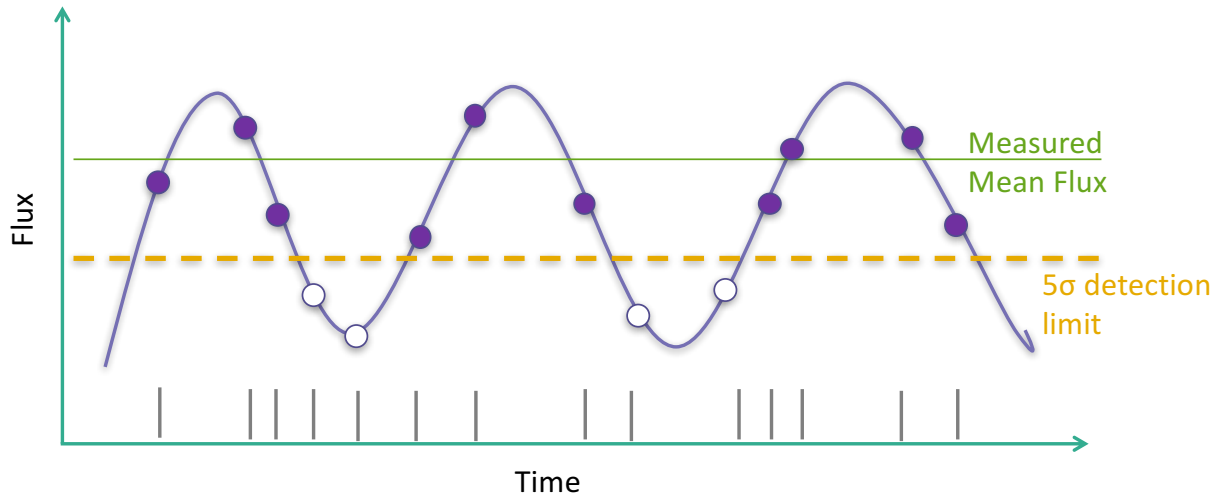

Figure 2.1 Cartoon of a periodic lightcurve near the single-epoch detection threshold. The purple solid line shows the intrinsic lightcurve, and the filled circles represent observations taken when the star is brighter than the detection threshold. Empty circles represent observations that would be excluded from a lightcuve generated by matching detections from epoch to epoch. Statistics on only the detections, such as the mean (green line) would be biased.

are derived from a lightcurve with missing epochs. Thus, detecting on single-epoch images automatically loses the information that some sources were, at times, too dim to be observed.

In the second method for combining information in multi-epoch surveys, an image co-add is created, and sources are detected and measured on this co-added image. However, measuring on a co-add also results in information loss, for three reasons. First, poor-seeing and bright-sky images do not contribute to the co-add's signal-to-noise ratio and are down-weighted or excluded (Section 2.2.2). Second, for accurate flux measurement, the co-added images require a good PSF model, but PSFs in co-added images are hard to characterize. PSFs in co-added images are irregular because they are the superposition of all the single-epoch PSFs. Each region of the co-add is an aggregate of a different set of input images,

which makes the PSF discontinuous at exposure or mask boundaries. This complication makes it difficult to construct a model PSF using star images from the co-add itself; the PSF is actually undefined for sources at these mask and chip boundaries. This problem has been addressed in various ways. Some researchers, for example, have homogenized the PSFs to a common co-add PSF by convolving them using a matching kernel (Huff et al. 2011). But because this convolution can only broaden the PSF, the PSF in the co-add is only as good as that in the worst-seeing image. Others have created a detection map or likelihood image by correlating each image with its own PSF (see Section 2.2.1 for a discussion) before stacking and measuring on the co-added detection maps (Kaiser 2001; Zackay & Ofek 2015) and (Lang et al. in prep, Bosch et al. in prep). But for any type of measurement other than PSF measurement, the co-add must be transformed back from a likelihood image to a real image. Others have used the `StackFit` algorithm (Jee et al. 2013), which makes a co-add of the the PSFs and maps PSF changes across the co-add. But despite all these techniques, one insurmountable challenge remains in measuring sources on a co-add: the sky is not static.

Stars move through the Galaxy. Quasars brighten and fade. Supernovae explode in galaxies. And once all the epochs have been stacked into a co-add, the temporal axis gets squashed. It becomes impossible to distinguish between, for example, stars with high proper motion, which appear blurred in a co-add, and extended sources such as galaxies (Lang et al. 2009). Furthermore, the co-add imprints the cadence of the telescope. For example, Stripe 82 was observed more times in the last 2 years of the survey than in the other 8 years combined. In a Stripe 82 co-add, the centroids of stars with high proper motion will thus be at the stars' actual position towards the end of the survey.

Acknowledging that the sky is not static requires a change in thinking. In one sense, every image taken with a telescope is a co-add. While the telescope shutter is open, the camera is recording a variable sky. The PSF is dancing around, and asteroids are zipping by. And yet we are comfortable saying that image represents the sky at the midpoint of the exposure. In a hypothetical case where all images have the same PSF, transparency, sky-backgrounds, exposure times, and pixelization, a co-add is like having the shutter open

longer. This is where the comfort with co-adds comes from. But these assumptions of homogenous conditions are never valid. And in realistic images, information is always lost.

We will therefore use a hybrid approach: we detect sources on the co-added images with a `StackFit` PSF, and then go back to the single-epoch exposures to measure those sources. Measurement on the single-epoch exposures can be done either by "Forced Photometry," forcing PSF flux measurement[1] at the co-add source centroids, or by "`MultiFit`," fitting the flux, centroid, and shape parameters simultaneously to the stack of images. This release provides a proof of concept of the forced photometry method. Because the centroids are fixed, high proper motion stars will look like they are increasing in brightness.[2]

The next section (§ 2.2.1) offers a detailed explanation of why an accurate PSF model is critical for accurate flux measurements, and section 2.2.2 explains how the procedure for optimal point source measurement can be used in constructing co-adds.

2.2.1 Point Source Detection

To guide intuition about point source detection, we first consider optimal point source measurement. Consider an image of an isolated star with a well-sampled PSF and sky-background-dominated noise. The atmosphere, telescope, and detector spread out light from the star onto multiple pixels. In the simplest form of photometry, known as aperture photometry, the counts (ADUs) of the pixels within a finite radius of the source centroid are added up, and the total background (the background level $\times$ the number of pixels) is subtracted. This can be reframed as a weighted mean, with weights equal to a tophat filter placed at the source centroid. Weights are 1 out to some radius and 0 outside. We know the PSF better than that, however. We know that the PSF usually directs more flux toward the center of the footprint and less flux toward pixels farther from the center. We can model this PSF using the profiles of point sources in the image. It turns out that the optimal point

[1] or model of the shape/profile of a galaxy.

[2] They will only *appear* to be increasing; because the Stripe 82 cadence more heavily weights the later years, it is then that the centroids are at the position of the star.

source photometry is mean weighted by the PSF profile, a kernel which sums to unity.[3] We call this quantity the PSF-flux.

We can also use the knowledge of the PSF to optimize point source detection by pre-computing the PSF-flux at every pixel. We often casually refer to this as "correlating the image by the PSF" or (when the PSF is symmetric) as "convolving the image by the PSF." We call the result of this cross-correlation a "detection map" or a "likelihood image." The term likelihood image is used because the PSF-weighted flux estimate is also the maximum likelihood estimate of the true flux of a point source.

It may be useful to think of this detection map as the result of applying a spatial filter (Lang et al. in prep). The best way to find a known signal in noise is to cross-correlate the image with a "matched filter," a kernel with the profile we hope to find. The name "filter" comes from the use of filters in the frequency domain, but these filters work the same way in the spatial domain.

This procedure, which is a test of the hypothesis that there is a point source centered at every pixel, is *not* the same as what the image-processing literature calls "image restoration" (see e.g. Gonzalez & Woods 2011). Image restoration repairs, or seeks to sharpen, blurred and noisy images; this is not the same thing as generating detection maps, which are even more blurry than the original image. Because degradations are modeled as being the result of a convolution, the restoration seeks to find the filters that apply the process in reverse: deconvolution. Therefore, the terms "image deconvolution" and "image restoration" are often used interchangeably. Because the degradation function must be estimated from the image itself, it is sometimes called "blind deconvolution." In astronomy, estimating the degradation function is feasible; astronomical images are filled with stars that we know to be point sources. Furthermore, estimating the degradation function by constrained least squares filtering requires only knowledge of the mean and variance and yields an optimal result for

[3]Optimal photometry is *not* weighted by the per-pixel variance, which depends on the flux (and would introduce biases). Whereas the per pixel flux reflects the actual noise, we want the optimal aperture flux, where the maximum-likelihood PSF-convolved model is the "aperture." (Jim Bosch)

each image (Gonzalez & Woods 2011). However, this process is computationally expensive. While promising, image restoration has not been used in astronomy, both because it is computationally expensive and and because the result is extremely sensitive to the accuracy of the PSF model.

Because the optimal point source photometry uses a PSF-weighted sum of fluxes, and because detection can be framed as a cross-correlation of the image with the PSF, it is clear that an accurate model of the PSF is critical for both optimal point source measurement and detection.

2.2.2 Co-addition

Although not optimal for deep source measurement, co-adds are still useful for deep source detection with which to seed forced photometry or for providing initial parameters to `MultiFit`. We therefore want to take care to construct these in a way that enables these goals. In constructing a co-add, a statistic (e.g. weighted mean) is computed as an estimator for the PSF-convolved sky in each pixel. Weighting the individual epochs that go into the co-adds can change the variance of the co-add, and this weighting can be optimized for different purposes.

Let $g(x, y)$ denote an image formed by the addition of noise, $\eta(x, y)$, to a noiseless image $f(x, y)$:

$$g(x, y) = f(x, y) + \eta(x, y), \tag{2.1}$$

where the assumption is that at every pair of coordinates (x, y) the noise is uncorrelated and has a zero average value. While we make no assumptions about the shape of the distribution that $\eta(x, y)$ draws from, it has an expected value $\mathrm{E}(\eta_i) = 0$ and variance $\mathrm{Var}(\eta_i) = \sigma^2_{\eta_i}$. The objective of co-addition is to reduce the noise, or increase the signal-to-noise ratio (SNR), by adding together the set of noisy images $g_i(x, y)$ to estimate $f(x, y)$.

Let $\bar{g}(x, y) = \frac{1}{K} \sum_{i=1}^{K} g_i(x, y)$, the average of K noisy images. It follows that the expected

value of $\bar{g}$, $\mathrm{E}\{\bar{g}(x,y)\} = f(x,y)$. It also follows that the variance of $\bar{g}$ can be expressed as

$$\mathrm{Var}[\bar{g}(x,y)] = \mathrm{Var}\left[\frac{1}{K}\sum_{i=1}^{K}\eta_i(x,y)\right] \tag{2.2}$$

$$= \frac{1}{K^2}\sum_{i=1}^{K}\mathrm{Var}\{\eta_i(x,y)\} \tag{2.3}$$

$$= \frac{1}{K}\mathrm{Var}\{\eta(x,y)\} \qquad \text{iff } \eta_i = \eta \tag{2.4}$$

$$= \frac{\sigma_\eta^2}{K} \qquad \text{iff } \eta_i = \eta. \tag{2.5}$$

The last step makes no assumptions about the noise model that $\eta(x,y)$ is drawn from, except that it is the same for all epochs. This assumption of equal-variance epochs yields the familiar result,

$$\sigma_{\bar{g}(x,y)} = \frac{1}{\sqrt{K}}\sigma_{\eta(x,y)}. \tag{2.6}$$

However, it is rarely the case that epochs have equal variance. In any real set of images,

$$\mathrm{Var}[\bar{g}(x,y)] = \frac{1}{K^2}\sum_{i=1}^{K}\mathrm{Var}(\eta_i(x,y)), \tag{2.7}$$

it is apparent that the variance of the unweighted mean co-adds is dominated by the variance of the worst, highest-variance epochs. Therefore, in astronomical co-addition, images are typically summed using a mean that is weighted to optimize a particular quantity. For the simplest two cases, we want find the co-addition statistic that minimizes the variance of the co-add or the statistic that maximizes the point source measurement SNR, which we will show later (§2.4.1.7) when comparing the minimum-variance weighting used in this work to other available Stripe 82 co-adds.

Methods for finding minimum-variance estimators are well established. The inverse-variance weighted mean is the minimum-variance unbiased estimator for non-constant η_i. The expectation of the weighted mean is again f.

$$\mathrm{E}\left[\frac{\sum g_i/\sigma_i^2}{\sum 1/\sigma_i^2}\right] = \mathrm{E}\left[\frac{\sum_{i=1}^{K}(f(x,y)+\eta_i(x,y)/\sigma_i^2}{\sum 1/\sigma_i^2}\right] \tag{2.8}$$

$$= f(x,y) \tag{2.9}$$

And the variance is:

$$\mathrm{Var}\left[\frac{\sum g_i/\sigma_i^2}{\sum 1/\sigma_i^2}\right] = \mathrm{Var}\left[\frac{\sum(f(x,y)+\eta_i(x,y))/\sigma_i^2}{\sum 1/\sigma_i^2}\right] \tag{2.10}$$

$$= \frac{1}{\sum 1/\sigma_i^2} \tag{2.11}$$

2.3 Observations: Stripe 82

In this section, we describe the observations used as inputs to the LSST Data Management pipeline. The area of interest for the Data Challenge is centered on the celestial equator, spanning 2.5° in Declination between -40° < *R.A.* < +55°, which was imaged in five photometric bands over ∼ 80 epochs spanning 10 yearsr.

The SDSS uses a dedicated wide-field 2.5 m telescope (Gunn et al. 2006) at the Apache Point Observatory (APO). The imaging instrument (Gunn et al. 1998) is a wide-field camera composed of 30 CCDs. Each CCD has 2048 × 2048 pixels, with a pixel scale of 0.396″. The telescope surveys the sky in drift-scan mode, such that while the sky passes over it, the CCDs continuously read out at the sidereal rate. The CCDs are laid out in six camera columns, or "camcols," and five rows. Each row has one of five filters (u, g, r, i, z) (Fukugita et al. 1996). The filters are imaged in the order *riuzg*, with an effective exposure time per filter of 54.1 s. The camera can image 18.75 deg^2 per hour. On average, point source 50% completeness limits for u, g, r, i, z are 22.5, 23.2, 22.6, 21.9, and 20.8, respectively (Abazajian et al. 2003).

Because of its unique camera, the SDSS uses unique terminology to describe the survey geometry. The camera geometry is closely tied to the survey strategy, which in turn strongly influences the organization of the output SDSS image files. Because we use terminology

throughout this paper, we offer the following summary of the SDSS survey geometry, which will aid in understanding how the geometry maps to the terminology of this data release.

A single continuous drift scan obtained on a single night is called a "run." A run is a single night's imaging of a "strip." In other words, a strip can be imaged by many runs. Each strip consists of six parallel "scanlines" (one for each of the camcols), each 13 arcmin wide, with gaps of about the same width. A survey "stripe" (not to be confused with a strip) is one camera-width wide, about 2.5°. Because of the gaps between the scanlines, two interleaving runs/strips are needed to cover a stripe. Runs are labeled N or S depending on whether the telescope is pointed half a CCD-width north or south of the stripe equator. In the case of Stripe 82, the stripe equator is the celestial equator.

Each scanline is a 13′ wide stream of data, which is partitioned along the scan direction into individual, overlapping images for processing. An individual image in a single filter is called a "frame." The set of frames from all filters covering the same region of sky is called a "field." A frame is 1489×2049 pixels, with the short side in the R.A. axis. Adjacent frames overlap by 128 pixels.

Stripe 82 is the SDSS stripe along the celestial equator in the Southern Galactic Cap and covers the $300\,\mathrm{deg}^2$ from $-50 <$ R.A $< +60$, $-1.256 <$ Dec. $< +1.256$. It was repeatedly imaged for a decade during the summer and fall months. In the seasons before 2005, Stripe 82 was imaged only under photometric conditions (the same conditions under which the main survey was imaged). There were 84 runs in this stage. In 2005–2007, there were 219 runs as part of the supernova survey (Frieman et al. 2008) designed to discover Type Ia supernovae at $0.1 < z < 0.4$. Because this survey operated on most usable nights, the dataset includes images taken on nights with poor seeing, bright moonlight, and non-photometric skies.

Because Stripe 82 is on the celestial equator, equatorial coordinates uniquely map onto pixel coordinates. Rows increase with R.A. and field number. Columns increase with Declination, as do camcol numbers. For more information on the SDSS nomenclature and technical terms, see (Stoughton et al. 2002).

2.3.1 SDSS Photometric Processing

The SDSS image processing pipeline determines astrometric calibration (Pier et al. 2003) with an average RMS of 45 mas (Abazajian et al. 2009). It then detects objects and measures their brightness, position, and shape (Lupton et al. 2001; Stoughton et al. 2002). Photometry is calibrated to the AB system (Oke & Gunn 1983), and zeropoints are known to 1–2% (Abazajian et al. 2003, 2004).

Photometric calibration is done with two methods. First, standard stars (Smith et al. 2002) are measured by a separate 0.5 m telescope on site (Tucker et al. 2006; Ivezić et al. 2004). Second, the overlap between adjacent imaging runs ties the photometry together in a process called "ubercalibration" (zeropoints have RMS errors $\sim 2\%$ in u and $\sim 1\%$ in $griz$).

We use the intermediate outputs from the SDSS DR7 image processing as inputs to our co-addition and photometry software. Details about the SDSS file formats can be found in `http://classic.sdss.org/dr7/algorithms/dataProcessing.html`. These outputs include:

- `fpC`: Corrected frame files (bias subtracted, flat-fielded and purged of bright sky)
- `fpM`: Frame binary masks
- `tsField`: Calibrated field statistics which include zeropoints
- `psField`: Initial photometric calibration and PSF fits

The special geometry required a few adaptations to our processing software. We make notes throughout this text on the steps that require adaptation.

2.4 Building the Co-adds and Catalogs

As part of the Summer 2013 Data Challenge, we—the LSST Data Management Team—reprocessed nearly all Stripe 82 imaging to test the accuracy and scalability of the prototype

data management software (Axelrod et al. 2010; Kantor et al. 2015). We processed only the 238 $\deg^2$ sub-region in the range: $-40° <$ R.A. $< +55°$ This range corresponds to $-32.5° < b < -41.3°$ and was to avoid areas close to the Galactic plane, which are crowded with stars and dust.

The reprocessing included the following steps:

1. Co-adding the imaging in all five bands

2. Detecting sources on the co-adds down to a 3σ detection threshold

3. Measuring the PSF flux at the positions of the i-band co-add detections on the *ugriz* single-epoch images

2.4.1 Co-add Construction

To detect deep sources with which to seed the forced photometry, we constructed co-adds in all filters. Figure 2.2 summarizes the steps undertaken to build a co-add, including sampling, zeropoint matching, background matching, and stacking the images. The result yields co-adds with a high SNR, with a background consisting of both diffuse sources and the temporal background of the reference exposure. These steps are described in more detail in the following subsection.

2.4.1.1 Single-Epoch Image Pre-processing

The first step was to reprocess the DR7 `fpC` frames to estimate the zeropoints and the shape of each frame's PSF. The PSFs were modeled as a sum of 2-D Gauss-Hermite shapelet basis functions (Bosch 2010) per single-epoch image. Sources were flux-calibrated per single-epoch image using zeropoints derived by fitting bright stars to a standard star catalog (Ivezić et al. 2007).

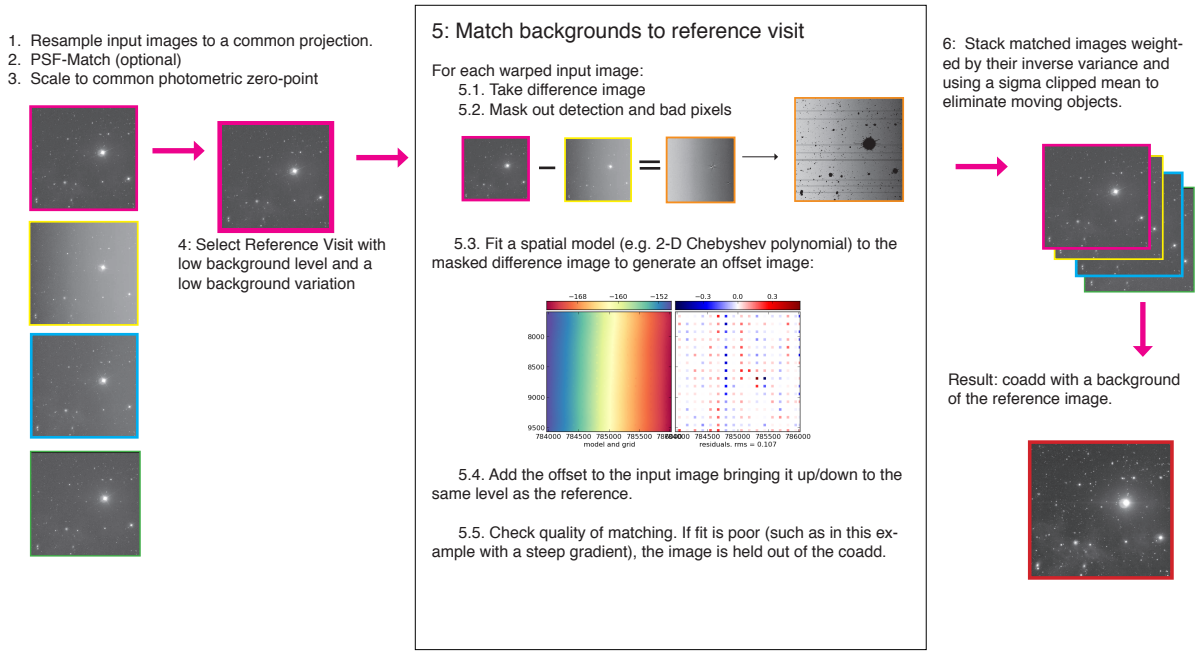

Figure 2.2 Steps to generate a background-matched co-add.

2.4.1.2 Co-add Run and Frame Selection

For inputs into the co-add, we used only frames with reasonable seeing, having a PSF with a FWHM $< 2''$, and an SDSS Quality Flag of 1 (`GOOD`) or 2 (`ACCEPTABLE`).[4] Input frames were visually inspected to flag major problems such as clouds or offsets in amplifier gains. The SDSS CCDs were read out with two amplifiers. In some runs, the two amplifiers had mismatched gains such that the two halves of an frame had obviously different background levels. We blacklisted some of these runs; the blacklisted runs are listed in §2.7.1. This filtering had a minimal effect, because the fraction of affected frames is $< 2\%$.

[4]Recall that we ran forced photometry in *all* frames, and all runs are represented in the lightcurves.

2.4.1.3 Co-add Projection and Geometry

Before stacking images, we must first resample them onto a common pixelization defined by a projection and geometry. Inherent trade-offs exist in the projection choices. Ideally, we would like the projection to be both conformal, which does not distort shape, and equal area, which preserves flux. Because the sky is curved, it is impossible to achieve both of these conditions at once, and thus there are two main families of projections: equal-area and stereographic.

Because it would take too much computer memory to stack 238 $\deg^2$ images, we define a geometry to split this projection into smaller images. We split the 238 $\deg^2$ region into two "tracts." The first tract, labeled 0, extends from $0° < R.A.$¡$90°$, and the second tract, labeled 3, extends from $270° < R.A. < 360°$.[5] We split these tracts into "patches" of 2000 ($13.2'$) $\times$ 1907 ($12.5862'$) pixels. A 30-pixel border was added to each patch to facilitate both mosaicking and source detection along the patch edges. In the declination axis, 12 patches correspond to the 6 camcols/scanlines and the 2 strips (N and S) that make up Stripe 82.

We used a cylindrical equal area (CEA) projection with native pixel scale ($0.396''$). Stripe 82 extends only $1.256°$ away from the equator, which means that the distortion of the pixels is at most 0.99976, 2.5 thousandths of a pixel.When choosing pixel scale, trade-offs between image quality and computational cost must be balanced. The scale of the resampled pixels must be at least the native pixel scale to preserve information. Pixels that are too big under-sample the original data; pixels that are too small over-sample the original data and carry around unnecessary bits.

2.4.1.4 Resampling to Co-add Projection

For each patch, we found all overlapping frames that met the seeing and quality criteria. The overlapping frames were resampled or warped onto the output projection—the most

[5]Geometrically, there are also tracts 1 and 2, which cover the North Galactic Cap, but do not intersect Stripe 82.

expensive step in the co-addition process. Theoretically, a sinc filter is the optimal reconstruction filter for band-limited signals of infinite extent. The Lanczos filter is one finitely supported, practical approximation of a sinc filter that is well-behaved (Duchon 1979) in terms of aliasing.[6] The SDSS pixel scale adequately samples the PSF at APO, but all ground-based images have finite bandwidth. We resampled using a 2-D, 4th-order Lanczos kernel: a sinc function windowed by the central lobe of a second, longer sinc function. In one dimension, the 4th-order Lanczos kernel is: $K(x) = \mathrm{sinc}(x)\mathrm{sinc}(x/4)$, for $-4 < x < 4$. The 2-D interpolation formula is $\hat{I}(x, y) = \sum_i \sum_j I_{i,j} K(x - i) K(y - j)$, where x, y are the output image pixel coordinates and i, j are the input image pixel coordinates.

We do not convolve the images to match the PSF to a common PSF, a process sometimes referred to as PSF-homogenization (Huff et al. 2011). PSF-homogenization would be needed if we wanted to get accurate measurements from the co-add, but we do not. Remember that we only intend to use the co-adds for detection and **not for measurement**.

2.4.1.5 Scaling to Common Zeropoint

The resampling step produced patch-sized images of each run, called a "co-add temporary exposures," `coaddTempExp`'s, or a "warps" by other projects. Next, these warps are scaled to a common zeropoint of 27th magnitude by multiplying by a scale factor computed from the frames' original measured zeropoint.

Zeropoints vary temporally and spatially because of cloud structure and movement. We assume that over a single patch ($13'$) there is no spatial structure in the zeropoint. However, we do account for temporal changes to the zeropoint.

Because the camera is a drift scan camera, the west side of the patch was scanned approximately 1 min before the east side. A minute is longer than the timescale on which the transparency of the sky varies. This variation manifests as a one-dimensional variation in the zeropoint along the R.A. axis or temporal axis. Therefore, we model the zeropoints

[6] In general, aliasing is the outcome when high frequency components of a continuous function masquerade as lower frequencies in a sampled function because the function was under-sampled.

fluctuation as a one-dimensional function of R.A.[7] Zeropoints are calculated in the single-epoch processing step before co-adds are generated. The calibrated zeropoint for each frame is assigned to the R.A. at the midpoint of the frame.

Figure 2.3 shows the variation in the i-band zero point over the course of 72 min, for three interpolation orders. The bottom line shows the zeropoints without interpolation, and the shading shows the measured uncertainties in the zeropoints. The middle line shows the same with linear interpolation. The top line shows the same for cubic spline interpolation. We modeled the zeropoints by interpolating between the midpoints using a natural cubic spline.

2.4.1.6 Matching Backgrounds

Exposures taken at different times exhibit temporal and spatial variation in background levels. The photometric pipeline aims to remove the non-astrophysical background (e.g., airglow, scattered moonlight), while preserving the common-mode astrophysical background (e.g., unresolved sources, wings of galaxies). In practice, however, the two are difficult to separate. For example, a diffuse astrophysical structure such as galactic cirrus can span an area wider than a single chip, and may be compromised by temporal effects like cloud cover. To minimize the non-astrophysical background when generating co-added images, we use a background-matching algorithm based on the method used by Huff et al. (2011).

The background-matching procedure starts with the selection of a reference image. The reference run locks the sky-level of a co-add patch to that of its adjacent neighbors, making it possible to mosaic the patches into a larger tract co-add; this ensures the tract co-add will be seamless at patch boundaries. We chose runs with good seeing and low background variation as a reference images: run number 5823 (6955) for the N (S) patches in tract 0 and 4207 (4192) for the N (S) patches in tract 3.

Then for each input science patch S_i of run i we perform the following steps:

[7] This is actually a 1-D model of zeropoint versus MJD as well, but the mapping between R.A. and pixel is already provided by the WCS which produced a simpler implementation.

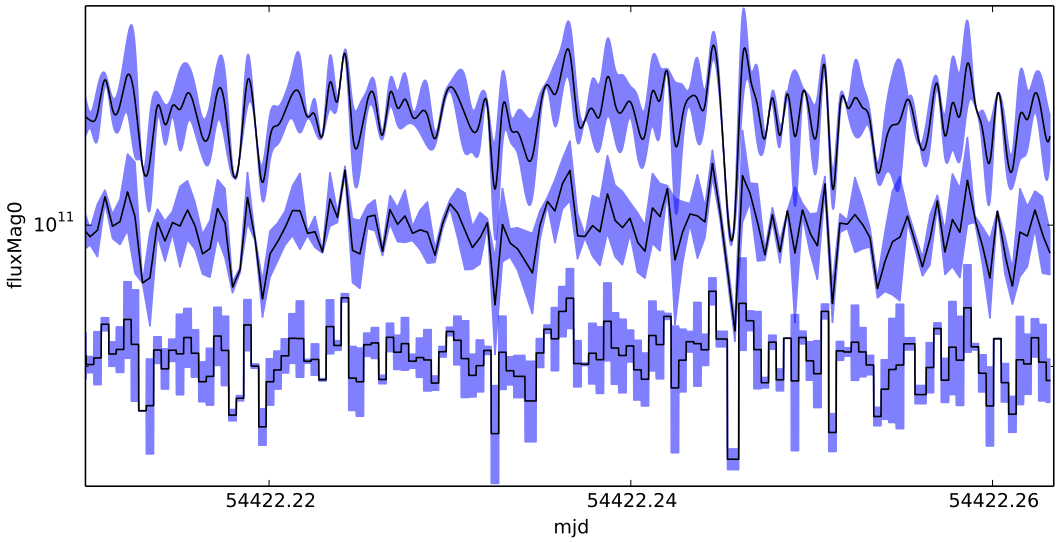

Figure 2.3 Variation in the i-band zeropoint for one camcol over 72min for three interpolation orders. The bottom line shows the frame zeropoints without interpolation, and the shading shows the measured uncertainties in the zeropoints. The middle line shows the same with linear interpolation. The top line shows the same for cubic spline interpolation. When rescaling images to a common co-add zeropoint, we used 1-D model of zeropoint (per camcol) as multiplicative scale factor. We interpolate the zeropoint variations as a function of R.A.

1. Take the difference image $(S_i - R)$, where R is the reference image.

2. Mask out detections and bad pixels in the difference image. $\text{Mask}_{S-R} = \text{Mask}_S \vee \text{Mask}_R$.

3. Fit a 2-D spatial model to the masked difference image to generate an offset image, O. We used a 4th order Chebyshev polynomial to model the background offset. We chose a smoothing model instead of an interpolating model such as 2-D splines because smoothing models are more robust to outliers; 4th degree was was chosen specifically for the size of a frame. See §2.7.2 for more detail on modeling these offset images.

4. Add the offset model to the input science image, matching the background level to that of the reference image. This produces a matched science image, $M_i = O_i + S_i$.

5. Check the quality of the match by computing two background-matching quality metrics: the RMS of residuals in modeling the offset, and the ratio of the mean squared error (MSE) to the variance of the background in the difference image. If one of these metrics does not meet a threshold, it is not included in the stack. The motivation behind the MSE/Variance metric follows. The difference of the background pixels of the reference image and the matched image $R - M$ should have a mean of zero. Its variance about zero should be within the expected variance of the matched image: $\sigma^2_{R-M} = \sigma^2_R + \sigma^2_M = \sigma^2_R + \sigma^2_S$, where σ^2_S is the variance plane of the science image and σ^2_R is the variance plane of the reference image.[8] The mean squared error

$$\begin{aligned} MSE &= \frac{1}{P}\sum_{i}^{P}(R(p) - M(p))^2 \\ &\approx \frac{1}{P}\sum_{p}^{P}(\sigma^2_S(p) + \sigma^2_R(p)) \\ &= \text{diffIMVar}. \end{aligned} \tag{2.12}$$

   where p represents unmasked pixels[9] and P the total number of unmasked pixels. Therefore the ratio (MSE/diffIMVar) is a proxy for how well the background of the matched image matches the background of the reference image. A larger value means that the the background of the matched image deviates more from that of the reference. The threshold for the MSE/diffIMVar quality-of-matching metric is given by the config parameter `config.matchBackgrounds.maxMatchResidualRatio`. A different threshold is required for co-adding different bands. These thresholds were chosen empirically.

[8] Assuming $\sigma^2_M = \sigma^2_S$. Because we add a smooth polynomial offset image to the matched image, adding the matching offset should not introduce any extra variance in the matched image. However, this offset image will have some uncertainty associated with it. Indirectly, that is actually what we are testing for here.

[9] The "background" constitutes the unmasked pixels.

See §2.7.2 for more details.

We also evaluated the process of simultaneous solving for the background offsets, given the stack of N input images: for N input images, find least-squared solution of all $N(N-1)/2$ difference image backgrounds (e.g. Huff et al. 2011). Barris et al. (2005) call this algorithm "NN2." NN2 offers better relative offsets in sky levels, but does not provide an *absolute* sky level. NN2 does not eliminate the need for a reference image for a large continuous co-add. If you tie the sky-level of each patch to the mean or to zero, you lose all astrophysical background at larger scales than a patch. We compared the performance of NN2 to that of matching every image to a reference run In the limit where a good reference run is available, the improvements do not justify the additional computational cost. In evaluations, we found a negligible difference in background offset models using NN2 vs. a reference run.

This result is seemingly counter intuitive, but it makes sense in the limiting case: when there are no transient detections (i.e. patches have identical coverage) and when a linear estimator is used for the background (i.e. mean instead of median). For example, when seeking the single scalar offset that matches the backgrounds, the background offsets determined by NN2 exactly match those achieved using a single reference run. NN2 only recovers more information when using a non-linear estimator that discards information. It was developed to estimate lightcurves from difference images generated from lossy procedures. But in co-addition, information loss arises from masking pixels with source detections. NN2 may be worth the computational cost in situations when a good reference run is difficult to find, such as on the ecliptic, which might be heavily masked by asteroids. In reference runs that are heavily masked because of asteroid detections, the background matching procedure has no information about the backgrounds behind those asteroids, and NN2 would recover the information lost to this heavy masking.

Traditionally, backgrounds are subtracted from the individual input images before co-addition. Figure 2.4 shows, for a single patch, the single-epoch, background-matched co-add, and background-subtracted co-add. Both co-adds have visibly higher SNRs than the

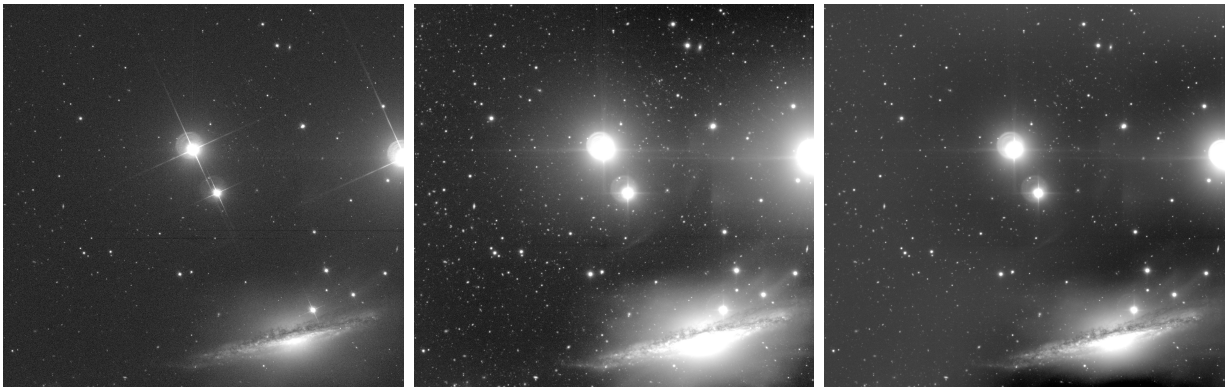

Figure 2.4 Single-epoch run (left), background-matched co-add (middle), and background-subtracted co-add (right) of an r-band, $13' \times 13'$ patch (tract: 0, patch: 225,8). The gradients in the background of the background-matched co-add better represents the sky. The background matched co-add has depressed background levels near the bright sources. These holes can bias the photometry in the vicinity of bright sources.

single-epoch image. However, the gradients in the background of the background-matched co-add better represent the sky. The background-subtracted co-add has depressed background levels (sometimes referred to as "holes") near the bright objects. These holes can bias the photometry in the vicinity of bright sources.

Background matching does not solve the problem of background *estimation*. One realization of the sky survives in the co-add, but because the co-add has a higher SNR than the individual input images, we can estimate and subtract this background more accurately. In other words, we are postponing the background estimation subtraction step until we have a co-add with a higher SNR.

Background estimation has serious implications for science done with photometric catalogs. For example, the sky-subtraction algorithm used by the DR7 photometric pipeline systematically underestimates the brightness of large galaxies (**?** and references therein). The sky background was over-subtracted in the wings of large galaxies, affecting the photometry of the fainter objects in their vicinity. The resulting proximity biases can affect

lensing signals (see Huff & Graves (2014) and references therein).

2.4.1.7 Stacking

For the final step, we stacked the ensemble of resampled and scaled runs, weighting each by the average inverse variance of the whole image. For each pixel, we sigma-clipped the epochs having flux counts that deviated more than 5σ from the mean to eliminate satellite trails, asteroids and small-scale instrumental scattered light. This 5σ cutoff is sufficiently high to not introduce problems from magnitude-dependent PSFs. Sigma-clipping with a high cutoff such as 5σ has the same effect as modeling each pixel as two Gaussians, a narrower one (representing the scatter in the photons emitted by that part of the sky over 100 observations) and a broader one (representing the possibility that an unusual event, such as an asteroid or spacecraft passing over, may happen).

The epochs vary in quality. Therefore, the images were weighted to optimize the SNR of the co-add, given the SNR_i of the individual epochs i. All weights are per image, not per pixel. Using different weights per pixel would produce a numerous bizarre biases, including a depression in the flux in sources relative to the background. Consider the ideal case in which we would have N images of identical SNR. In this case, the SNR of the resulting co-add would be $\sqrt{N} \times SNR$. This ideal case is never realized in practice, however, because some images have better SNR than others. For example, imagine an opposite scenario: two exposures were taken, but in one the dome light was accidentally left on. The SNR for this bad image is negligible compared to the good one: $(\mathrm{SNR})_{bad} << (\mathrm{SNR})_{good}$. In this case, the SNR of the resulting co-add should be approximately equal to that to the good image because the bad one adds no information.

Assuming that all noise sources are independent, the SNR of the measurement of flux from a star is:

$$SNR \propto \frac{N_{\mathrm{photons}}}{\sigma_{\mathrm{sky}}\sqrt{A}}, \tag{2.13}$$

where N_{photons} is the number of photons detected from the star, A is the area in pixels covered

by the star. The per-pixel sky noise σ_{sky} includes all sources of noise: dark current, read noise and sky-background, and it coded in the variance plane of the image. For the epoch i, $\sigma^2_{i,\mathrm{sky}}$ is the average of the variance plane. We emphasize again that $\sigma^2_{\mathrm{sky},i}$ is **not** the variance per pixel, but the variance of the whole image. The N_{photons} is proportional to transparency T, and the area that the stellar photons cover depends on the seeing: $A \propto \mathrm{FWHM}^2$.

Therefore, a co-add that was optimized for point-source detection would, to weight each image by the SNR^2, use the following as weights:

$$w_i = \mathrm{SNR}^2 \propto \frac{T_i^2}{\mathrm{FWHM_i}^2 \sigma_i^2}. \tag{2.14}$$

This weighting prefers good-seeing epochs taken when the sky is transparent and dark. Co-adds with variants of this weighting were produced by Annis et al. (2014) and Jiang et al. (2014). Jiang et al. (2014) used simulations to empirically find a preferred weighting of $w_i = \frac{T_i}{\mathrm{FWHM_i}\sigma_i^2}$. The SDSS co-adds (Annis et al. 2014) were released with an effective weighting of $w_i = \frac{T_i}{\mathrm{FWHM_i}^2\sigma_i^2}$.[10] Because we use our co-adds for a multiple purposes (among these point source detection), but also extended source detection and recovery of low surface brightness features, we chose the usual inverse-variance weighting which produces the minimum-variance co-add. The weights are given by,

$$w_i = \begin{cases} T_i^2/\sigma_i^2 & \mathrm{FWHM} < 2'', \\ & \texttt{Quality} \leq 2, \\ & \text{Bkgd-match successful} \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

As implemented, the factor of T_i^2 incorporated into the variance when flux-scaling the single-epoch images to a common zeropoint. This step *multiplies* the image by a scale-factor, which increases the variance of the image by the square of the scale factor. The scale factor is inversely proportional to the transparency, so that $\sigma_{scaled} = \sigma_{sky}/T$. We then weighted each epoch by the inverse-scaled-variance $1/\sigma^2_{i,scaled}$.

[10] According to the authors, they intended to use equation 2.14, but the pipeline used this weighting instead because of a bug.

Figure 2.5 shows the number of images that contributed to each patch in the $238\,\mathrm{deg}^2$ co-add. There were more N runs than S runs on average. The patches with $N < 30$ are patches with very bright stars, galaxies and clusters.

2.4.1.8 Auxiliary Co-Added Data Products

`CoaddPsf`: A challenge in producing quality measurements from non-PSF-matched co-adds is the complexity of the effective point-spread function on the co-add. Because the PSF has discontinuities at the location of chip boundaries, modeling approaches based on interpolating with smooth functions cannot be used. It is difficult to construct a model PSF using star images from the co-add itself.[11] Instead, when creating a co-add image, we also combine the PSF models of all the input exposures, using an approach similar to `StackFit` devised by Jee et al. (2013). We call this construction `CoaddPsf`. With each co-add, we store the single-epoch PSF models, their bounding boxes, the weight w_i of the epoch, and the coordinate transforms that relate them to the co-add pixel grid. Then, when measuring an object on the co-add, we can evaluate the co-add PSF model at that location by co-adding the PSF models of all exposures that contributed to the relevant part of the co-add, after warping them by the appropriate coordinate transform and weighting them by the same weighting used in assembling the co-add. We approximate the PSF as spatially constant over the footprint of a source.

2.4.2 Catalog Construction

To construct catalogs, we detected and measured shapes, centroids, and fluxes on the co-adds and force PSF-flux measurement on the single-epoch images at the centroids of the co-add detections. This process resulted in two separate catalogs: flux measurements of the averaged sky and lightcurves of the variable sky. On the co-adds, we then detected and measured PSF-fluxes and model-fluxes for all sources above 3σ on all five bands using co-

[11] This problem is the reason that the LSST plans to measure extended objects by simultaneously fitting the sources on the individual input images, rather than directly on the co-add.

added PSF models, `CoaddPsf`. We call each row in this catalog a `DeepSource`. If an object is detectable in all five bands, then it will have five `DeepSources`.

Then, to produce lightcurves and colors we measured flux on the single-epoch images. For each *i*-band `DeepSources`, we measured the PSF-flux on each of the *ugriz* single-epoch images at that centroid of the *i*-band `DeepSource`, *regardless of whether a source detected in that single-epoch image.* Each *i*-band `DeepSource` was measured, on average, 80 times in each of five bands, producing a catalog of `DeepForcedSources` with 16 billion rows.

2.4.2.1 Detection on the Co-adds

With the `coaddPSF`, we measured both PSF fluxes and multi-shapelet model fluxes after running a deblending algorithm. To detect sources, the image was first cross-correlated with a PSF profile. Pixels above a 3σ threshold in this smoothed image are flagged, and groups of contiguous pixels, called "detection footprints," are measured to determine the centroid positions and fluxes of possibly blended sources. These detection"footprints," were deblended using a deblending algorithm. The deblender decomposes footprints by finding peaks and assuming that galaxies are symmetrical. We limited the deblender to at most 40 children. This limit was both high enough to capture all sources in crowded areas, such as galaxy clusters, and low enough to prevent the shredding of brighter stars and glints into an unmanageable number of fake sources. We found that the deblender behaved poorly in these footprints near bright stars (§2.5.2). Model fluxes and shapes were measured using a multi-shapelet approach wherein galaxies are approximated as a sum of low-order elliptical shapelet expansions (Bosch 2010). Both the PSF fluxes and model fluxes require an accurate model of the PSF.

Star/Galaxy Separation: We reuse from the SDSS a simple metric to separate point sources and extended sources: $2.5\log$[modelFlux/psfFlux], which we label `extendedness`. For true point sources, the best-fitting source model should approximate the PSF, and the model flux and PSF flux should be equal. Therefore, we expect point sources to have an `extendedness`' of 0 and extended sources to have an `extendedness` of > 0. However, the

model fluxes are not dependable in all scenarios. We noticed that for point sources close to bright stars ($i < 17$) the deblending algorithm assigns some flux from the bright star to the children sources. This makes the children have larger model fluxes and appear extended (§2.5.2).

2.4.2.2 Forced Photometry on Single-Epoch Images:

After detecting sources on the co-adds, we measured the PSF flux at the centroid of each i-band co-add detection on each of the $5\times \sim 80$ single-epoch exposures. This produced a 16-billion row forced-photometry catalog of lightcurves in 5 bands of 40M objects. The positions were forced and were not recentered from epoch to epoch. The PSFs were modeled as a sum of 2-D Gaussian shapelets (Bosch 2010) for each single-epoch image. Sources were flux calibrated for each single-epoch image using zeropoints derived by fitting bright stars to a standard star catalog (Ivezić et al. 2007). Backgrounds were estimated by binning the backgrounds and interpolating the bin means with a 2-D spline.

The camera and processing procedure imparted a few unique characteristics to these lightcurves. First, the timestamps can be up to 27 s off, and a single lightcurve can contain data points 27 s apart. Sources that fell on the 128-pixel overlap region were force-measured twice, appearing in a lightcurve as two data points with timestamps seconds apart. These duplicate measurements were useful for validating the processing outputs. In the `DeepForcedSource` catalog, if two forced sources in a single run appear, then in post-processing we marked as "not-primary" the row that had measurement flags; if neither row was flagged, we denoted the row closer to the edge of the frame as "not-primary." Additionally, with a drift-scan camera on the equator, time increases with R.A. Thus, technically, every pixel column has a slightly different midpoint timestamp. For simplicity, however, the timestamps in the catalogs correspond to the center of each frame. When studying objects that vary on timescales longer than a minute, the timestamp discretization is inconsequential.

Second, some forced flux measurements are negative. This effect is *expected* in images that have been properly background-subtracted and calibrated. For example, consider a red

source that was detected in the i-band, but that was well below the detection threshold in the u-band. If the expected flux is zero and the flux is measured at the position of the i-band detection in 100 u-band epochs, we would expect 50 positive measurements and 50 negative measurements. The median of these 100 measurements would be nearly zero. Furthermore, in any set of null detections, half of the medians would be slightly negative as well. Although negative forced-flux measurements require a change in thinking in using catalog data, distribution of non-detections about zero offer more information than the source simply missing from the catalog.

There are two major benefits of using forced photometry for producing lightcurves: the lightcurves are more complete and the resulting aggregates (mean fluxes) are less biased. First, the lightcurves are more complete. By detecting on the deep co-add and measuring on the single-epoch exposures, we get lightcurves of objects that are too faint to be detected in the single-epoch exposures. Lightcurves generated the traditional way, by detecting on individual single-epoch images and matching them, would miss some epochs. Even a flux measurement with large uncertainties (or an upper limit for measured fluxes that are negative) offers some information about the epoch that would otherwise be missed, and some information is better than no information. Figure 2.6 illustrates this idea with lightcurves for two variable quasars close to the detection limit. For each quasar, the lightcurves derived from reprocessed forced photometry contain more epochs than the lightcurves derived by matching the single-epoch detections.

Second, the flux aggregates are less biased. Flux aggregates are median (or sigma-clipped, weighted-mean) measured fluxes, and are sometimes called a "catalog co-add." Figure 2.6 shows that the average fluxes for the two faint quasars are both lower and less biased in the more complete forced-photometry lightcurves than in the lightcurves derived from matched single-epoch detections. These two characteristics of forced-photometry lightcurves are very important for studies that examine variable sources near the single-epoch detection threshold. In Section 2.5.1.1, we show that the flux aggregates provide better *color estimates* of point sources than does matching measurements on the co-adds. For color measurements, we

used flux aggregates, or "catalog co-adds," for three reasons. First, the PSF flux is only as good as the PSF estimate, and PSFs on co-adds are difficult to estimate, as mentioned above. Second, flux aggregates have higher SNRs than fluxes measured from the image co-adds, because images with poor seeing and backgrounds (which degrade co-add quality) were excluded from the co-adds. Therefore, the catalog co-add contains more epochs than the image co-add. Third, colors derived from co-add measurements would need to be matched between the bands, and matching process would add additional uncertainty in the colors.

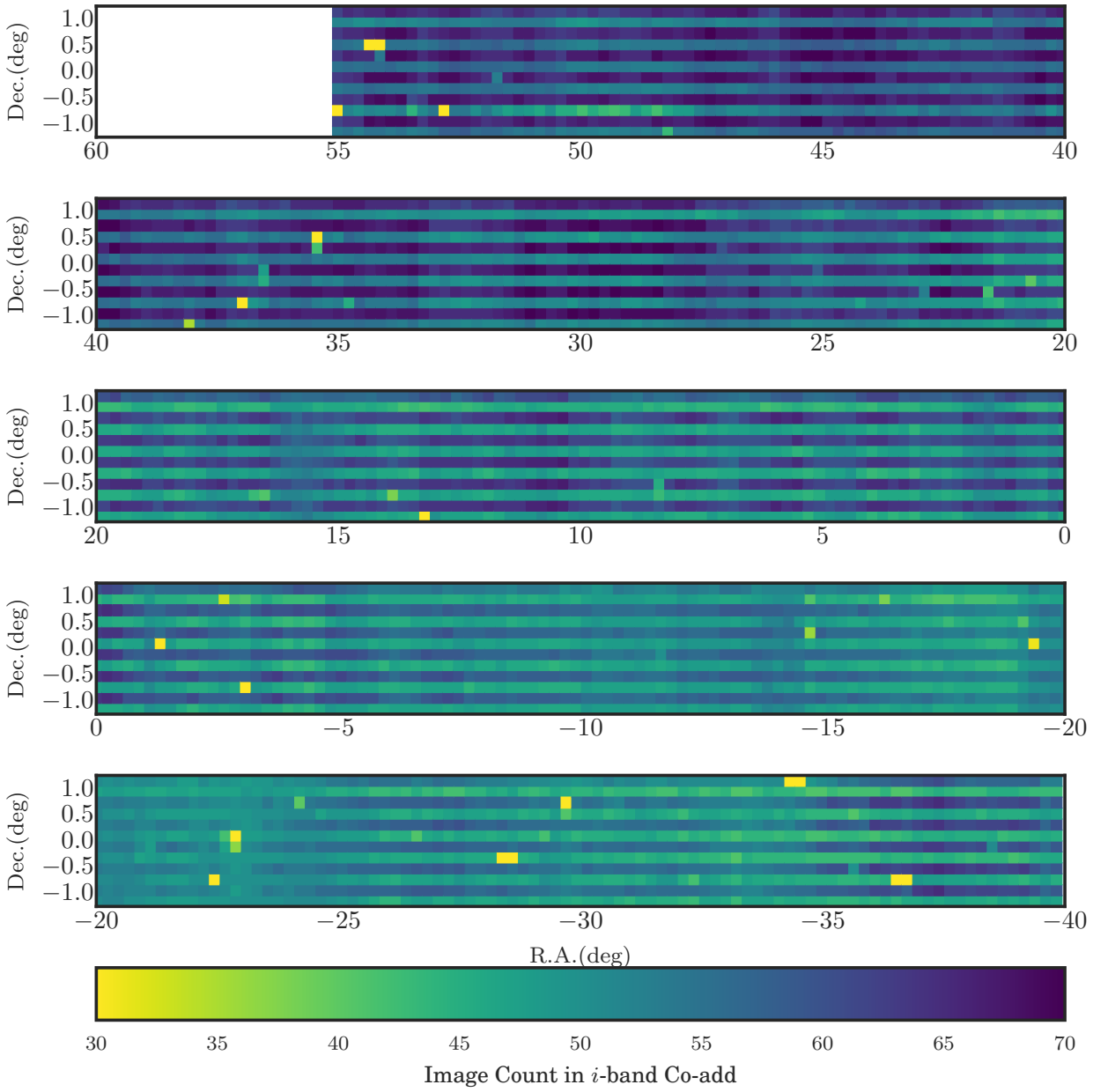

Figure 2.5 Number of epochs contributing the i-band co-add by patch. On average, 55 epochs contributed to the the co-adds, with as many as 71 and as few as 0. Fewer epochs were contributed to the co-adds than were observed by the telescope because of filters on seeing, image, and background-matching quality. The telescope scanned the North scan lines and later R.A. more frequently. The result is that the South patches in the tract 0 (R.A. $>0°$) have comparable depth to the North patches in tract 3 (R.A. $< 0°$)

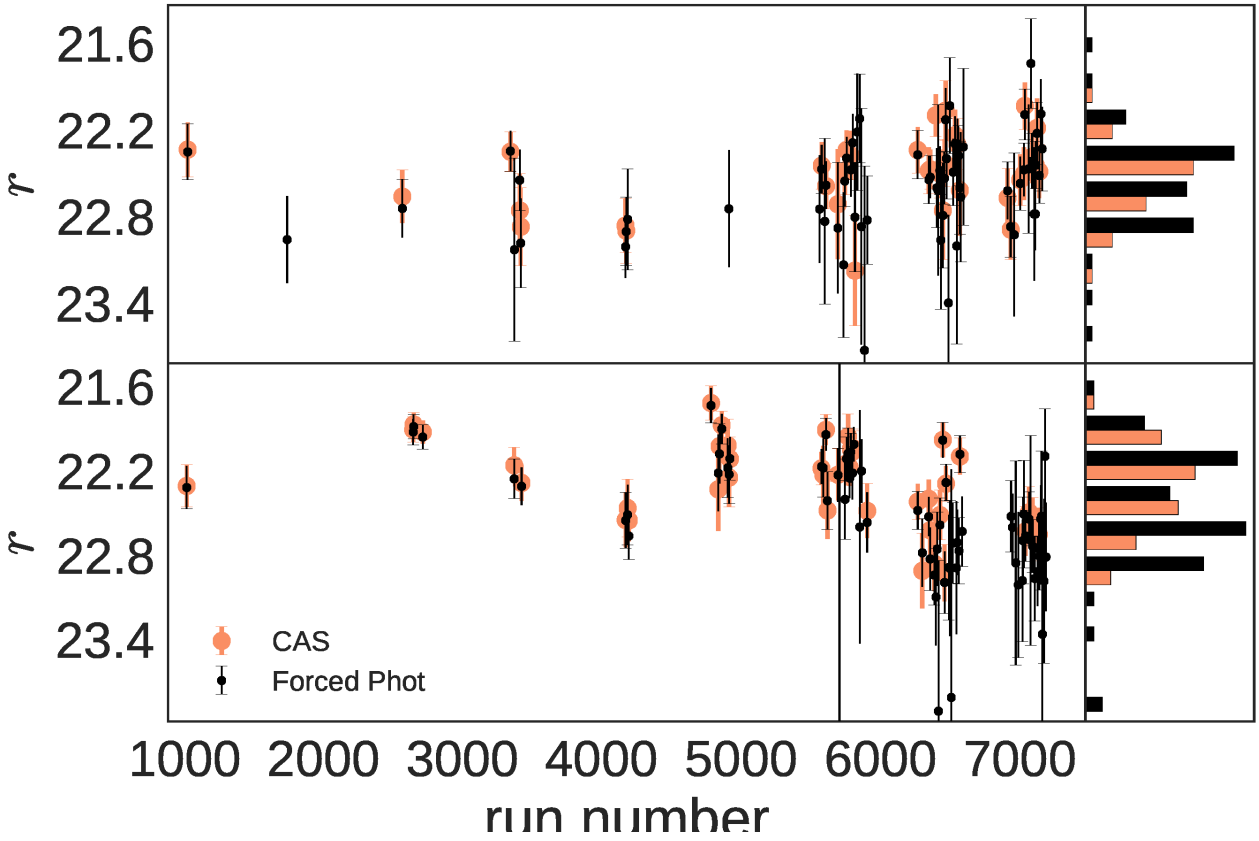

Figure 2.6 Lightcurves for two $z \sim 4$ quasars near the single-epoch detection limit as measured from forced photometry (black) and downloaded from skyserver.sdss3.org/CasJobs/ (orange). At this average magnitude, there are 50% more epochs available in the forced-photometry data set.

2.5 Results: Characterizing the Data Products

We validate the photometry by comparing it to external catalogs where deeper surveys exist and by checking the internal consistency. Externally, we compare the photometry to two deeper surveys' catalogs, 8 sq. arcmin. from HST and 0.7 $\deg^2$ from CFHT, to analyze star-galaxy separation and completeness. For most of the 238 $\deg^2$ area, no deeper catalog exists. Internally, we assess the photometric quality of the algorithms independently from other surveys by examining the measured colors of stars and comparing repeated observations.

2.5.1 Internal Validation

2.5.1.1 Stellar Locus

Most bright point sources are main sequence stars.[12] Stars populate a well-defined locus in optical color-color space nearly independent of magnitude. Therefore, color-color diagrams are useful as an independent assessment quality of the photometry.

Figure 2.7 shows the color-color diagrams of point sources for sources in the range $10 <$ `R.A.` < 55 for bins in r-band magnitude. We defined point sources as sources with `extendedness` ≤ 0.1; see Section 2.5.2.3 for more detail on this threshold. We defined colors using the median forced PSF-magnitude, though other flux estimators perform well (see below). The scatter in the stellar locus increases for fainter bins because of photometric noise and increasing contamination from compact galaxies.

The scatter in the stellar locus can be better quantified by transforming the colors to the principal stellar colors following the method of Ivezić et al. (2004). These principal colors include w and x (defined in the $r-i$ vs. $g-r$ diagram), s (defined in the $g-r$ vs. $u-g$ diagram), and y (defined in the $i-z$ vs. $r-i$ diagram). We examine the scatter, specifically the root mean square (RMS), of the stellar locus distributions for principal colors as a function of r magnitude. The distribution of points around the principal colors s, w, x, y

[12] Galaxies are more numerous at $r > 19$ and appear compact at fainter magnitudes

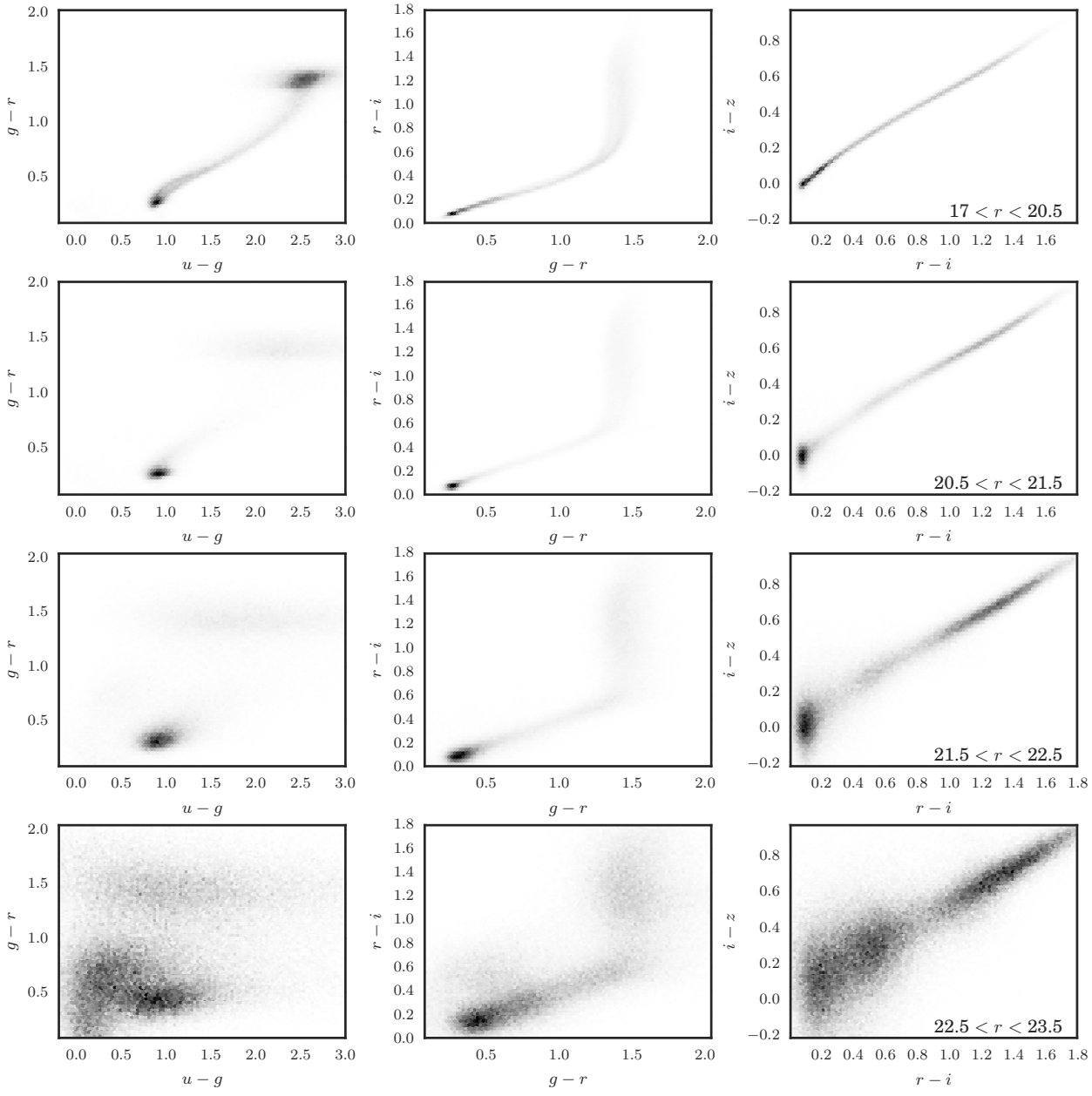

Figure 2.7 Color-color diagrams of stars (`extendedness` < 0.1) from $17 \leq r \leq 23.5$. Colors are the median forced PSF-magnitude. For $r > 21.5$ the scatter increases due to photometric noise and the increasing number density of compact galaxies.

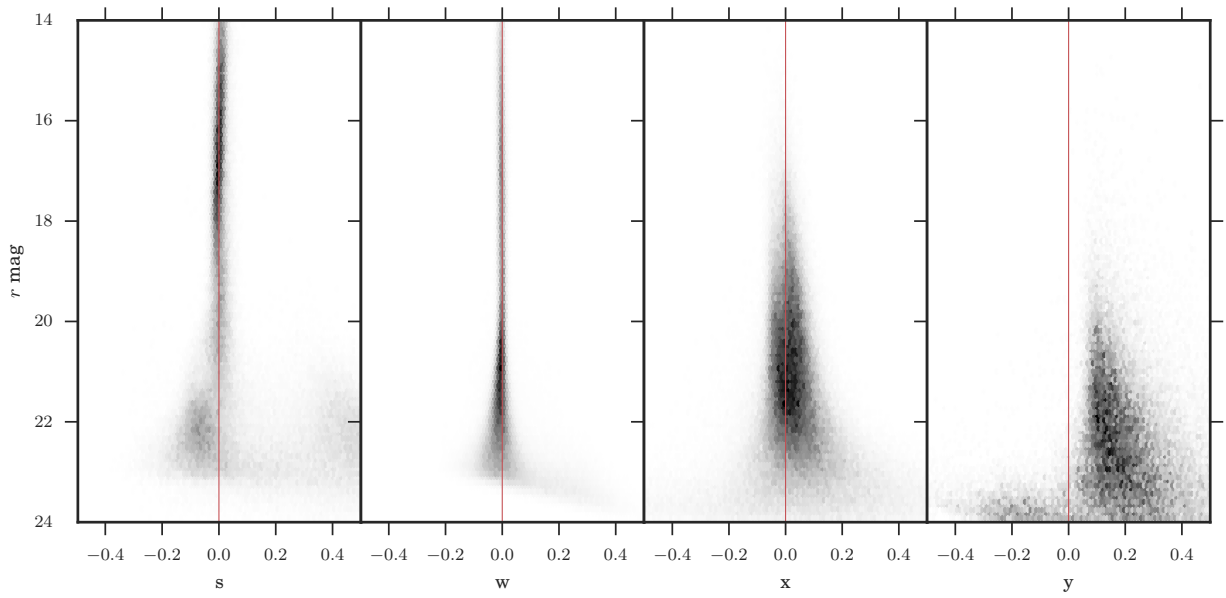

Figure 2.8 The width of the principal colors s, w, x, y (left to right) as a function of r-band magnitude. Colors are the median forced PSF-magnitude for each point source with `extendedness` < 0.1. Colors s, w, and x are unbiased, confirming the calibration of the u, g, r, and i-bands. Conversely, the z-band forced photometry was biased faint, pulling the distribution off the expected locus.

is shown in figure 2.8. For $r < 21.5$ the width of the stellar locus is dominated by the intrinsic scatter in stellar colors.

Figure 2.9 shows the RMS of the principal colors w and x defined in the $g - r$ vs. $r - i$ color diagram for a variety of estimators as a function of r-magnitude. The RMS results from a combination of the intrinsic width of the stellar locus and the photometric errors. The RMS approximates the intrinsic width for bright bins and increases for fainter magnitudes as the photometry errors become more pronounced. A lower RMS corresponds to a tighter stellar locus, which is indicative of more precise photometry.

As expected, the robust estimators, e.g. the median, sigma-clipped weighted mean and

weighted mean, in general perform better—they produce a lower RMS in the principal colors—than the non-robust mean Robustness is important because many of the observations were taken in non-photometric conditions producing extreme flux outliers. Of the robust estimators, the median performed better in w than the weighted mean at bright magnitudes only, $r < 20$. Furthermore, the sigma-clipped weighted mean performed better than the non-clipped mean. At fainter magnitudes ($20 < r < 23$), the difference between sigma-clipped and non-clipped weighted means were negligible and they both had $0 - -15\%$ lower RMS than the medians. This behavior indicates that outliers are more prominent for brighter sources. That the sigma-clipped weighted means performed better than the weighted means for sources $i < 17$ confirms that the errors for these very bright stars are sometimes underestimated. Therefore, when working with this catalog we recommend removing all $i < 17$ sources, and their children from the deblender algorithm.

2.5.1.2 Repeatability

Do measurement errors reflect the scatter in repeated observations? In the Stripe 82 survey, each astronomical object is measured on average 80 times. These repeated measurements provide an opportunity to check if the measurement errors reported on the PSF forced photometry errors accurately reflect the scatter observed. If a source is intrinsically variable, the observed scatter in flux will be greater than the average measurement errors. But if a source is static (as most are on the cadence of Stripe 82), the measurement errors should approximate the observed scatter. We are essentially modeling the flux of the source as a steady value, with measurement errors pulled from a Gaussian distribution.[13]

For each object, we computed the standard deviation of the forced photometry measurements, the robust standard deviation ($0.742 \times$ IQR), and the mean of the reported measurement errors on the individual forced photometry measurements. The measurement errors reported by the forced photometry algorithm are determined by integrating the pixels

[13]Since the measurement errors are heteroscedastic a better way of looking at this is checking that $\chi^2_\nu \sim 1$. We do this section over with χ^2.

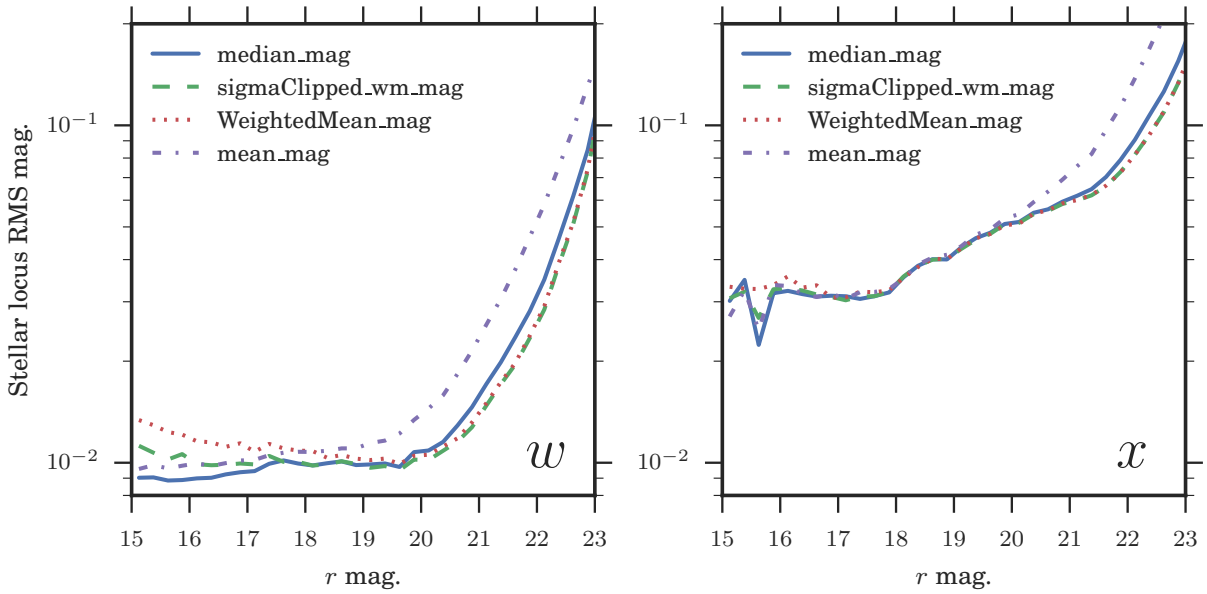

Figure 2.9 RMS of the stellar locus for 5 color estimators: flux measured on the image co-adds and 4 flux aggregates, or ways to compute a catalog co-add of the individual forced photometry measurements. We plot the RMS of the principal components w (left) and x (right), as defined in the $g-r$ vs. $r-i$ color diagram per Ivezić et al. (2007), as a function of magnitude. As expected, the robust estimators (median, sigma-clipped weighted mean and weighted mean) perform better than the non-robust mean. The median performs better at bright magnitudes, and the weighted means perform better at fainter magnitudes.

in the variance plane. We then binned these objects by their r-band co-add magnitude; we show the means of these 3 quantities in figure 2.10. The mean of the measurement errors closely tracks the robust RMS of flux measurements. The non-robust RMS is greater than the measurement errors because of outlier measurements. For $r > 24.5$ the errors appear to be over-estimated.

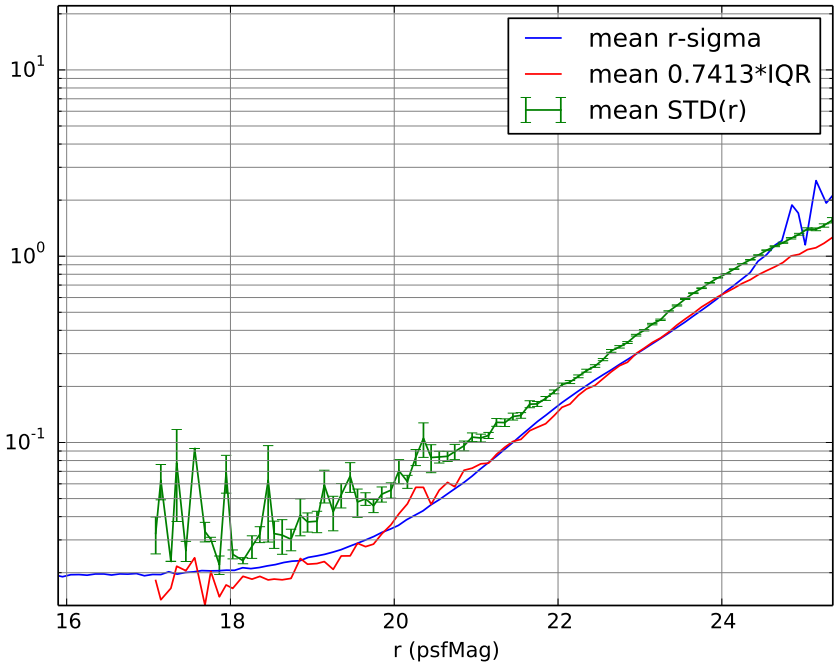

Figure 2.10 The mean of the reported measurement uncertainties (red line) tracks the observed scatter—as measured but the standard deviation—of repeated single-epoch observations (blue line) down to $r \sim 24$. Fainter than $r \sim 24$, the measurement uncertainties are higher than the observed scatter. Units are in magnitudes.

2.5.1.3 Depth

Depth can be measured in terms of both detection completeness—the magnitude at which 50% of known sources are detected— and measurement SNR—the magnitude at which the RMS in the stellar locus passes some threshold or the SNR flux/flux_err drop below some threshold.

2.5.2 External Validation

At the time of writing, two deeper optical surveys intersect Stripe 82. The Cluster Lensing and Supernova survey with Hubble Project (CLASHES; Postman et al. (2012)) imaged the galaxy cluster RXJ2129+0005 with the Hubble Space Telescope (HST) Advanced Camera for Surveys (ACS).[14] The DEEP2 project (Newman et al. 2013) imaged fields with the Canada France Hawaii Telescope (CFHT).[15] We compare our $i-$band `DeepSource` catalog with fields 3.1 and 3.2.

2.5.2.1 Completeness: DEEP2

We matched the i-band `DeepSource` catalog to the DEEP2 DR4 photometric catalog `pcat.31.fits` as a "truth" sample. The truth sample was limited to `BAGFLAG` = 0 or 86 and $351.4 <$ R.A. < 352.0 and $-0.05 <$ Dec. < 0.35 to ensure consistent coverage when comparing the LSST reprocessed Stripe 82 results. LSST sources with the following flags were excluded: `flags_pixel_edge` , `flags_badcentroid`, `flags_pixel_edge`, `flags_pixel_saturated_any`, `flux_psf_flags`, `correctfluxes_apcorr_flag`.

Figure 2.11 shows the completeness/contamination of the co-add detections, tested using `pcat.31.fits` from DEEP2 as the reference catalog. The SDSS and CFHT filter sets are similar but not identical, and we report completeness estimates in CFHT I-mag. The solid blue line shows the completeness for fiducial deblender settings that allow a maximum of 40 peaks (`maxPeaks`=40). The DeepSource catalog reaches a 50% completeness at $I = 24.1$. The yellow line (`maxPeaks`=1) shows without the deblender the completeness is 50–90% even for bright source $I < 22$. Statistically, maxPeaks parameters of 20, 30, 40 do not differ much at $i < 25$.

The DeepSource catalog was also examined for contamination, which is defined as the fraction of sources that were detected in the DeepSource catalog but not in the DEEP2

[14] `http://archive.stsci.edu/prepds/clash/`

[15] `http://deep.ps.uci.edu/DR4/home.html`

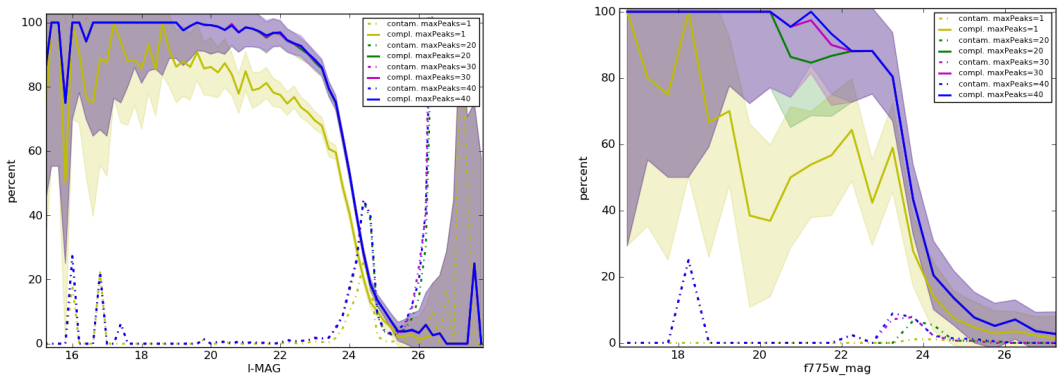

Figure 2.11 Completeness and contamination of the i-band DeepSource detection catalog, using both the DEEP2 catalog `pcat.31.fits` (LEFT) and HST catalog `archive.stsci.edu/pub/hlsp/clash/catalogs/hlsp_clash_hst_acs-ir_rxj2129_cat.txt`(RIGHT) as reference catalogs. The DeepSource detection catalog reaches a 50% completeness limit at $I = 24.1$ and $F775W = 23.8$. Contamination is $< 5\%$ at the limiting depth of the survey. .

catalog. The dashed dark blue line shows the contamination for the fiducial deblender setting (`maxPeaks`=40). The first peak of contamination, at I-mag=24.2, corresponds to a SNR of 3, which is the detection threshold. The second fainter peak at 27 mag is the result of deblended children. The contamination looks higher at $I = 27$ only because the the DEEP2 catalog is not complete beyond $i = 25$. In the end, the contamination relative to the limiting magnitude was found to be less than 5% at the limiting depth of DEEP2.

2.5.2.2 Completeness: HST ACS CLASH cluster RXJ2129+0005

Figure 2.11 (right panel) shows the completeness and contamination using the overlapping HST catalog, `archive.stsci.edu/pub/hlsp/clash/catalogs/hlsp_clash_hst_acs-ir_rxj2129_cat.tx` We considered only sources from the rxj2129 catalog with `flag5sig == 0`.

.

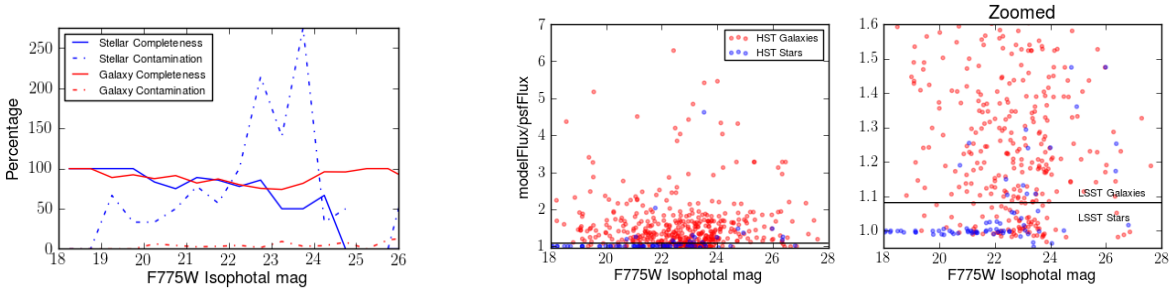

Figure 2.12 (left) Completeness/Contamination of stars and galaxies as compared to HST. (right) HST stars and galaxies (Blue/Red) according to LSST (ratio of model-flux to psf-flux). A SExtractor "stellarity" cutoff of 0.5 was used to define the HST Star-Galaxy separation.

2.5.2.3 Star-Galaxy Separation

The HST ACS imaging of the CLASH cluster RX2919 also is excellent for validating the morphological star/galaxy separation in the DeepSource detection catalog. The PSF has a FWHM of 0.1″, an order of magnitude smaller than ground-based imaging. Recall that we resuse a simple SDSS metric for morphology called `extendedness`, defined as (`PsfMag` - `modelMag`). As sources become more compact, `extendedness` approaches zero. The photometric catalog released for the CLASH cluster was produced using SExtractor (Bertin & Arnouts 1996), which offers a "stellarity" metric. As the name suggests, compact sources have a stellarity close to 1 and extended sources have a stellarity close to 0. Both extendedness and stellarity are continuous.

Figure 2.12 compares the star-galaxy separation from the HST imaging as measured by SExtractor and the Stripe82 i-band DeepSource detection catalog. The left plot shows the continuous stellarity vs. $F775W$-mag (both measured on the HST images) color-coded by the discretized extendedness parameter (measured on Stripe 82 i-band coadd). The right plot shows the inverse: the continuous extendedness (measured on the Stripe 82) vs. $F775W$-mag, color-coded by the HST stellarity of matched sources. The two catalogs are in good

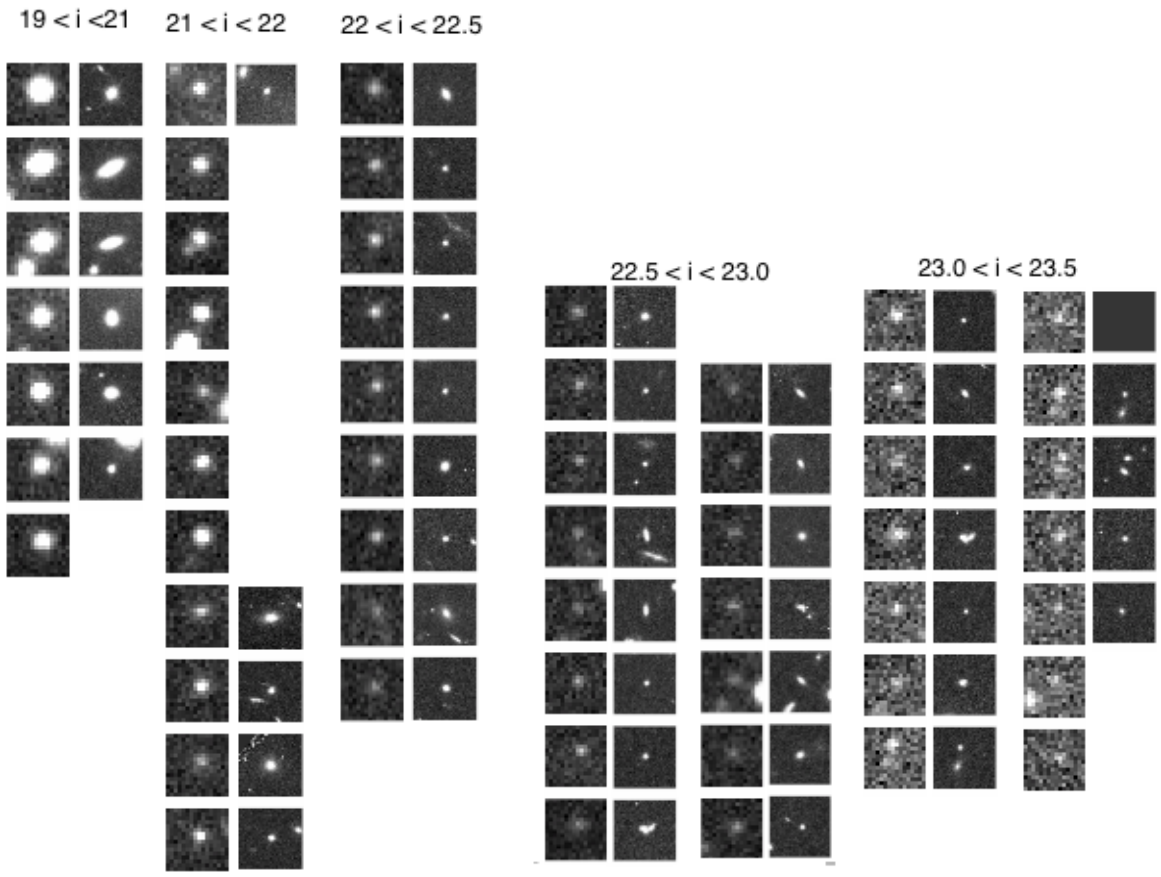

Figure 2.13 Pairs of Stripe 82 (left thumbnail per column) and HST (right thumbnail per column) images of false stars—sources with `extendedness` = 0 and "stellarity"< 0.5—ordered by i-mag. Most are more compact than the ground-based PSF, and thus we lack the resolution/depth to identify these as extended.

agreement for $F775 < 23$; many fainter galaxies, however, appear compact in the co-add.

Thumbnails of these "false stars" are shown, sorted by i-magnitude, in figure 2.13. These sources appear compact in the Stripe 82 co-add but extended in the HST imaging. Each row has two thumbnails side by side. In each row, the left thumbnail is the Stripe 82 co-add, and the right thumbnail is the HST image. Most of these galaxies are simply compact and appear as point sources from the ground. The six sources $21 < i < 22$ with missing HST

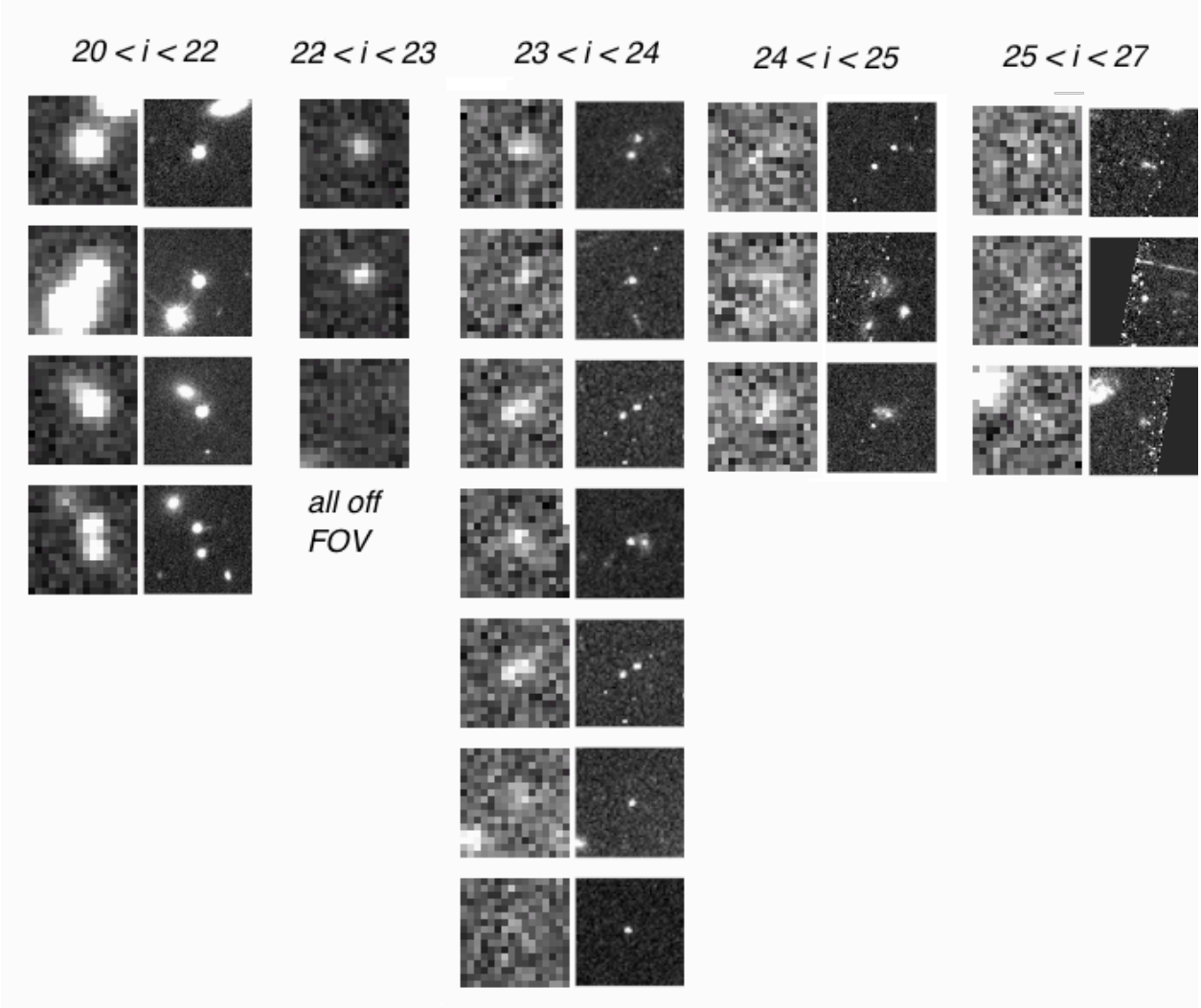

Figure 2.14 Pairs of HST and Stripe 82 thumbnails for false galaxies—sources with extendedness = 1 and "stellarity"> 0.5—ordered by i-mag. Most are clearly multiple sources that blend into an apparent single extended source from the ground.

thumbnails appeared in the CLASH catalog but not the archival HST image inspected here.

Thumbnails of these "false galaxies " are shown in figure 2.14, which is also sorted by i-magnitude. These sources appear extended in the Stripe 82 co-add but compact in the HST imaging. Most are clearly multiple sources that blend into an apparent single extended source from the ground. Of the sources that were not blends, the source of the failure is

uncertain. The two faintest sources in the $23 < i < 24$ column appear isolated. Whereas the sources $20 < i < 22$ were clearly blends, the sources in the top and bottom columns appear separable by the deblending algorithm. The completeness in this magnitude range happens to be the most affected by the deblender `maxPeaks` parameter in this field of view. These are likely the peaks that were not measured because there were more peaks than `maxPeaks` in the large cluster footprint.

We also validated the star/galaxy separation on the whole $238\,\mathrm{deg}^2$ for the brightest sources by cross-matching the catalog with SDSS spectroscopy. Because the SDSS spectroscopic catalogs are biased, comparing the measured extendedness to spectroscopic designation does not offer statistical validation of the extendedness. However, we found the exercise useful because it illustrated the limits of the deblending algorithm used.

We cross-matched the deepSource catalog with the SDSS "spAll"[16] file, yielding matches with spectroscopically confirmed QSOS (N=147), stars (N=231), and galaxies (N=71). In general (but not always), QSOs and stars are point sources. Fifty-four stars or QSOs appear extended in the catalog, having modelFlux/psfFlux > 1.08—equivalent to extendedness > 0.1. While no conclusions can be drawn from the statistics, a closer look at these false galaxies reveals that many are children of bright ($i < 17$) stars, shown in figure 2.15. The figure displays four columns of five ($ugriz$) 40×40 pixel thumbnails. The deblender struggles to assign flux to the children in large, bright footprints. We recommend that parent sources brighter than 17th mag and their children sources be filtered from statistical catalogs.

2.5.2.4 Recovery of Low Surface Brightness in Co-adds

[16] `http://data.sdss3.org/datamodel/files/BOSS_SPECTRO_REDUX/RUN2D/spAll.html`

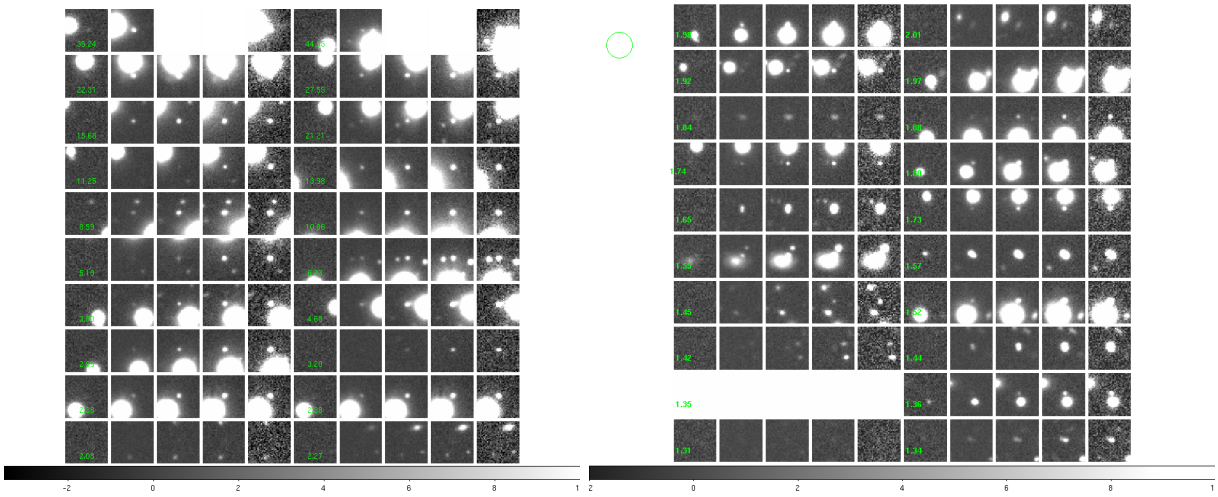

Figure 2.15 40×40 pix thumbnails of stars and QSOs with modelFlux/psfFlux$>$ 1.3. The green label shows the modelFlux/psfFlux, and is placed over the u-co-add. From left to right, images are u-co-add, g-co-add, r-co-add, i-co-add, z-co-add. Note that the none of these were flagged for bad model flux.

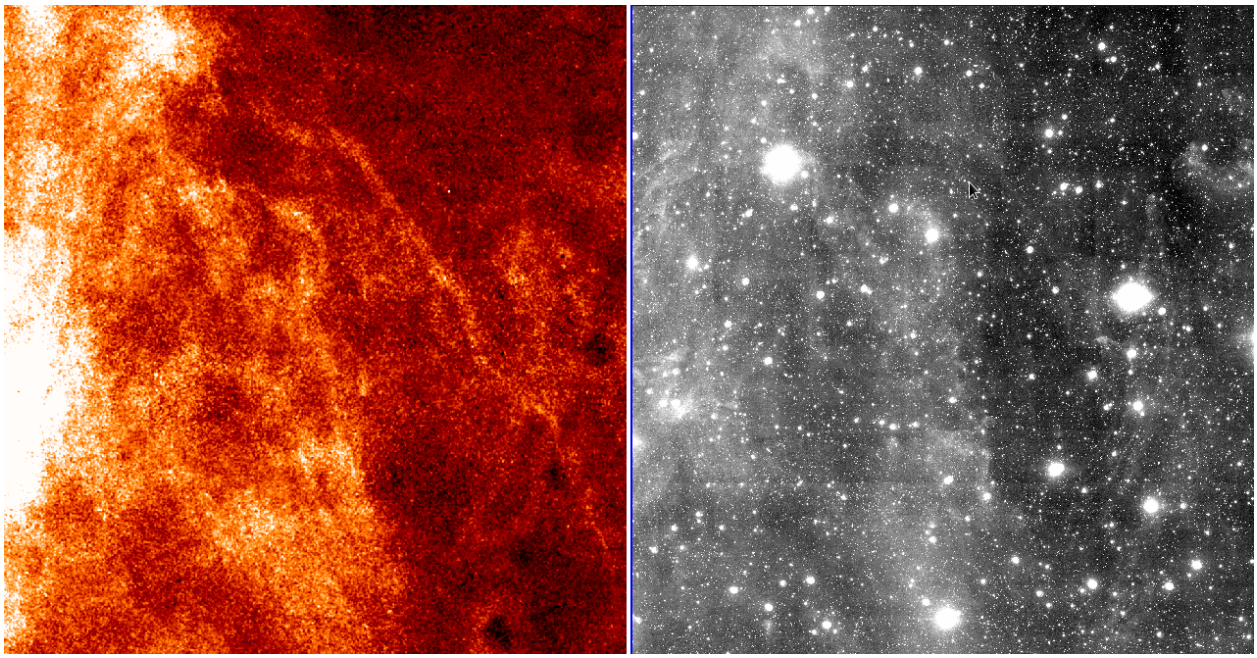

Figure 2.16 Comparison of a co-add of patches 150-160 (right) with the WISE dust maps from Meisner & Finkbeiner (2014) (left). The dust we see reflecting light in the optical (right) is seen emitting light in the IR (left), confirming that the backgrounds in the co-adds are real astrophysical backgrounds. This mosaic of 64 co-add patches is $1.4^\circ \times 1.4^\circ$.

2.6 Conclusion

Starting with 238 $\deg^2$ of repeat imaging, we have extracted the data necessary for efficient quasar selection: lightcurves, colors, and morphologies. We verified the scientific reliability of this data set. At the magnitude limit of this QLF $i = 22.0$ (i=22.5),

- the $(g - r)$ and $(r - i)$ colors have uncertainties less than 2% (4%), as determined by the width of the stellar locus;

- the completeness is greater than 95%(90%) as compared to the DEEP2 CFHT survey;

- the `extendedness` parameter provides sufficiently accurate star-galaxy separation as validated by comparing with HST imaging, with the exception of sources that are children of bright $i < 17$ stars; these can be easily removed from the survey and the affected area can be subtracted from the total survey area.

The resulting catalog of 40 million lightcurves (10 million $i < 23.5$) is the largest catalog of lightcurves produced to date, and it is an ideal data set for finding high-z quasars close to the single-epoch detection threshold. High-z quasars vary on long timescales, $10^{\sim 2}$ days; the Stripe 82 lightcurves have a 10-year baseline. High-z quasars are point sources; PSF photometry is the optimal photometry for point sources. Furthermore, forced photometry aggregates, or "catalog-co-adds," achieve more precise photometry than measuring directly on co-adds. High-z quasars are relatively rare objects; we expect a few hundred $z \sim 4$ quasars on Stripe 82. With this dataset, we can reach sufficient depth to measure the faint end of the $z \sim 4$ QLF, while at the same time surveying the wide area ($238\,\deg^2$) needed to also measure the bright end of the QLF, which is made up of rarer bright quasars; this allows us to constrain its overall shape.

2.7 Supplementary Information

2.7.1 SDSS Runs

Several runs were excluded from the co-adds, based on visual inspection. Table 2.1 lists the blacklisted runs.

2.7.2 Background-Matching Detail

The background-matching algorithm used in co-addition was introduced in section 2.4.1.6. We made several decisions on the low-level components of the algorithm.

Modeling Image Difference Backgrounds: We modeled the difference in background between the science image and the reference image as follows. First we binned the image into 128-pixel bins. We computed the median of all background pixels in each bin. Background pixels were defined as any pixel that was not masked as a detection, cosmic ray, or saturated. These median values were placed at the bin centers, and a fourth order Chebyshev polynomial was fit to these points.

We chose the Chebyshev polynomial instead of an interpolating function because it was important to model the low-frequency temporal changes. Note that the maximum frequency scale can be controlled in both polynomial fits (by choosing the order) and in interpolating spline fits (by choosing the bin size of interpolating points). The bigger the bins, the larger the spatial scale probed. We found empirically that for `coaddTempExposures` of size 2K $\times$2K, a pixel scale of $0.396''$, and bin sizes of 128 pixels that 4th order Chebyshev polynomials were sufficient. Second-order polynomials could not model the rapid gradients backgrounds caused by transparency fluctuations, and 6th-order polynomials started modeling out smaller features.

Why not NN2? We used the algorithm that subtracted each science image from the reference run because we found that the specific problem of co-addition provided simplifying conditions that made the NN2 algorithm (Barris et al. 2005) unnecessary. Recall that for N input images, the the NN2 algorithm fits the additive background offset to all $N(N-1)/2$

Table 2.1. Blacklisted Runs

Run	Filter	Camcols	Fields	Notes
125	ugriz	123456	all	Some fields exhibit a minor WCS offset, and twin run 94 not available.
4905	ugriz	123456	all	Some (but not all) fields exhibit a minor WCS offset
4917	ugriz	123456	39-48	
4933	ugriz	123456	18-28	
6441	ugriz	123456	353-362	
6568	u	4	436, 474, 516, 530, 596, 613, 708, 730	AsTransConversion fails
4263	i	2	110-114	Exception thrown at src/KernelSumVisitor.cc 16
6433	i	2	464	Exception thrown at src/KernelSumVisitor.cc 17
6430	i	2	446-7	Exception thrown at src/KernelSumVisitor.cc 18
7080	i	2	442-3	Exception thrown at src/KernelSumVisitor.cc 19
4263	i	3	132-133	Exception thrown at src/KernelSumVisitor.cc 20
5744	i	3	230-232	Exception thrown at src/KernelSumVisitor.cc 21

image differences.

We tested both algorithms on co-adding SDSS fields. We toggled three different configurations:

- Which algorithm: NN2 vs. reference image only
- Which bin estimator: medians vs. means computed for each bin.
- Which mask: Pair-wise (where only the pixels masked in the reference image and science image are masked) vs. co-add mask (where any pixel masked in any of the input images is masked in the difference being modeled)[17]

Figure 2.17 shows the resulting 2^3 models of the background differences between two example runs, 7136 and 4145, in field 110 imaged in r. The filled circles represent the bin-center estimator of the background pixels in the difference image. Their size represents their statistical weight that was used in the weighted least-squares fit. The background rainbow colored image shows the resulting Chebyshev polynomial model fit to these bin centers. The red and blue squares show the difference between the values estimated at the bin-centers toggling the three configurations. The left panel shows results for the co-add mask, and the right panel the (more appropriate) pair-wise mask.

For the pairwise mask (figure 2.17 right panel) the median and mean estimator produced more similar results for the NN2 algorithm than for the reference-image-only algorithm, which suggests that NN2 can offer more robustness. Medians are less sensitive to outliers than means.

For the co-add mask (figure 2.17 left panel), the NN2 and reference-image-only algorithms produce identical background estimates for the mean and not the median. This is because,

[17]We do not advocate using a co-add mask for co-addition because it masks good pixels, in images which had a temporal problem in another epoch such as cosmic rays, satellite trails, and asteroids. We implemented it, however, to demonstrate its effect on NN2.

unlike the median, the mean is both a linear and commutative operation. To see this in more detail, the NN2 algorithm fits a vector V, the relative background estimates of single bin across N epochs. Following Barris et al. (2005), this vector is fit to the data in the form of a matrix A in which each element A_{ij} is the background estimate of the bin in the image difference between epochs j and i. A is fit to the form $A_{ij} = V_j - V_i$ by minimizing

$$\chi^2 = \sum_{i<j} \frac{(-A_{ij} + V_j - V_i)^2}{E_{ij}^2} + \frac{\sum V_i}{< E >^2} \tag{2.16}$$

where E_{ij} represents the observed error matrix and $< E >$ the average error. If the elements of the data matrix A are measured with linear and commutative operators, there will be an exact solution V, such that $(-A_{ij} + V_j - V_i) = 0$ for all ij.

Thresholds for the background-matching metrics: We used two background-matching metrics to determine whether the background matching was successful: the MSE/Var Ratio and the Residual RMS, discussed in Section 2.4.1.6. We chose constant thresholds across all of Stripe 82. Because the sky backgrounds manifest differently in different bands, we chose thresholds by inspecting the distribution per filter (figure 2.18). See table 2.2 for the threshold values used.

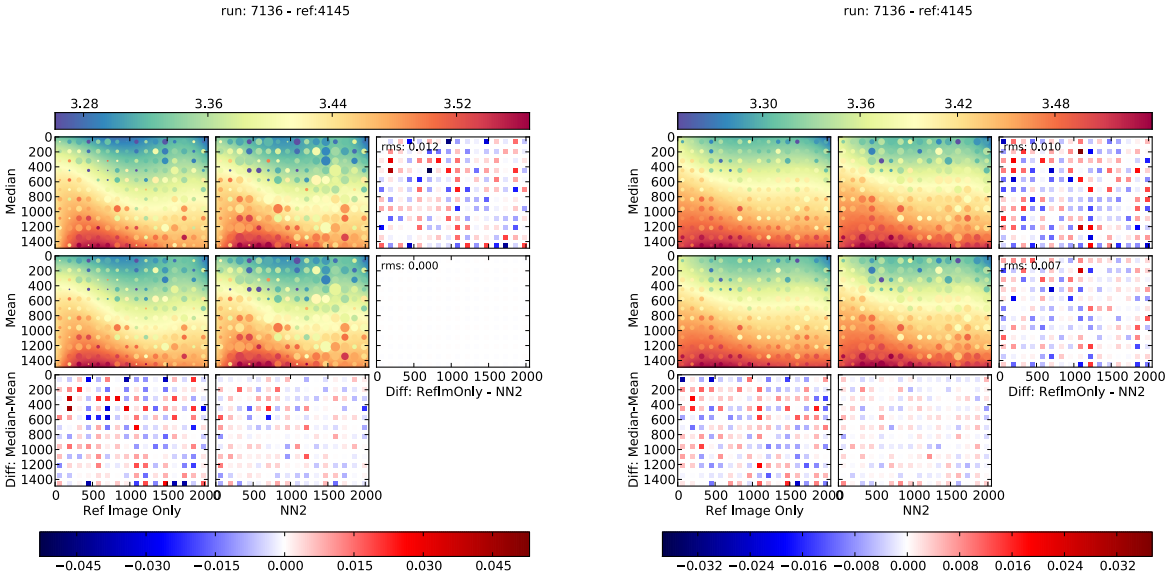

Figure 2.17 Comparison of four different ways to model the background difference between run 7136 and run 4145 in field 110. Two mask types are used are used to mask detections: the **co-add mask (left)** (an `OR` of the masks in every epoch) and the **pair-wise mask (right)** corresponding to just the `OR` of the two image masks. In combination with the co-add mask and the mean (a linear estimator of the binned background pixels), the reference-run only algorithm used in this paper uses the exact same information that the NN2 algorithm would use. Therefore, the two algorithms yield identical models of the background offset needed to match run 7136 to 4145 (RMS=0). However, with the pair-wise mask, using NN2 slightly reduces the difference between using medians and means suggesting that NN2 is the more robust algorithm. However for the median bin estimator the reference run matching algorithm and NN2 produced nearly the same result with a maximum difference in a given bin of ~ 0.05 ADU and and RMS difference of ~ 0.01 ADU.

Table 2.2. Threshold for background-matching success metrics.

Filter	`MSE/Var Ratio` (`maxMatchResidualRatio`)	`Residual RMS` (`maxMatchResidualRMS`)
u	2.5	0.5
g	2	0.5
r	1.2	0.5
i	1.6	0.5
z	2	1

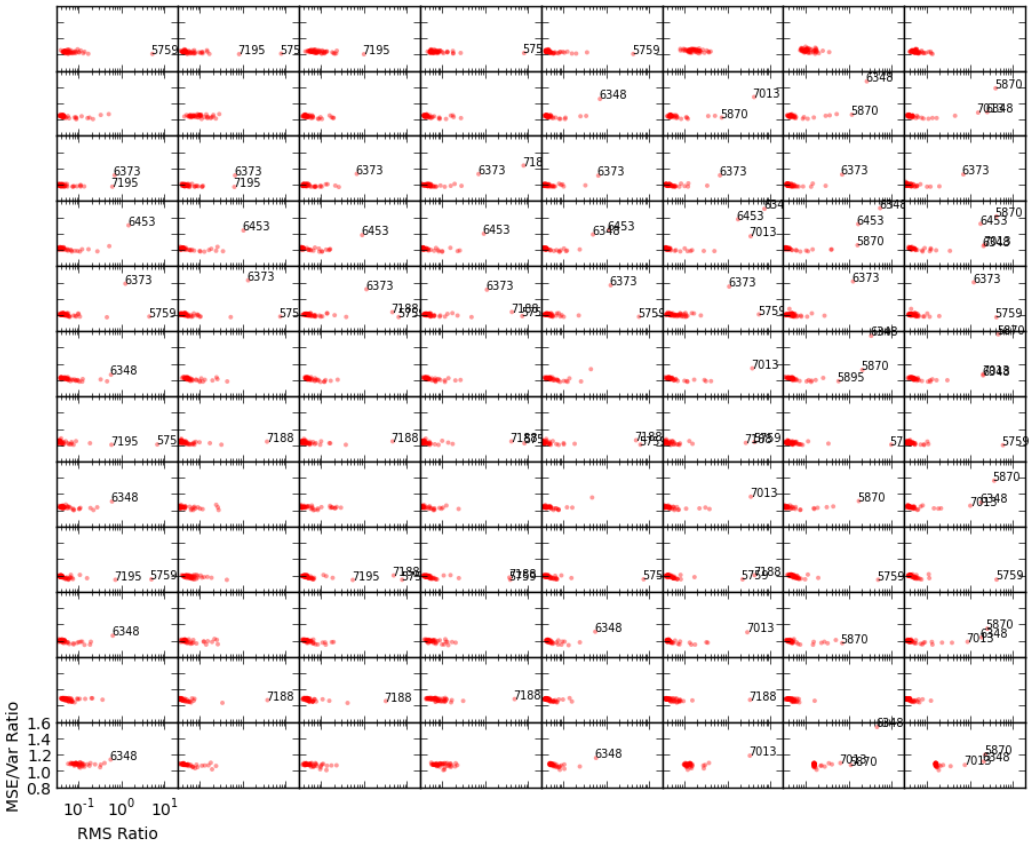

Figure 2.18 MSE/Var Ratio vs. RMS of the residuals of the polynomial fit to the background difference image for 108 patches (each panel) in the region covering DEEP2. Each point is a run that contributed to the co-add. Points that are labeled had a RMS Ratio above the threshold and were left out of the co-add during assembly.

Chapter 3

SPECTROSCOPY: 102 NEW QUASAR SPECTRA $3.4 < Z < 4.6$

3.1 Introduction

To measure the faint-end $z \sim 4$ QLF, we need an efficient procedure to classify $z \sim 4$ quasars using features that can be extracted from the lightcurves described in Chapter 2. But first we need to know what characteristics $z \sim 4$ quasar lightcurves exhibit. We can label lightcurves that have already been identified in previous spectroscopic surveys, and we can then use this labeled data set to train a classifier to identify more quasars.

In addition to helping identify quasars, spectra also provide highly accurate redshifts, and therefore can be used to produce more accurate luminosity estimates than those calculated from photometric data alone. The spectroscopic redshifts have an uncertainty of $\Delta z_{spec} \sim 0.02$, whereas photometric redshifts have an uncertainty of$\Delta z_{photo} \sim 0.15$.[1]

In this chapter, we describe the compilation of a catalog of spectroscopic identifications and redshifts from both previous surveys and new spectroscopy used to train the selection algorithm presented in Chapter 4. We targeted candidates for this new spectroscopy based on a preliminary iteration of the selection algorithm. The final selection algorithm presented in Chapter 4 incorporates the information from these new 102 new spectra $3.4 < z < 4.6$.

3.2 Spectroscopic $z \sim 4$ Quasar Catalog on Stripe 82

To build and cross-validate the classifier, we needed a training sample of spectroscopic redshifts (§ 4.5.1). To build an initial training set, we cross-matched our point source catalog with published quasar catalogs that cover Stripe 82 and included quasars with redshifts

[1]These uncertainties are as measured by the RMS of the residuals in the photometric redshift regression model in Chapter 5.

greater than 3.5. These catalogs included the SDSS I/II/III quasar catalogs from DR7, DR9, and DR12 (Schneider et al. 2010; Pâris et al. 2014); the $z \sim 6$ surveys of Jiang et al. (2006, 2008); the $z \sim 5$ survey of McGreer et al. (2013); the 2dF-SDSS LRG and QSO Survey (Croom et al. 2008, 2SLAQ); and the Stripe 82 variability survey of Palanque-Delabrouille et al. (2011). This resulted in a total of 406 spectroscopically confirmed quasars between $3.5 < z < 4.5$, although not all objects fell within the survey area. See Chapter 4 for details on the contribution of each survey to the training set.

Previous spectroscopic catalogs were composed entirely of objects brighter than $i \sim 22.0$ (corrected for extinction), and 93% of quasars in these catalogs were brighter than $i \sim 21.5$. It is important that the training sample of labeled sources have similar characteristics to the candidates for classification. Because we want to classify quasars that are fainter than those from previous wide-field surveys, we needed to augment our training set with fainter spectra. To extend the training set to include fainter objects, we obtained single-slit spectroscopic identifications of 130 candidates from multiple facilities.

To optimize the efficiency of the spectroscopic observations, we used the following targeting strategy. We used MMT and Gemini to obtain spectra down to $i \sim 22.5$. At MMT, we focused on creating a spectroscopically complete sample down to $i \sim 22$ in a $55\,\mathrm{deg}^2$ subregion. This $55\,\mathrm{deg}^2$ subregion covers the north strips in the range $-40° < \mathrm{R.A.} < +10°$. Because this region covers a wide range of R.A., part of the spectroscopic region was observable at any given time of night. The spectroscopic survey area of six discontiguous scanlines was not ideal for clustering measurements, which benefit from larger contiguous areas. However this strategy uses all six rows of CCDs, reducing potential, though unexpected, systematics such as quantum efficiency, amplifier offsets, and astrometric bias.

We used the APO 3.5m telescope to fill in the brighter, but previously unconfirmed, $z \sim 4$ candidates. We also used some time at Gemini North to fill in some of the later candidates.

This complete set of spectra was taken over the course of two years with evolving selection criteria. All spectra contributed to the dataset, which was used both to train a classifier and empirically model the variability characteristics. However, not all spectra are in the resulting

uniformly selected luminosity-function sample. We present all spectra, whether or not they are used in the luminosity function measurement.

3.2.1 Spectroscopic Observations

3.2.1.1 MMT Observations

We observed 74 $z > 3.4$ quasars and 17 contaminants at the MMT 6.5 m telescope with the Red Channel spectrograph. We used the 270 mm^{-1} grating centered at 7000 Å, providing coverage from 5250–8800 Å. We used either $1''$ or $1.5''$ slits based on the seeing conditions, providing resolutions of $R \sim 640$ and $R \sim 430$, respectively. We also used an L-42 filter to block the higher-order, bluer wavelengths.

We conducted observations on three half-nights (UT: 130906, UT141017, UT141018) and two whole nights (UT141019, UT141020).

3.2.1.2 APO 3.5m Observations

Brighter targets—15 $z > 3.4$ quasars and 14 contaminants—were observed with the Apache Point Observatory (APO) 3.5m with the Dual Imaging Spectrograph (DIS). In the standard low-resolution configuration (R300/B400), we used the 300 mm^{-1} grating centered at 7500 Å, providing coverage from 5250 Å to 9800 Å.

We observed during the fourth quarter of 2013 (2013Q4) and the first (2014Q1) and fourth (2014Q4) quarters of 2014. We had one half-night of director's discretionary time on UT131224. During Q1, we had three half-nights (UT140101, UT140102, UT140103). UT140101 and UT140102 had long exposure times because of poor seeing. During Q4, we observed for fractions of half-nights (UT141003, UT141216, UT141216, UT141220, UT141230).

3.2.1.3 Gemini Observations

Brighter targets—13 $z > 3.4$ quasars and 1 contaminant—were observed on Gemini North on UT131024 and UT131025.

Table 3.1. Summary of Spectroscopic Observations

UT Date	Half[a]	Observatory	Instrument	Seeing (arcsec)	Aperture (arcsec)	Wavelength Range (Å)	Resolution (Å per pixel)
2013 September 6		MMTO	Red Channel	0.5-1.3	1, 1.5	5250–8800	3.53
2013 October 24	Partial	Gemini	GMOS N			5300–9500	1.37
2013 October 25	Partial	Gemini	GMOS N			5300–9500	1.37
2013 December 24	A-Half	APO 3.5m	DIS	1		5200–9800	2.34
2014 January 1	A-Half	APO 3.5m	DIS	1.3		5200–9800	2.34
2014 January 2	A-Half	APO 3.5m	DIS	2.8		5200–9800	2.34
2014 January 3	A-Half	APO 3.5m	DIS	1		5200–9800	2.34
2014 October 3	B-Half	APO 3.5m	DIS	1		5200–9800	2.34
2014 October 17	A-Half	MMTO	Red Channel	0.77	1, 1.5	5250–8800	3.53
2014 October 18	A-Half	MMTO	Red Channel	0.7-1.4	1	5250–8800	3.53
2014 October 19		MMTO	Red Channel	0.7-1.2	1	5250–8800	3.53
2014 October 20		MMTO	Red Channel	1.1-1.6	1.5	5250–8800	3.53
2014 December 16	A-Half	APO 3.5m	DIS	1.2	1.5	5200–9800	2.34
2014 December 20	A-Half	APO 3.5m	DIS	1.5	1.5	5200–9800	2.34
2014 December 30	A-Half	APO 3.5m	DIS	1.5	1.5	5200–9800	2.34

[a]Time on the APO 3.5m is scheduled per half night: A-half, from sundown to midnight, and B-half, from midnight to sunrise. Some nights at MMT were pooled and split with another PI also resulting in half-night observing runs.

3.2.1.4 Spectra Reduction

Spectra were processed using standard longslit reduction techniques through a combination of Pyraf[2] and Python functions. The instrument signature removal process subtracted bias, corrected for flat fields generated from internal lamps, and subtracted sky background using a polynomial fit along the slit direction. Cosmic rays were identified and interpolated over using the LACOS routines (van Dokkum 2001). The wavelength solution was initially calibrated from an internal HeNeAr lamp and then corrected on a per-spectrum basis using night sky lines. Wavelength calibration was corrected using night sky lines in two steps: first, a bulk offset was made using 5577 Å alone. Second, six other night sky lines (including 5577 Å) were fit and compared to known wavelengths, and the residuals were minimized. The error in the wavelength calibration was then given by the standard error of the six lines.

Multiple exposures were co-added. Standard stars were observed between one and three times per night and used for flux calibration without extinction correction; therefore, the absolute flux calibration of the spectra is not reliable. We did not correct for telluric lines when processing the spectra.

3.2.1.5 Redshift Estimation

We estimate the redshifts of the spectra by maximizing the likelihood of the fit to a quasar template spectrum from Vanden Berk et al. (2001) as a function of redshift. The likelihood for n data points, each drawn from a Gaussian distribution around the model $f(x_i, z)$ with a heteroscedastic measurement uncertainty σ_i, is given by

$$-\log L = \frac{n}{2}\log(2\pi) - \sum_i \log\sigma_i - \sum \frac{y_i - f(x_i, z)}{2\sigma_i^2}. \tag{3.1}$$

Maximizing this likelihood is equivalent to minimizing $\chi^2(z)$.

The model $f(x_i, z)$ is a linear scaling of the template T_λ, to fit the measured spectrum S_λ with two parameters: β_0, the additive offset, and β_1, the multiplicative amplitude. $f(x_i, z) =$

[2]Pyraf is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

$\beta_1(z)T_{\lambda+\Delta} - \beta_0(z)$. Two parameters are needed to account for both the overall normalization of the spectrum with respect to the template and the strength of the emission lines with respect to the continua, which vary from quasar to quasar. The continua and broad emission lines emanate from two physically distinct components of a quasar. Therefore,

$$\chi^2(z) = \sum_{\lambda} \frac{(S_\lambda - \beta_1(z)T_{\lambda+\Delta} - \beta_0(z))^2}{\sigma_\lambda^2}. \tag{3.2}$$

For every trial redshift, we computed $\chi^2(z)$, creating a χ^2 curve, and found the minimum of the curve: $\hat{z} = \text{argmin}(\chi^2(z))$. The 68% confidence interval of $\hat{z}$ was bounded by the z's that correspond to one χ^2 above the minimum χ^2 (full χ^2, not reduced), and provided the statistical uncertainty on the measurement, $\sigma_{z,\text{statistical}}$. We took the range of redshifts in the interval where $\chi^2 < 1 + \min(\chi^2)$. The best-fit redshift was taken to be the median, and the σ_z was taken to be the edges of the interval. The best-fit redshifts were visually inspected, because there were only 102.

In addition to $\sigma_{z,\text{statistical}}$, other sources of uncertainty on the redshift include: the redshift of the template spectrum, $\sigma_{z,\text{template}}$; the wavelength calibration computed when processing the spectra, $\sigma_{z,\text{wavelength-calibration}}$; and intrinsic systematic-offsets $\sigma_{z,\text{systematic-offset}}$. The wavelength calibration uncertainties were small (5–75 km s^{-1}), as were the statistical errors on the template fitting (300–1500 km s^{-1}). In fact, the dominant source of uncertainty, $\sigma_{z,\text{systematic-offset}}$, is not statistical at all; it stems from systematic shifts between the emission lines with respect to the host galaxy (1500 km s^{-1}) and Hubble flow ($\delta z = 0.02$ or 6000 km s^{-1}). We assumed that the sources of error on the redshifts were independent. Therefore, the total uncertainty is the sum in quadrature of $\sigma_{z,\text{statistical}}$, $\sigma_{z,\text{template}}$, $\sigma_{z,\text{wavelength-calibration}}$, and $\sigma_{z,\text{systematic-offset}}$.

Seeking a large quantity of spectra identifications of high-z quasars, we set exposure times for sufficient SNR to identify emission lines for a redshift estimate. The spectra of contaminants, which lacked clear emission lines, often did not have sufficient signal to make a definitive identification. In cases where the SNR afforded an identification, we included the identification in our training set. In cases where the SNR was too low to identify the

contaminants, the spectrum was not included in the training set.

To assess our confidence in the spectroscopic identification we used templates for the most common contaminants from the SDSS DR5 spectral cross-correlation template repository:[3]

- `spDR2-009.fit`: G Star 2
- `spDR2-010.fit`: K Star 1
- `spDR2-011.fit`: M1 Star
- `spDR2-017.fit`: Carbon Star 1
- `spDR2-018.fit`: Carbon Star 2
- `spDR2-019.fit`: Carbon Star (Late-type star)
- `spDR2-023.fit`: Early-type galaxy
- `spDR2-024.fit`: Galaxy 1
- `spDR2-025.fit`: Galaxy 2
- `spDR2-026.fit`: Galaxy 3
- `spDR2-027.fit`: Late-type Galaxy
- `spDR2-028.fit`: Luminous Red Galaxy
- `spDR2-029.fit`: QSO
- `spDR2-030.fit`: BAL QSO 1
- `spDR2-031.fit`: BAL QSO 2.

We then measured the χ^2 and cross-correlation for all of these templates in an appropriate range of redshifts. Most contaminants were galaxies and carbon stars.

As can be seen in figures 3.1–3.6, all identified quasars are consistent with a Type I classification based on their broad line widths (FWHM $> 1000\,\mathrm{km\,s^{-1}}$). Of these quasars, a higher proportion than average were broad absorption line quasars (BALQSOs). This is not surprising, because we selected based on g-, r-, and i-band variability (section 4.3). The r-band covers the Lyα line, and the i-band covers the C IV broad emission line. Broad absorption lines are known to be variable (Lundgren et al. 2007; Gibson et al. 2008, 2010). Therefore, selecting based on variability means we have a higher likelihood of selecting objects with variable broad absorption lines.

[3] http://classic.sdss.org/dr5/algorithms/spectemplates

3.2.2 Resulting Spectroscopic Quasar Catalog

Table 3.2 presents the spectroscopic information for all $z > 3.4$ quasars observed, regardless of whether they ended up in the final luminosity-function sample. Because quasars' number density drops off with increasing redshift, there are many more low-z objects than high-z objects. The most frequent contaminants of $z \sim 4$ quasars are $z \sim 3.5$ quasars. Rows are sorted by redshift and include the best-fit redshift, the uncertainties (above), and the luminosities.

Figures 3.1 – 3.6 present the spectra for the objects in this catalog, split by observatory and sorted by redshift. These spectra were smoothed by a 13-pixel Hanning window. The broad telluric line at 7600 was not corrected, which eats the C IV emission line at $z \sim 4$. Spectra were not photometrically calibrated, so flux density should be taken as relative.

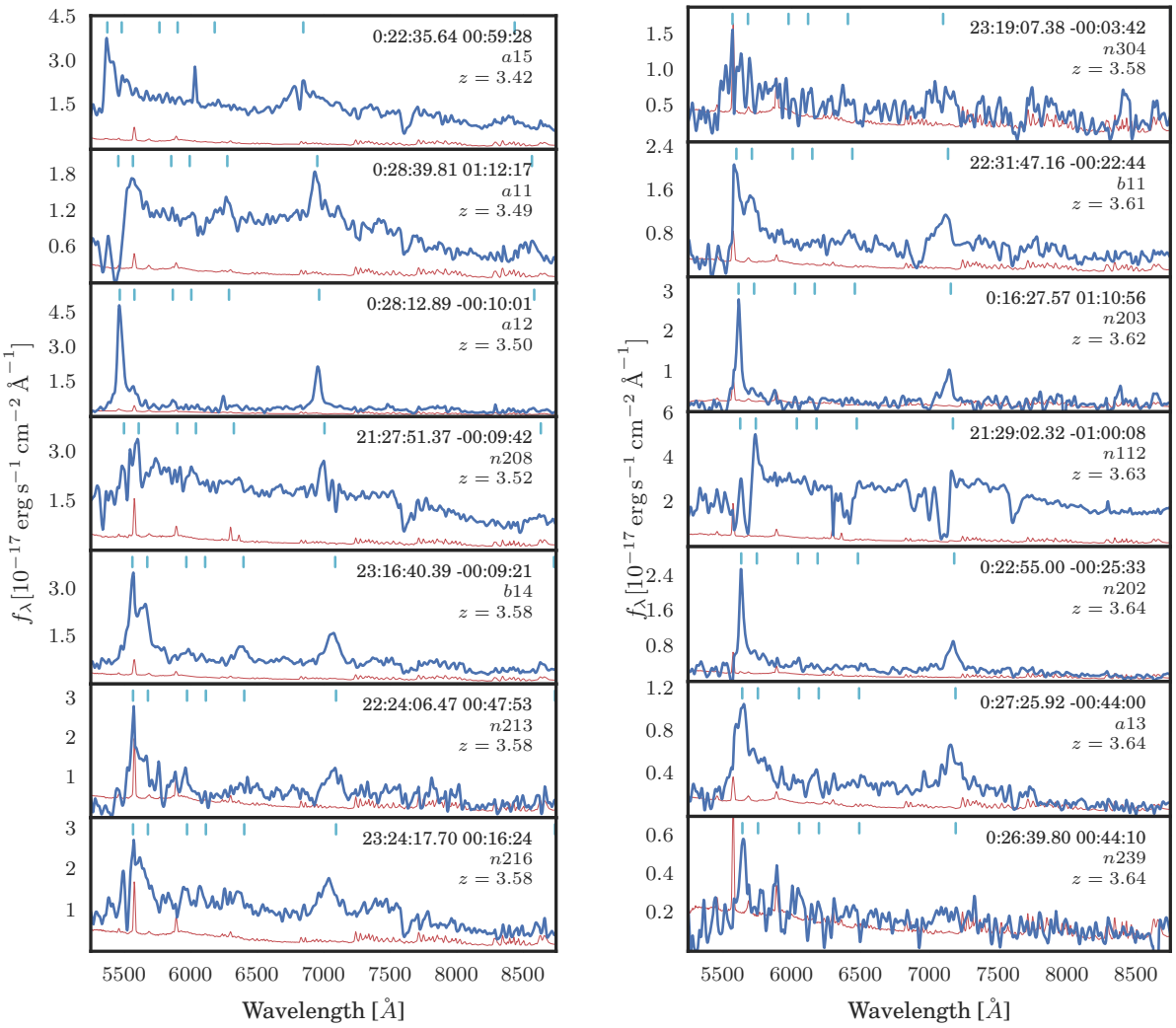

Figure 3.1 Spectra of Stripe 82 quasars observed with MMT Red Channel Spectrograph, smoothed with a 13-pixel Hanning window. They have been flux calibrated from standard star observations; however, the absolute flux calibrations are uncertain. Telluric line at 7600 Å has not been removed. The red line represents the uncertainty, and vertical ticks mark wavelengths of redshifted emissions lines: Lyβ, Lyα, Si IV, N V, O I, C II, Si IV, C IV and C III.

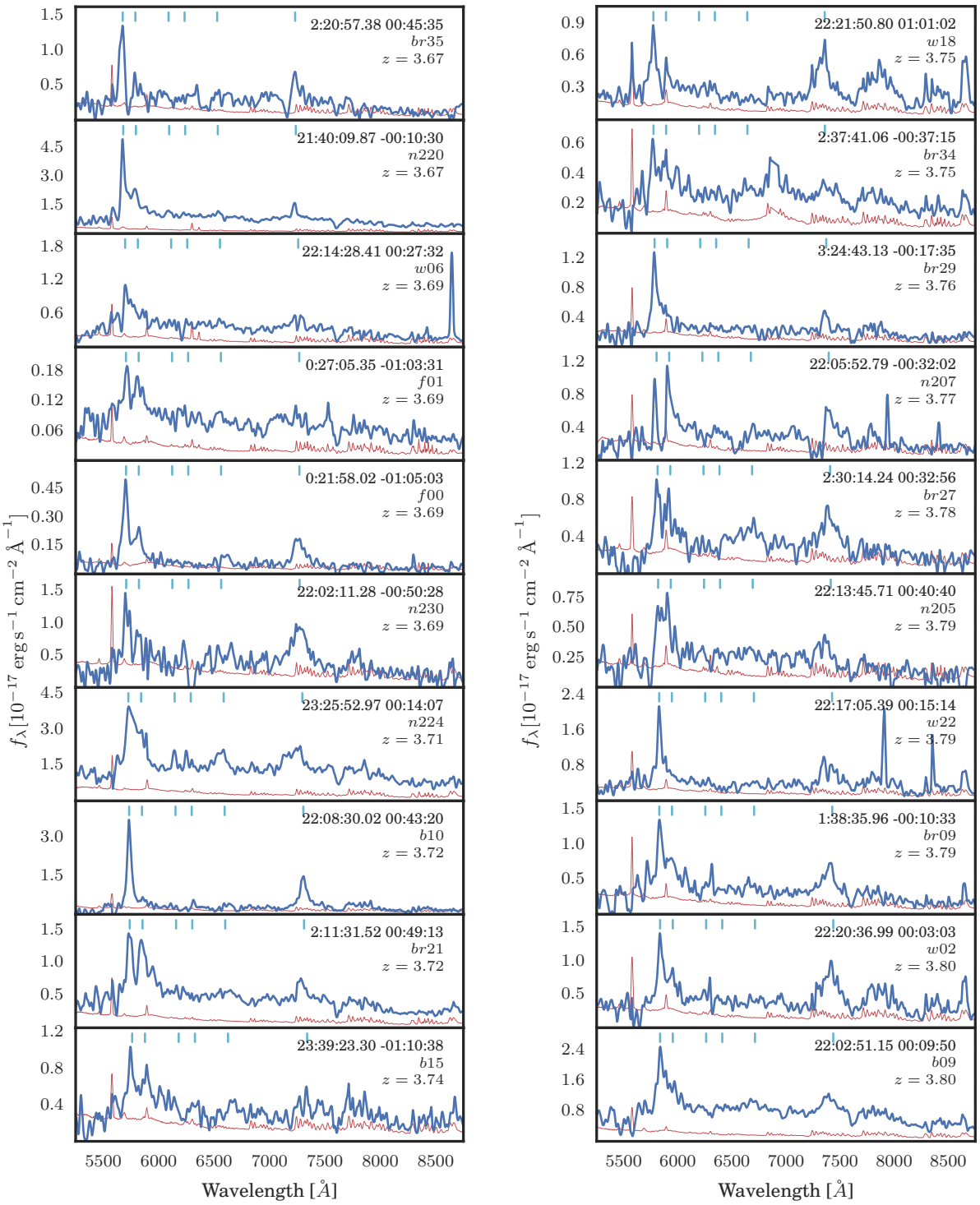

Figure 3.2 Spectra of Stripe 82 quasars observed with MMT Red Channel, continued.

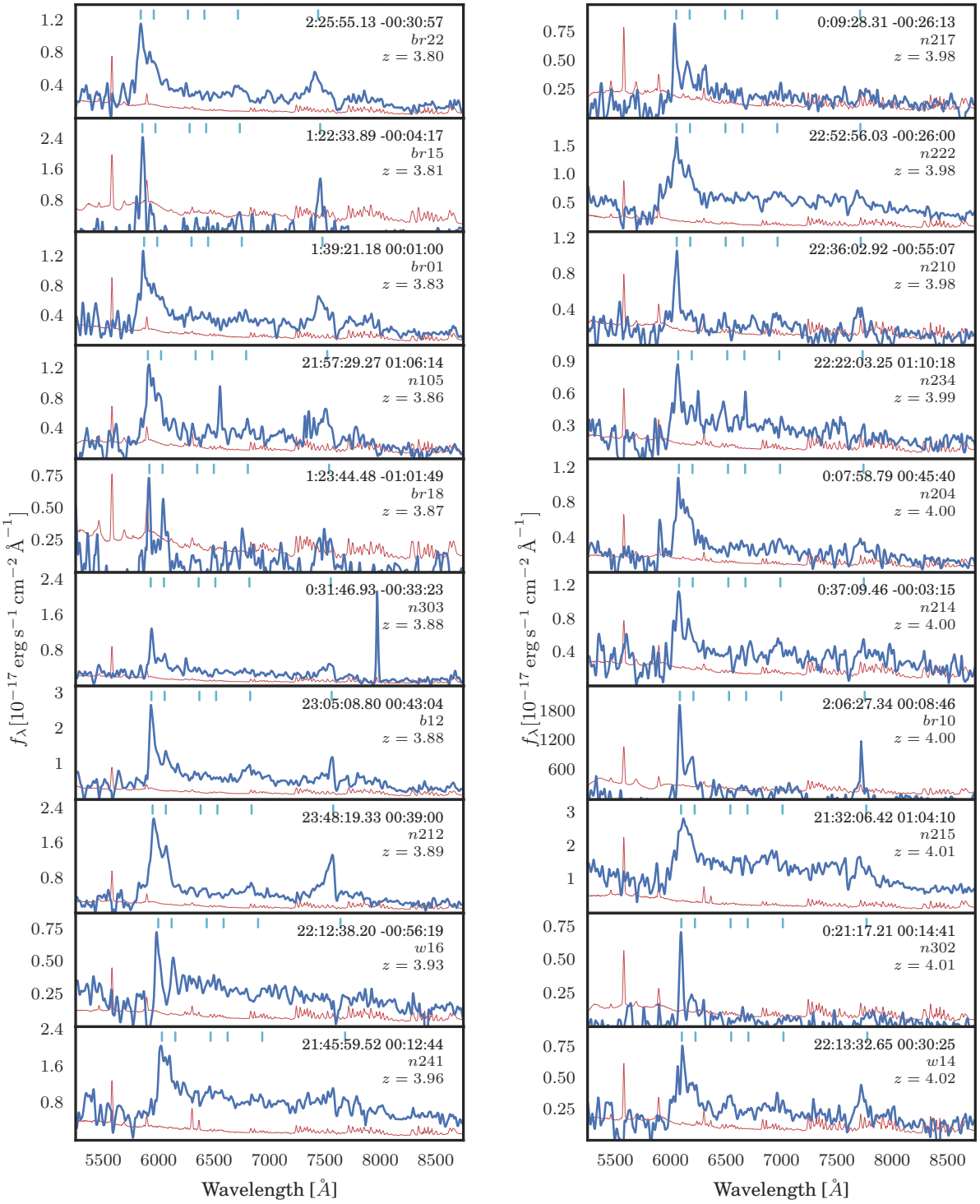

Figure 3.3 Spectra of Stripe 82 quasars observed with MMT Red Channel, continued.

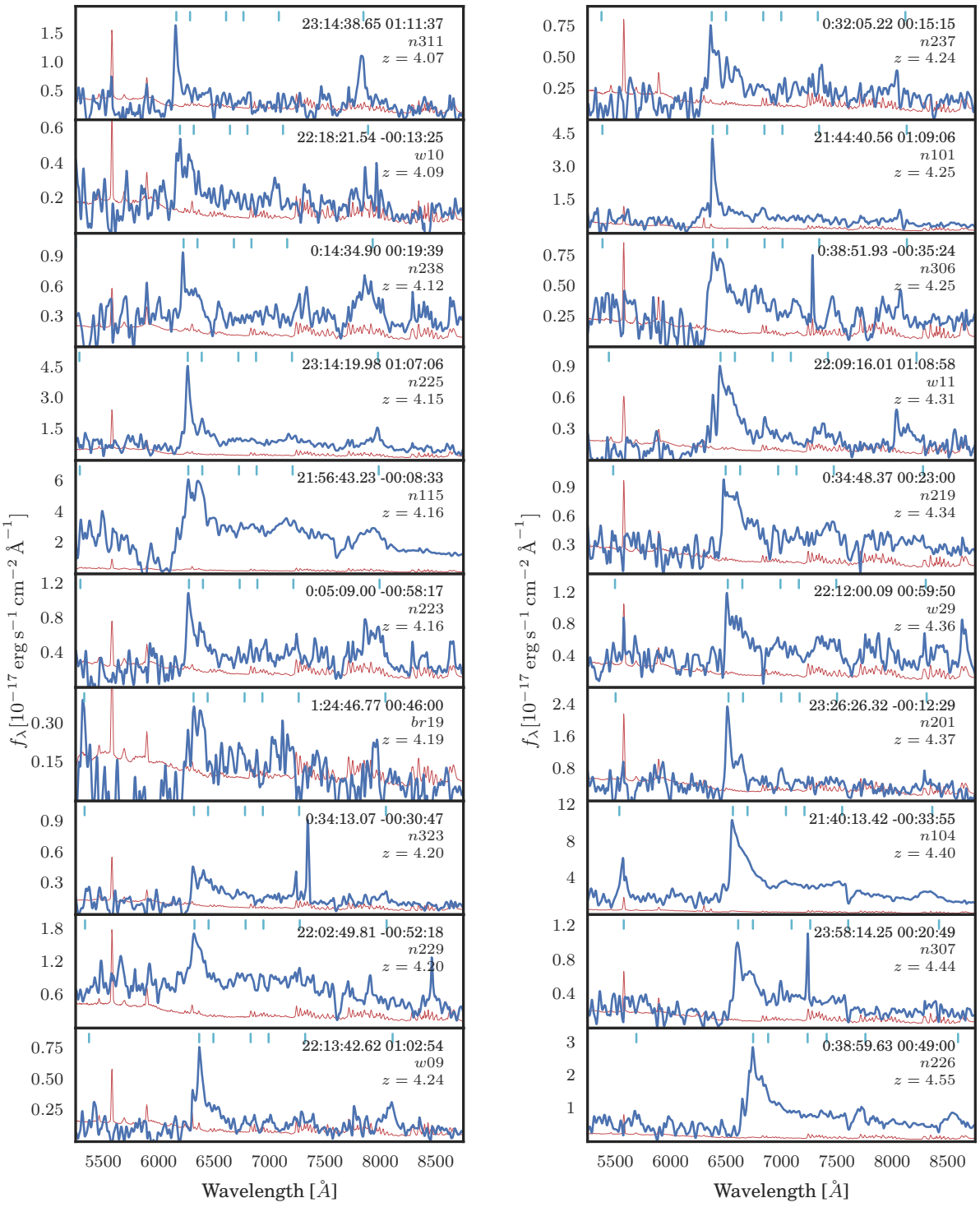

Figure 3.4 Spectra of Stripe 82 quasars observed with MMT Red Channel, continued.

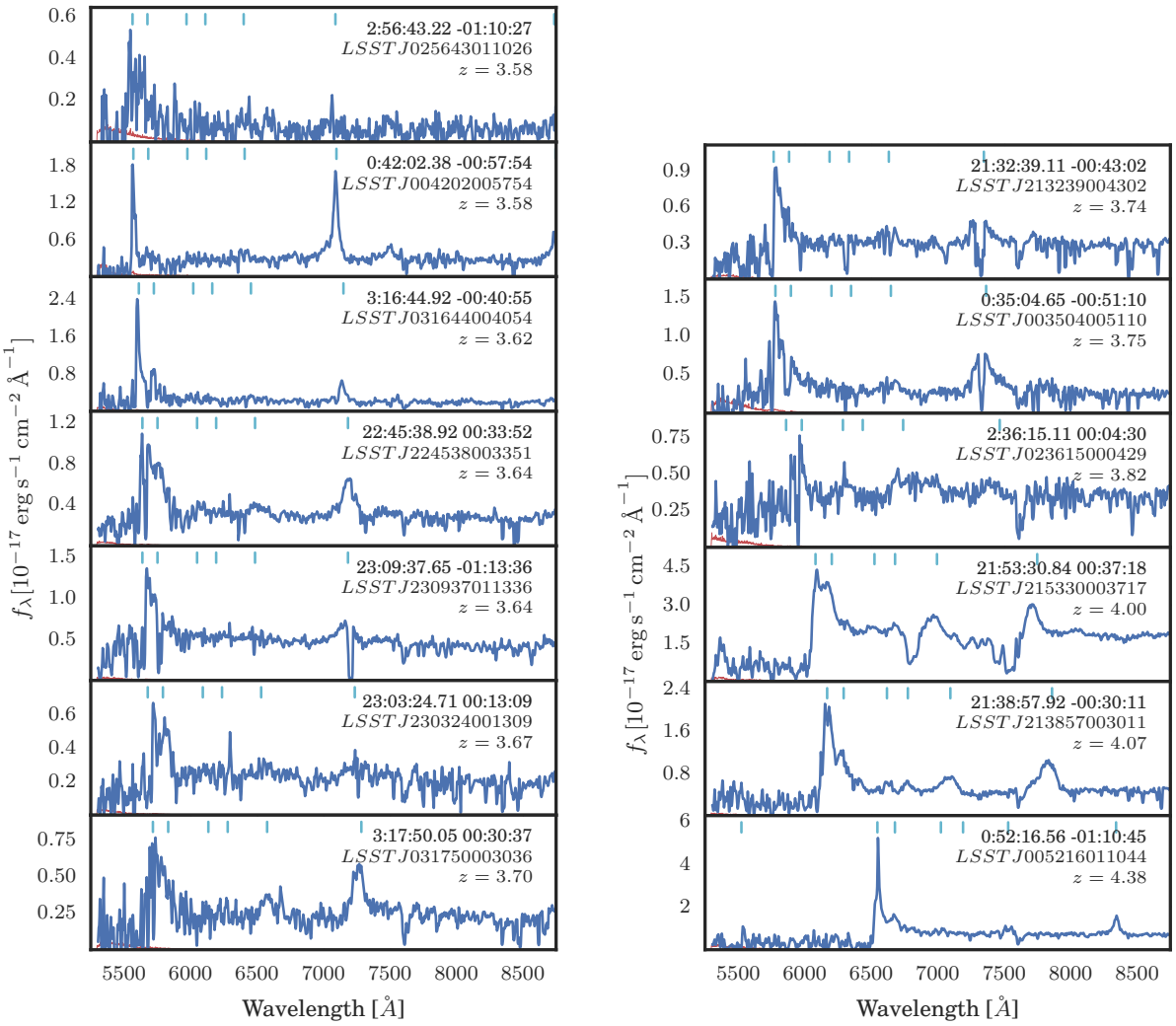

Figure 3.5 Spectra of Stripe 82 quasars observed with Gemini GMOS-N, smoothed with a 13-pixel Hanning window. Markings follow figure 3.1.

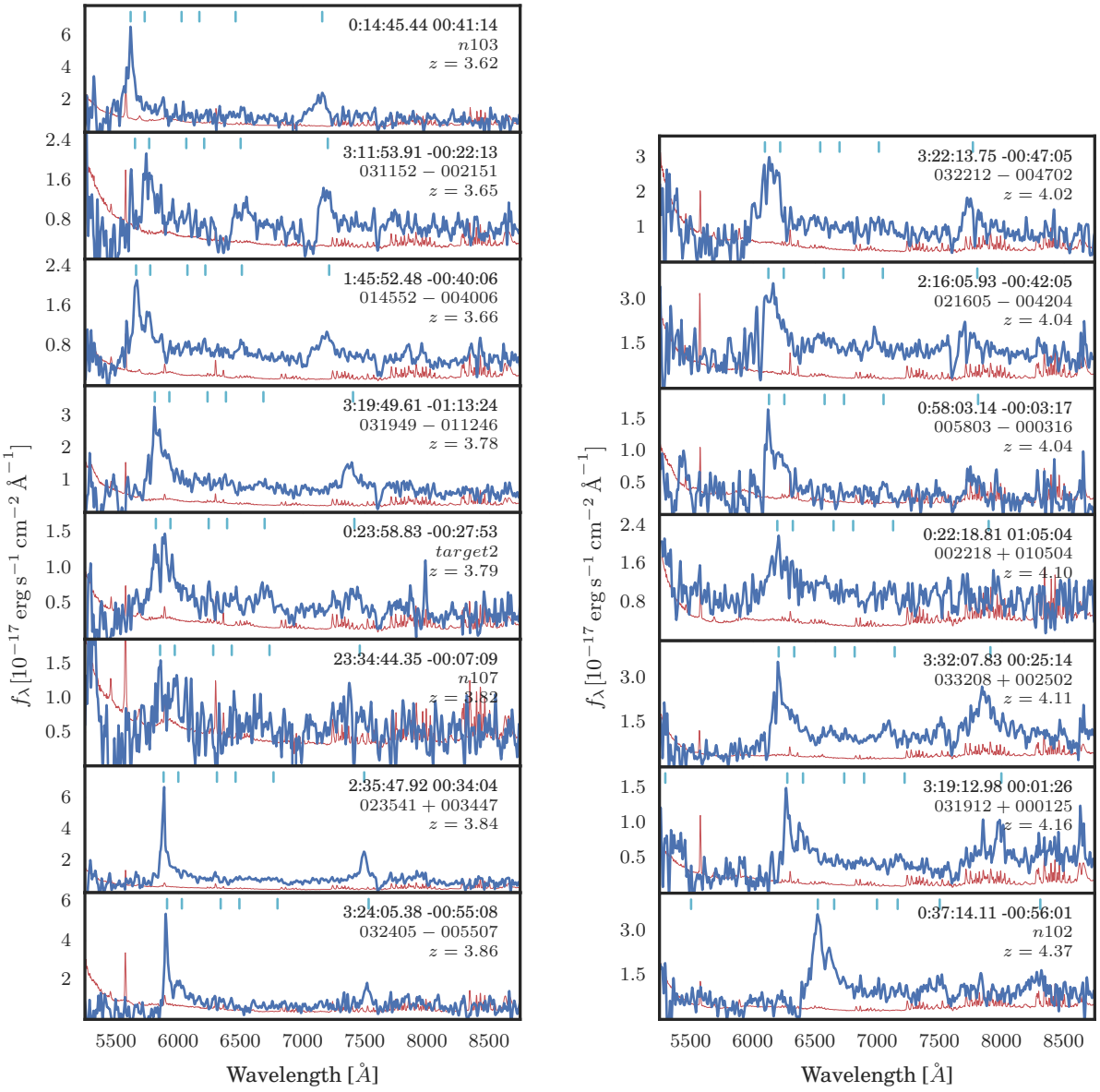

Figure 3.6 Spectra of Stripe 82 quasars observed with APO DIS, smoothed with a 13-pixel Hanning window. Markings follow figure 3.1.

Table 3.2. Stripe 82 $3.4 < z < 4.5$ Quasar Catalog

RA (J2000)	Dec (J2000)	$i_{\rm AB}$	M_{1450}	$z \pm \sigma_z$	Observation Date	Exposure time (s)	Seeing (")	Object Name
0:22:35.64	00:59:28	20.33	-25.15	3.421 ± 0.020	2013-09-06	450		a15
0:28:39.81	01:12:17			3.489 ± 0.020	2013-09-06	600		a11
0:28:12.89	-00:10:01	21.26	-24.22	3.498 ± 0.020	2013-09-06	600		a12
21:27:51.37	-00:09:42	21.40	-24.07	3.524 ± 0.020	2014-10-17	450	0.77	n208
23:16:40.39	-00:09:21			3.576 ± 0.020	2013-09-06	450		b14
2:56:43.22	-01:10:27	21.94	-23.53	3.577 ± 0.020	2013-10-25	399		LSSTJ025643011026
23:24:17.70	00:16:24	21.80	-23.68	3.580 ± 0.020	2014-10-17	450	0.77	n216
22:24:06.47	00:47:53	21.50	-23.99	3.580 ± 0.020	2014-10-17	450	0.77	n213
0:42:02.38	-00:57:54	21.42	-24.08	3.582 ± 0.020	2013-10-25	299		LSSTJ004202005754
23:19:07.38	-00:03:42	22.12	-23.35	3.584 ± 0.020	2014-10-17	750	0.77	n304
22:31:47.16	-00:22:44	20.69	-24.85	3.607 ± 0.020	2013-09-06	450		b11
3:16:44.92	-00:40:55	21.72	-23.78	3.616 ± 0.020	2013-10-24	599		LSSTJ031644004054
0:16:27.57	01:10:56	21.73	-23.77	3.620 ± 0.020	2014-10-20	600	1.43	n203
0:14:45.44	00:41:14	20.54	-25.01	3.621 ± 0.020	2014-12-20	1351		n103
21:29:02.32	-01:00:08	20.59	-24.97	3.630 ± 0.102	2014-10-17	450	0.77	n112
0:22:55.00	-00:25:33	21.94	-23.56	3.636 ± 0.020	2014-10-19	600	1.21	n202
22:45:38.92	00:33:52	21.29	-24.24	3.638 ± 0.020	2013-10-24	599		LSSTJ224538003351
23:09:37.65	-01:13:36	21.26	-24.28	3.638 ± 0.020	2013-10-24	599		LSSTJ230937011336
0:27:25.92	-00:44:00	21.34	-24.19	3.642 ± 0.022	2013-09-06	1050		a13
0:26:39.80	00:44:10	21.68	-23.83	3.643 ± 0.021	2014-10-19	600	1.21	n239
3:11:53.91	-00:22:13	20.74	-24.83	3.650 ± 0.028	2013-12-24	1200		031152-002151
1:45:52.48	-00:40:06	21.00	-24.56	3.657 ± 0.020	2014-01-01	2400		014552-004006
2:20:57.38	00:45:35	21.50	-24.04	3.668 ± 0.020	2014-10-19	900	0.97	br35
23:03:24.71	00:13:09	21.56	-23.98	3.671 ± 0.020	2013-10-24	599		LSSTJ230324001309
21:40:09.87	-00:10:30	21.14	-24.42	3.672 ± 0.020	2014-10-19	450	0.61	n220
22:14:28.41	00:27:32	22.09	-23.43	3.686 ± 0.020	2014-10-19	900	0.61	w06
0:27:05.35	-01:03:31	22.05	-23.48	3.692 ± 0.020	2013-09-06	3600		f01
0:21:58.02	-01:05:03			3.693 ± 0.020	2013-09-06	1800		f00
22:02:11.28	-00:50:28	21.81	-23.73	3.694 ± 0.020	2014-10-17	900	0.77	n230
3:17:50.05	00:30:37			3.703 ± 0.020	2013-10-24	599		LSSTJ031750003036
23:25:52.97	00:14:07	20.99	-24.60	3.710 ± 0.020	2014-10-17	450	0.77	n224
22:08:30.02	00:43:20	21.34	-24.23	3.716 ± 0.020	2013-09-06	450		b10
2:11:31.52	00:49:13	21.92	-23.63	3.720 ± 0.020	2014-10-19	900	0.97	br21
23:39:23.30	-01:10:38	21.29	-24.30	3.738 ± 0.020	2013-09-06	450		b15

Table 3.2 (cont'd)

RA (J2000)	Dec (J2000)	$i_{\rm AB}$	M_{1450}	$z \pm \sigma_z$	Observation Date	Exposure time (s)	Seeing (")	Object Name
21:32:39.11	-00:43:02	21.21	-24.39	3.743 ± 0.020	2013-10-24	599		LSSTJ213239004302
22:21:50.80	01:01:02			3.750 ± 0.020	2014-10-20	1800	1.12	w18
2:37:41.06	-00:37:15	21.85	-23.73	3.751 ± 0.020	2014-10-19	1200	0.97	br34
0:35:04.65	-00:51:10	21.40	-24.20	3.754 ± 0.020	2013-10-25	299		LSSTJ003504005110
3:24:43.13	-00:17:35	21.64	-23.95	3.759 ± 0.020	2014-10-19	900	0.97	br29
22:05:52.79	-00:32:02	21.65	-23.95	3.774 ± 0.020	2014-10-18	600	1.41	n207
3:19:49.61	-01:13:24	20.60	-25.05	3.781 ± 0.020	2013-12-24	1800		031949-011246
2:30:14.24	00:32:56	21.69	-23.91	3.782 ± 0.021	2014-10-20	600	1.50	br27
22:13:45.71	00:40:40	21.34	-24.28	3.786 ± 0.022	2014-10-18	900	1.41	n205
0:23:58.83	-00:27:53	21.59	-24.02	3.788 ± 0.020	2013-12-24	3604		target2
22:17:05.39	00:15:14	21.43	-24.20	3.794 ± 0.020	2014-10-20	600	1.41	w22
1:38:35.96	-00:10:33	21.83	-23.78	3.795 ± 0.020	2014-10-19	600	0.97	br09
22:20:36.99	00:03:03	21.67	-23.95	3.801 ± 0.020	2014-10-20	600	1.41	w02
22:02:51.15	00:09:50	21.07	-24.58	3.801 ± 0.020	2013-09-06	450		b09
2:25:55.13	-00:30:57	21.99	-23.61	3.802 ± 0.020	2014-10-19	900	0.97	br22
1:22:33.89	-00:04:17	21.72	-23.90	3.814 ± 0.020	2014-10-20	600	1.50	br15
23:34:44.35	-00:07:09	20.92	-24.74	3.816 ± 0.020	2014-12-20	1500		n107
2:36:15.11	00:04:30	21.63	-24.01	3.820 ± 0.020	2013-10-25	599		LSSTJ023615000429
1:39:21.18	00:01:00	21.53	-24.11	3.827 ± 0.020	2014-10-19	600	0.97	br01
2:35:47.92	00:34:04	21.10	-24.57	3.839 ± 0.020	2014-01-03	1800		023541+003447
21:57:29.27	01:06:14	20.44	-25.26	3.855 ± 0.020	2014-10-18	900	0.91	n105
3:24:05.38	-00:55:08			3.862 ± 0.020	2014-01-02	1200		032405-005507
1:23:44.48	-01:01:49	21.54	-24.13	3.866 ± 0.020	2014-10-20	600	1.50	br18
0:31:46.93	-00:33:23	22.05	-23.60	3.876 ± 0.020	2014-10-19	900	0.97	n303
23:05:08.80	00:43:04	21.12	-24.58	3.881 ± 0.020	2013-09-06	450		b12
23:48:19.33	00:39:00	21.45	-24.25	3.890 ± 0.020	2014-10-18	600	0.69	n212
22:12:38.20	-00:56:19			3.932 ± 0.020	2014-10-19	2400	0.68	w16
21:45:59.52	00:12:44	21.26	-24.50	3.959 ± 0.020	2014-10-19	300	0.61	n241
0:09:28.31	-00:26:13	21.71	-24.05	3.977 ± 0.020	2014-10-19	600	0.71	n217
22:52:56.03	-00:26:00	20.83	-24.95	3.979 ± 0.020	2014-10-18	600	1.41	n222
22:36:02.92	-00:55:07	21.68	-24.08	3.981 ± 0.020	2014-10-18	600	1.41	n210
22:22:03.25	01:10:18	21.53	-24.24	3.992 ± 0.020	2014-10-18	900	1.41	n234
0:07:58.79	00:45:40	21.28	-24.51	3.997 ± 0.020	2014-10-19	1200	0.71	n204
0:37:09.46	-00:03:15	21.76	-24.01	3.999 ± 0.020	2014-10-20	600	1.43	n214

Table 3.2 (cont'd)

RA (J2000)	Dec (J2000)	$i_{\rm AB}$	M_{1450}	$z \pm \sigma_z$	Observation Date	Exposure time (s)	Seeing (")	Object Name
21:53:30.84	00:37:18	19.88	-25.95	4.002 ± 0.020	2013-10-25	299		LSSTJ215330003717
2:06:27.34	00:08:46			4.004 ± 0.020	2014-10-20	900	1.50	br10
21:32:06.42	01:04:10	21.30	-24.50	4.013 ± 0.020	2014-10-17	450	0.77	n215
0:21:17.21	00:14:41	22.07	-23.71	4.015 ± 0.020	2014-10-18	1800	0.69	n302
3:22:13.75	-00:47:05	20.20	-25.63	4.016 ± 0.020	2013-12-24	900		032212-004702
22:13:32.65	00:30:25			4.019 ± 0.020	2014-10-19	1200	0.68	w14
2:16:05.93	-00:42:05	20.75	-25.09	4.039 ± 0.020	2014-01-02	2400		021605-004204
0:58:03.14	-00:03:17	20.83	-25.01	4.043 ± 0.020	2014-01-03	2400		005803-000316
23:14:38.65	01:11:37	22.23	-23.58	4.066 ± 0.020	2014-10-17	900	0.77	n311
21:38:57.92	-00:30:11	20.92	-24.94	4.074 ± 0.020	2013-10-24	599		LSSTJ213857003011
22:18:21.54	-00:13:25	21.88	-23.97	4.093 ± 0.020	2014-10-19	1200	0.71	w10
0:22:18.81	01:05:04			4.098 ± 0.020	2014-10-03	1800		002218+010504
3:32:07.83	00:25:14	20.66	-25.23	4.107 ± 0.020	2014-10-03	1800		033208+002502
0:14:34.90	00:19:39	21.86	-24.02	4.120 ± 0.021	2014-10-20	1200	1.43	n238
23:14:19.98	01:07:06	21.30	-24.62	4.151 ± 0.020	2014-10-17	450	0.77	n225
21:56:43.23	-00:08:33	19.56	-26.39	4.155 ± 0.020	2014-10-18	450	1.41	n115
0:05:09.00	-00:58:17	21.40	-24.52	4.160 ± 0.020	2014-10-20	600	1.43	n223
3:19:12.98	00:01:26	20.66	-25.28	4.164 ± 0.020	2014-01-01	4200		031912+000125
1:24:46.77	00:46:00	21.67	-24.28	4.194 ± 0.020	2014-10-20	600	1.50	br19
0:34:13.07	-00:30:47	22.05	-23.88	4.197 ± 0.020	2014-10-19	1800	0.97	n323
22:02:49.81	-00:52:18	21.18	-24.78	4.201 ± 0.021	2014-10-17	450	0.77	n229
22:13:42.62	01:02:54	22.42	-23.54	4.236 ± 0.020	2014-10-19	1200	0.68	w09
0:32:05.22	00:15:15	21.85	-24.13	4.241 ± 0.020	2014-10-19	600	1.21	n237
21:44:40.56	01:09:06	20.70	-25.30	4.249 ± 0.020	2014-10-19	300	0.61	n101
0:38:51.93	-00:35:24	22.06	-23.92	4.250 ± 0.020	2014-10-19	900	0.97	n306
22:09:16.01	01:08:58	21.71	-24.32	4.306 ± 0.020	2014-10-20	1200	1.55	w11
0:34:48.37	00:23:00	21.45	-24.60	4.344 ± 0.021	2014-10-19	600	1.21	n219
22:12:00.09	00:59:50	21.60	-24.46	4.360 ± 0.028	2014-10-20	600	1.41	w29
23:26:26.32	-00:12:29	21.84	-24.22	4.365 ± 0.020	2014-10-17	450	0.77	n201
0:37:14.11	-00:56:01	20.33	-25.74	4.366 ± 0.020	2014-10-03	1802		n102
0:52:16.56	-01:10:45	20.55	-25.53	4.385 ± 0.020	2013-10-25	299		LSSTJ005216011044
21:40:13.42	-00:33:55	19.90	-26.18	4.398 ± 0.020	2014-10-17	450	0.77	n104
23:58:14.25	00:20:49	21.93	-24.15	4.437 ± 0.020	2014-10-18	900	0.69	n307
0:38:59.63	00:49:00	21.20	-24.89	4.548 ± 0.020	2014-10-18	900	0.69	n226

Table 3.2 (cont'd)

RA (J2000)	Dec (J2000)	$i_{\rm AB}$	M_{1450}	$z \pm \sigma_z$	Observation Date	Exposure time (s)	Seeing (")	Object Name

3.3 Conclusion

To ensure that the training set of labeled $z \sim 4$ lightcurves resembles the data to be classified, we obtained spectroscopic redshifts for 102 $z \sim 4$ quasars that extend fainter than the set of previously known $z \sim 4$ quasars (figure 3.7). The observing runs were focused on obtaining spectroscopic completeness on a $55\,\mathrm{deg}^2$ subregion. This work increases the set of spectroscopically confirmed $3.75 < z < 4.5$ quasars by 33% on the full $238\,\mathrm{deg}^2$ Stripe 82 region and by 78% on the $55\,\mathrm{deg}^2$ subregion.

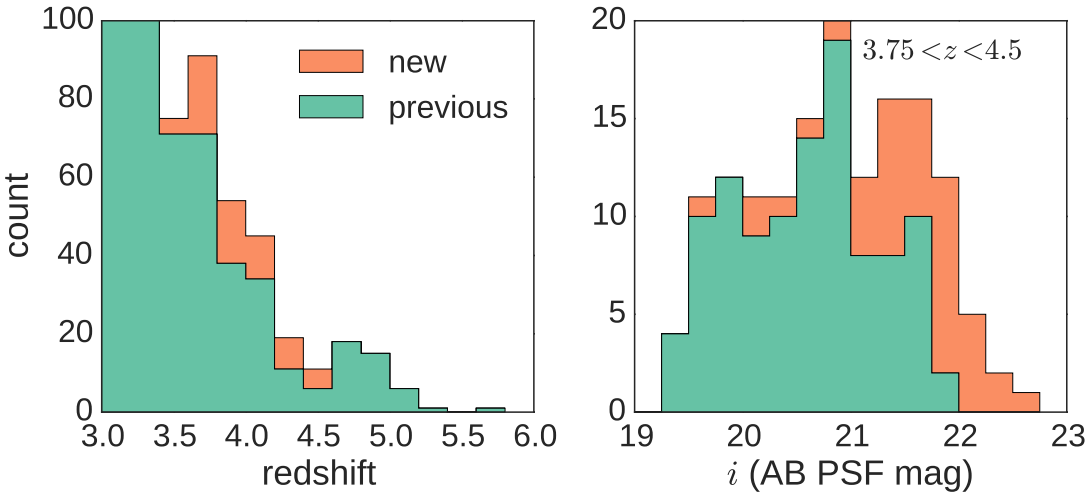

Figure 3.7 Quasars spectroscopically confirmed at MMT and the APO 3.5m 2013–2014 are shown in orange. Previously confirmed quasars found are shown in green.

3.4 Supplementary Information

3.4.1 Contaminant Spectra

For completeness, we present in figures 3.8, 3.9, and 3.10 the spectra obtained at MMT, APO, and Gemini that were not positively identified as quasars.

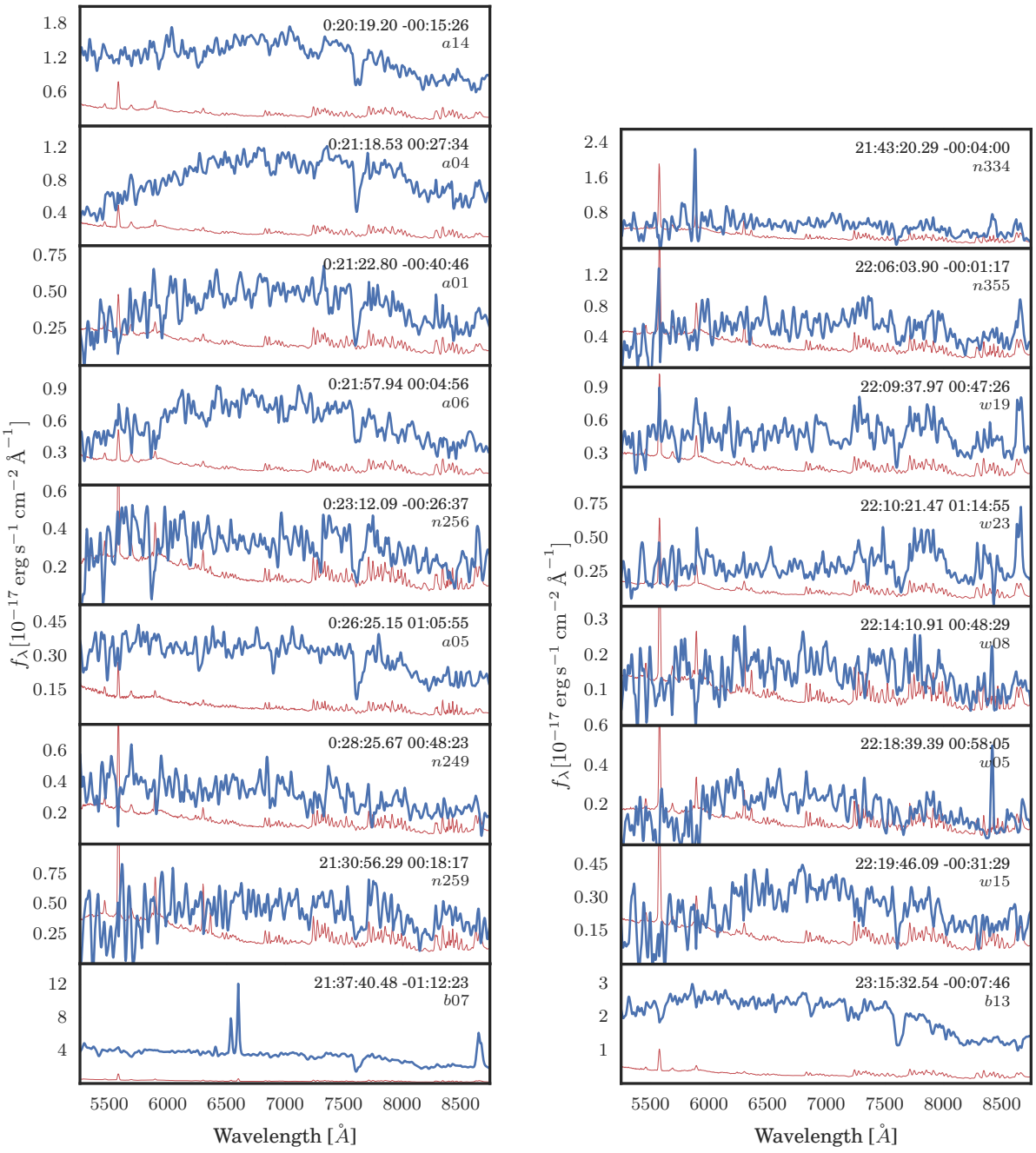

Figure 3.8 Spectra of Stripe 82 contaminants observed with MMT Red Channel Spectrograph, smoothed with a 13-pixel Hanning window.

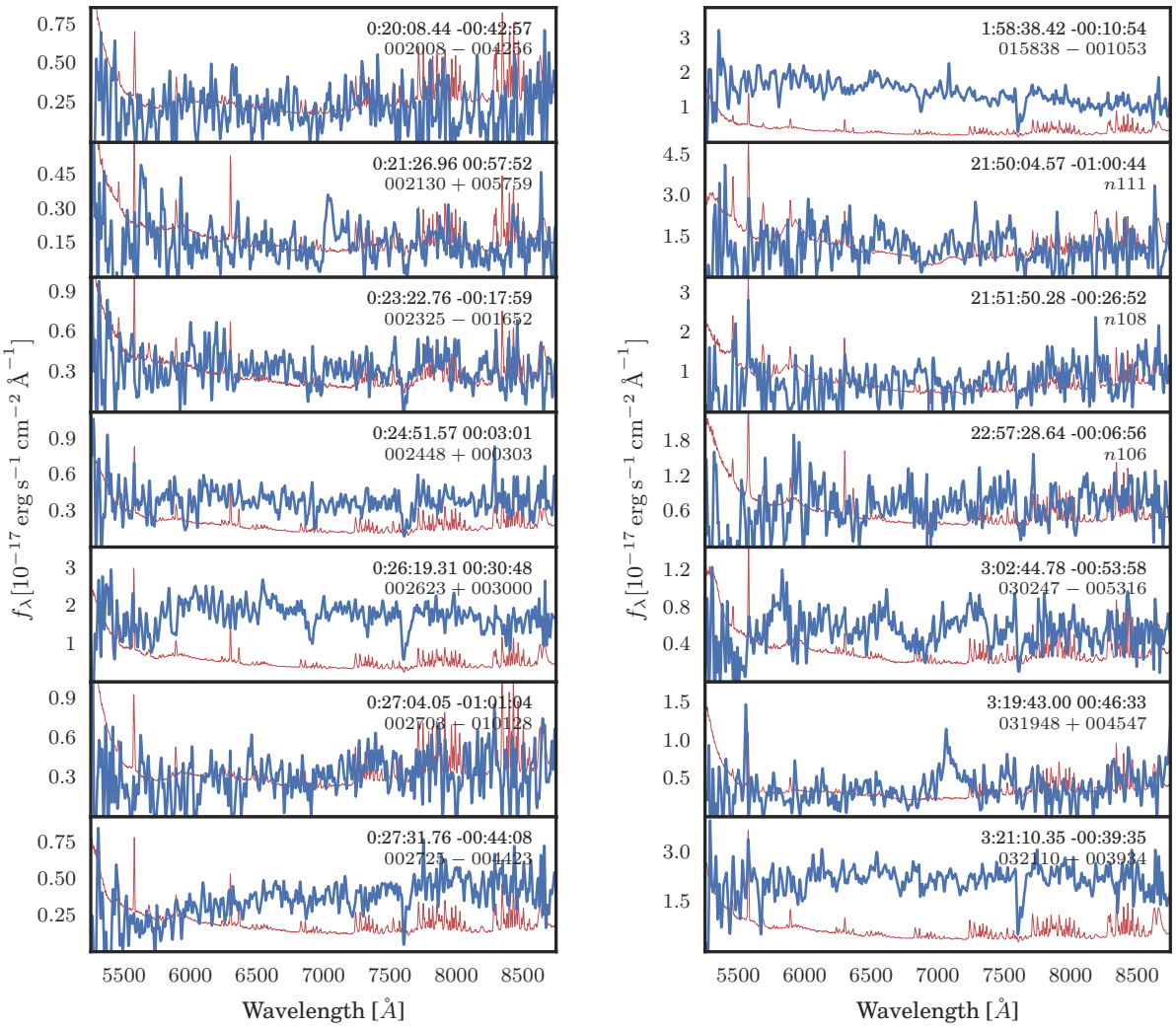

Figure 3.9 Spectra of Stripe 82 contaminants observed with APO DIS, smoothed with a 13-pixel Hanning window. Markings follow figure 3.1.

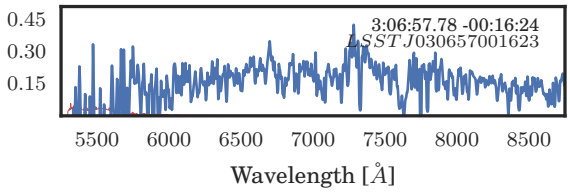

Figure 3.10 Spectrum of Stripe 82 contaminant observed with Gemini GMOS-N, smoothed with a 13-pixel Hanning window. Markings follow figure 3.1.

Chapter 4

VARIABILITY AND CLASSIFICATION OF $Z \sim 4$ QUASARS

4.1 Introduction

With the Large Synoptic Survey Telescope (LSST), we expect to find over a million quasars with redshifts between 4 and 7. However, we will not be able to get spectroscopic follow-up for all of them. To estimate statistics like luminosity functions and clustering measurements, we will need purely photometric methods to classify quasars, estimate redshifts, and estimate selection functions. Photometric selection methods are also critical for efficiently selecting targets for spectroscopic surveys. Quasars exhibit well-studied characteristics that distinguish them from stars and galaxies, allowing them to be accurately selected. These characteristics include both their morphology (Richards et al. 2002), their colors (Sandage et al. 1965; Fan 1999; Richards et al. 2001; Croom et al. 2009; Kirkpatrick et al. 2011; Bovy et al. 2012), their lack of proper motion (Sandage & Luyten 1967; Kron & Chiu 1981; Richards et al. 2002), and their variability Ulrich et al. (1997); Koo et al. (1986); MacLeod et al. (2011); Palanque-Delabrouille et al. (2011); Peters et al. (2015). With the proliferation of new and future time-domain surveys (e.g. LSST, Pan-STARRS, the Dark Energy Survey, the Intermediate Palomar (and Zwicky) Transient Facility) variability selection methods will become fundamental to identifying quasars, bolstering less reliable color selection methods. As with other areas of astronomy, large time-domain surveys will advance studies that take advantage of time-series photometry and that rely less on spectroscopy.

In response to the development of these large time-domain surveys, research on quasars' variability signatures has been especially active (Graham et al. 2014; Kelly et al. 2009; Kozłowski et al. 2010; MacLeod et al. 2010; Schmidt et al. 2010; Zu et al. 2013). Whereas there is a wealth of recent studies of quasar variability, these studies have focused primarily

on low-redshift quasars at $z < 3.5$, because these are most plentiful. Methods that are used to classify low-redshift quasars do not work on high-redshift quasars; the published efficiencies for quasar-star separation must be adjusted for use on high-redshift quasars. For example, MacLeod et al. (2011) demonstrated that g-band DRW parameters efficiently separated stars and quasars for the DR7 quasars in the Schneider et al. (2010) catalog, 97% of which were $z < 3.5$. At $z \sim 4$, a quasar "drops out" of the g filter, which covers the part of the spectrum absorbed by intergalactic HI, the so-called Lyman-alpha forest, and the i band probes a similar rest wavelength as the g band at $z \sim 2$, which is the peak of the redshift distribution in DR7.

To develop a pure and complete sample of $z \sim 4$ quasars based on photometric information alone, we investigate (color and) variability-based classification methods for selecting high-redshift, $z > 3.5$, quasars. We compare the performance of classification methods in terms of three major factors: feature quality, training-set and label definition, and classification algorithms.

Features are central to classifier performance. Variability-based quasar classification relies on the ability to distinguish between lightcurves. Identifying the differences directly from the lightcurves themselves is challenging, because lightcurves are sampled at irregular intervals, observed with different numbers of epochs and SNRs, and have high dimensionality. Therefore, we standardize the data by transforming each lightcurve into a set of "features" with statistics (parameter estimates such as χ^2) and fitting variability models. Features are the measurable properties (independent variables) from which the label for the data point is predicted (e.g $(g-r)$ color, `extendedness`, or χ^2_g), and are the first major factor we use to evaluate classifier performance. Some features are more useful than others in discrimination, and features are also often correlated and redundant. Investigators thus must prepare and prune the features they use to classify, a process called *feature engineering*; this is where most of the work in developing classification methods lies.

We focus on two aspects of feature engineering that make variability-based classifiers useful in selecting quasars: accuracy in dealing with noisy data and high discrimination

power. First, how well do the variability estimators recover the "true" lightcurve parameters that we want to measure in the presence of noise? Because the color and morphology estimators were validated in chapter 2, in this chapter we focus specifically on variability estimators. We use simulated lightcurves with realistic noise added to investigate the bias and variance of these estimators. Because faint sources are more affected by noise, we specifically focus on how these estimators degrade with increasing magnitude. Second, which combination of features yields the best discriminating power? Using information theory, we perform feature selection using a training dataset of spectroscopically labeled lightcurves.

The quality of this training set is the second factor that affects classification performance. Building a classifier requires many instances of previously labeled lightcurves, grouped into what is called a training set. It is imperative that the training set resemble the actual data set that will be classified, with respect to the distribution in feature space.[1] Astronomers are familiar with the problem of training sets not matching actual datasets. Because new telescopes are more sensitive and can observe fainter objects, training sets (based on old, well-studied data) are often brighter than the new data to be classified. Populations may be fainter because they either have dimmer intrinsic luminosities or have higher redshifts; there is no guarantee that low-luminosity and high-luminosity populations are identical, or that nearer/later and farther/younger populations have similar characteristics.

The third factor that affects classification performance is the classification algorithm. It has been shown that combining multiple classifiers/models performs better than (and is less subject to overfitting than) using one model in isolation (Opitz & Maclin 1999). One interpretation of this concept of using a committee of classifiers, sometimes called "bagging," trains several classifiers and averages their predictions. Random Forest (Breiman 2001), which is based on randomized decision forests, is a popular bagging implementation in the astronomy community. Instead of averaging the predictions of a set of classifiers, some ensemble classifiers combine models by "boosting." Boosting involves training classifiers in

[1]The density of points in the N-dimensions defined by the lightcurves' features such as median $g - r$, median $r - i$, `Extendedness`, and $\chi^2_\nu(i)$.

a sequence in which the error function used to train each model depends on the accuracy of the previous iterations in the sequence. A popular adaptive boosting algorithm is Adaboost (Freund & Schapire 1997; Schapire 2003). Random Forest and Adaboost are the two ensemble classification algorithms we test and evaluate.

In addition to evaluating photometric features and classification algorithms for high-z quasar selection, in this chapter we also present a catalog of variability features for the 10 million point sources in the Stripe 82 forced photometry data set and the final classifier based simultaneously on colors and variability. We conclude with the description of the final classifier used to select the $z \sim 4$ quasar sample that was used to compute the luminosity function in chapter 5.

4.2 Background

At $z \sim 4$, quasars exhibit a number of characteristics that allow them to be selected efficiently and nearly completely from a dense foreground of stars and galaxies. These characteristics include morphology, color, and variability.

Morphologically, at $z \sim 4$ we expect all quasars to appear as compact point sources, because to be observed at $i < 22.5$, the absolute magnitude of any galaxy must be brighter than $M_{1450} = -23.0$. Therefore, we expect quasars to completely outshine their host galaxies and appear compact in our survey. Furthermore, if the quasar host galaxy were bright enough to be observed, it would likely also appear compact. However, simply selecting point sources would not eliminate galaxies from the survey. Although all $z \sim 4$ quasars appear compact in our survey, not all galaxies are extended; faint galaxies also appear compact.

The second characteristic that allows for efficient quasar selection is color. Quasar candidates have been selected from compact stars and galaxies using their optical colors ever since the first quasar was discovered, in 1963. For example, (Sandage et al. 1965) used UV Excess to select star-like objects with unusually blue broadband colors.

Color selection techniques change depending on the redshift range probed, because quasar colors depend strongly on redshift. With increasing redshift, the characteristic broad emis-

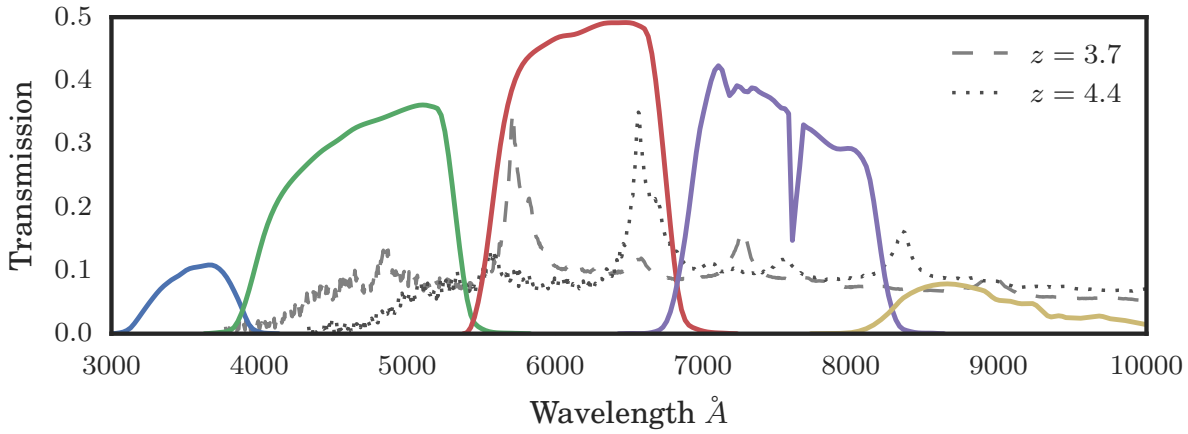

Figure 4.1 SDSS transmission curves and composite quasar spectrum (Vanden Berk et al. 2001) with $z_{average} \sim 2$ redshifted to $z \sim 3.7$ and $z \sim 4.4$.

sion lines and intergalactic absorption (the Lyα-forest and the Lyman break) march redward. In certain redshift ranges, specifically at $z \sim 2.7$, color selection techniques fail because the optical colors of quasars resemble those of stars (Fan 1999; Richards et al. 2001; Ross et al. 2013). At high redshift, however, colors are well separated from the stellar locus (Figure 4.2), and strong color dependence on redshift allows for the computation of photometric redshifts (photo-z's) for quasars.

IR colors are important for high-z redshift selection as well. At $z \sim 6$, the Lyα emission line is covered by the z-band, the reddest band available in most optical surveys.[2] Any quasars $z > 5$ require near-IR bands for identification. Additionally, galaxies and quasars both have dust that emits light in the mid-IR, whereas stars do not. Therefore, in general, mid-IR colors are useful for eliminating stellar contaminants. However, we do not use IR in this study because at $i \sim 22$ most sources contaminating high-z quasar samples are *galaxies*, which resemble quasars in the IR, and no sufficiently deep IR surveys cover all of Stripe 82.

At $z \sim 4$, quasar colors ($g-r$ and $r-i$) are well separated from those of stars. Figure 4.1

[2]Many future ground-based optical surveys (e.g. LSST) plan to include a redder y-band.

shows a composite quasar spectrum (Vanden Berk et al. 2001) (with $z_{average} \sim 2$) redshifted to $z = 3.7$ and $z = 4.4$ (the lower and upper edges of this $z \sim 4$ survey), overlaid with the filter transmission curves from SDSS. At $z > 3.7$, the u-band is blueward of the Lyman limit, and quasars will be faint or not detected in the u-band. In this target redshift range, the Lyα line is covered by the r-band and Ly*alpha* forest covered by the g-band, producing very red $(g-r)$ colors. As redshift increases, the Lyman break moves through g band, producing fainter g fluxes. The $(g - r)$ colors thus become redder with increasing redshifts as the Lyman break moves across the g-band; $(g - r)$ strongly correlates with redshift providing uncertainties in photometric redshifts of $\sigma_z < 0.15$. As seen in color tracks of model quasar SEDs (Figure 4.2), the redder $(g - r)$ colors of quasars move redward of the stellar locus in this redshift range. Colors in this redshift range are also luminosity dependent and subject to large intrinsic scatter; this is especially true in $(g - r)$ because of absorption in the Lyα forest, which varies from sightline to sightline.

The third characteristic that allows for efficient selection of quasars is variability; quasars are highly variable at all wavelengths. Of the known quasars, 90% have an RMS variability of at least 0.03 mag (Sesar et al. 2007). However, a simple RMS cannot distinguish quasars from other variable objects such as RR Lyrae stars. Instead, we must distinguish quasars using the timescales of their flux variability: quasar fluxes vary on longer timescales ($\sim 10^{2-3}$ days) than do variable stars. Indeed, the RMS *as a function of time lag between observations*—a quantity called the structure function, which is an empirical estimate of the lightcurve's power spectrum—*can* efficiently separate quasars from variable stars (Schmidt et al. 2010, 2012; Palanque-Delabrouille et al. 2011; Peters et al. 2015).

Indeed, quasars' variability characteristics, and particularly their spectral density, have been actively researched in recent years. Kelly et al. (2009) showed that quasar lightcurves are well described by a continuous-time autoregressive process of order 1 (CAR(1)), also referred to as a damped random walk (DRW) or an Ornstein-Uhlenbeck process. A CAR(1) model can be simply parameterized by a characteristic damping timescale (τ) and amplitude of variability ($\hat{\sigma}$). Here, $\hat{\sigma}$ does not represent the estimate of σ (the overall standard deviation

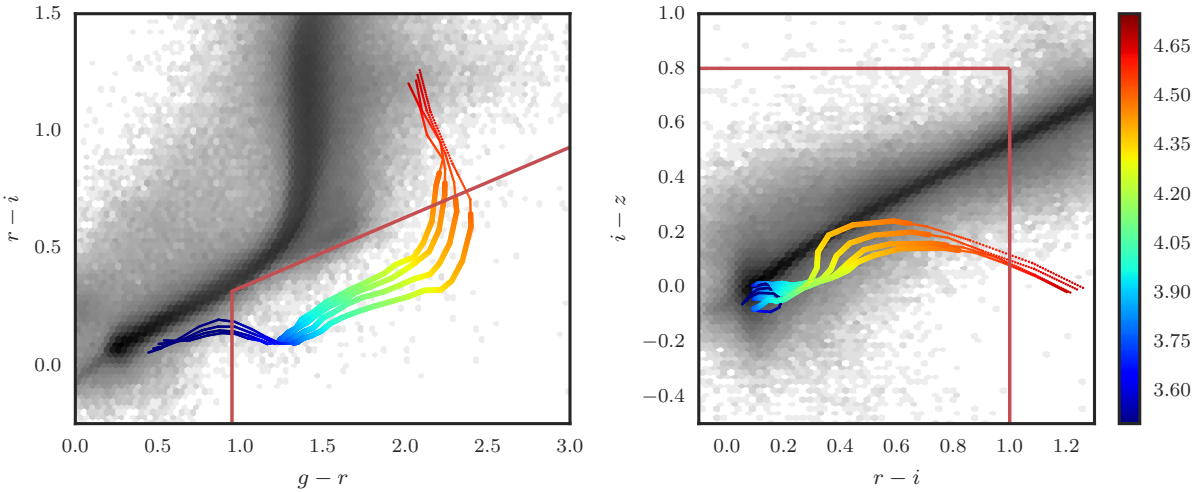

Figure 4.2 Mean color tracks for simulated quasars (McGreer et al. 2013) in four luminosity bins for $3.5 < z < 4.75$. Bold lines denote the main redshift range $3.75 < z < 4.5$. The background image shows the logarithmically scaled distribution of colors for point sources $i < 22$. The red lines show our color cuts. Quasars at $z \sim 4$ are well separated from the stellar locus. The simulated models take into account the luminosity dependence.

in the flux), but instead follows the notation of Kelly et al. (2009), wherein the amplitude of short-term driving variability $\hat{\sigma}$ is related to the overall standard deviation in the flux σ by $\hat{\sigma} = \sqrt{2\sigma^2/\tau}$. Kozłowski et al. (2010) tested this model using a larger sample of lightcurves and concluded that it is efficient for AGN selection; MacLeod et al. (2011) showed that the CAR(1) model parameters provide efficient and complete selection of quasars on Stripe 82 lightcurves. Additional studies with Stripe 82 have also demonstrated the power of using a CAR(1) variability-selection model to select quasars (MacLeod et al. 2011; Butler & Bloom 2011; Choi et al. 2014).

There are other metrics describing quasar variability as a function of timescale that do not depend on fitting a full CAR(1) model. Graham et al. (2014) recently showed that spectral density can be estimated using wavelets. This is analogous to the more basic binned-χ^2

metric that we evaluate in this study. A simple χ^2 metric formally tests the hypothesis that an object has a steady flux that can be modeled by a single Gaussian with a mean and variance. If the reduced χ^2 is large, then the steady-flux model fits the lightcurve poorly; this also indirectly tests the converse hypothesis, that the source is variable on the timescales on which the lightcurve is sampled. Nominally, this timescale is determined by the cadence of the observations. However, the lightcurve can be binned to transform the cadence from frequent low-SNR observations to less frequent observations with higher SNRs. The reduced χ^2 of this new binned lightcurve tests whether the source is variable on timescales equal to or *longer* than the time between bin centers. If the object is variable over a timescale *shorter* than this bin size, the variability will average out within the bin, and the binned lightcurve will appear constant. Because quasars vary on timescales on the order of 10^{2-3} days, we bin the lightcurves annually, and compute the reduced χ^2 on these annually averaged lightcurves. We call this metric the annually averaged χ^2. Stellar annually averaged lightcurves will appear constant and thus have a low annually averaged χ^2. Next, we discuss how we computed this metric and the other features in the catalog.

4.3 Building the Variability Catalog

After reprocessing the Stripe 82 images and obtaining lightcurves, as described in chapter 2, we computed the features for selection criteria: colors, extendedness, and variability. See section 4.4 for a description of how these features were applied and how selection criteria were ultimately chosen; this section will describe how features were computed.

Colors: For color measurements, we used catalog co-adds, aggregates of the forced PSF-fluxes that were measured on the single-epoch images. For the flux estimator we use a median; for the 1σ uncertainty estimator we chose $\hat{\sigma} = 0.742\text{IQR}/\sqrt{(N)}$, where IQR is the interquartile range. These estimators were chosen from a variety of options (see chapter 2) that were evaluated based on the RMS of the stellar locus (Figure 2.9), which provides an independent measure of a survey's photometric uncertainty. We ultimately chose medians because they were the most robust over the full range of magnitudes.

Morphology: We use a very simple metric to separate point sources from extended sources: $2.5\log$[modelFlux/psfFlux], which we label `extendedness`. For true point sources, the best-fitting source model should approximate the PSF, and the model-flux and PSF-flux should be equal. Therefore, we expect point sources to have an "extendedness" of 0, and extended sources to have an "extendedness" of > 0.

Variability Metrics: We computed a number of variability metrics for each of the six lightcurves: u, g, r, i, z, and a gri-merged lightcurve. The idea behind the gri-merged lightcurve is that quasar lightcurves co-vary in all optical bands.[3] In theory, the variability signal should be higher in a multi-band lightcurve than in any single lightcurve.

For each epoch j, the gri-lightcurve consists of $gri_j = 0.15g_j + 0.45r_j + 0.4i_j$, where g, r, i are the fluxes in the g, r, and i band respectively at epoch j. Note that this formulation is only possible with a camera such as SDSS, where the photometry in each passband was taken simultaneously (1.5 min apart in this dataset) relative to the quasar variability timescale of 10^{2-3} days. The weighting was derived from model spectral energy distributions (SEDs) of quasars at $z \sim 4$. Whereas we expect the intrinsic variability of a $z \sim 4$ quasar to have a higher amplitude in the g-band, the flux in g is absorbed by the Lyα-forest, and is thus subject to photometric uncertainties. We did not weight each band by the measured inverse variance, because then each gri-lightcurve would depend on the colors of the source.[4] The idea of combining bands using the weightings based on the target SEDs is not new. Lang et al. (in prep) points out that this idea can be framed as a multi-band-detection matched-filter.

In addition to computing variability statistics on the original u, g, r, i, z lightcurves and derived gri- lightcurves, we also computed variability statistics on annually averaged lightcurves, which we denote $\{y_1, y_2, \ldots y_{10}\}$, where y_j is the averaged PSF flux for the jth season, and $\{\sigma_{y_1}, \sigma_{y_2}, \ldots \sigma_{y_10}\}$ are the uncertainties on y_j. As introduced above, there are benefits to

[3]This idea also led to a trial of using the correlation between the bands as metric for intrinsic variability, but we found this correlation to be very weak and not useful.

[4]We also found empirically that the power to discriminate quasars from stars and galaxies was best when we used annually averaged χ^2 metrics derived from the model-weighting used rather than per-observation inverse-variance weighting.

computing annually averaged lightcurves to find quasars in Stripe 82 data. First, quasars vary on long time scales, and we therefore expect to see variability from season to season. In fact, we want to separate quasar lightcurves from lightcurves that show intra-season variability. Second, binning by year standardizes the Stripe 82 cadence in two ways. Recall that lightcurves from the North scan lines contain more (and different) epochs than those from the South scan lines. Also recall that in the early years there was only one observation per year, and in later years there were as many as 20. Binning makes the N/S lightcurves consistent; it also makes the cadence throughout the decade of observations consistent.

Using annual averaging, each 80-epoch lightcurve is reduced to 10 epochs (1 per season), and the average flux for a given season is computed as the mean weighted by the inverse variance. The uncertainty is given by the weighted standard deviation corrected for the observed dispersion. Note that these annually averaged metrics were computed *in addition to*, not in place of, the metrics computed over the original lightcurves.

Table 4.1 lists a subset of the variability metrics computed, where x_i is the calibrated PSF flux, `psfFlux`, at each epoch i, and σ_i is the measurement uncertainty, `psfFluxErr`, from the variance plane of the single-epoch image.

Table 4.1. Lightcurve Aggregates and Variability Metrics

Name	Formula	Description
`N`	N	Count of epochs in lightcurve
`e50`	e_{50}	Median of the `psfFluxErr`
`median`	$x_{q_{50}}$	Median of the `psfFlux`
`mean`	$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$	Sample mean of the `psfFlux`
`sigma`	$s = \sqrt{\frac{1}{N-1}\sum(x_i - \overline{x})^2}$	Sample standard deviation of the `psfFlux`
`sigmaG`	$\sigma_G = 0.7314(x_{q_{75}} - x_{q_{25}})$	Robust sample standard deviation of `psfFlux` derived from IQR
`e_mean`	$\overline{\sigma} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i$	Mean of the `psfFluxErr`
`intrinsic_sigma`	$\sigma_0 = \sqrt{\sigma_G^2 - e_{50}^2}$	See Section 4.9.1.1 for motivation.
`chi2`	$\chi^2_\nu = \frac{1}{N-1}\sum_{i=1}^{N} \frac{(x_i - \overline{x})^2}{\sigma_i^2}$	Reduced χ^2
`WeightedMean`	$\overline{x} = \frac{\sum_{i=1}^{N}(x_i \sigma^{-2})}{\sum_{i=1}^{N} \sigma_i^{-2}}$	Weighted mean of the `psfFlux`
`WeightedStdCorr`	$\sigma^2_{\overline{x}} = \frac{1}{\sum_{i=1}^{N} \sigma_i^{-2}} \times \frac{1}{(n-1)}\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\sigma_i^2}$	Dispersion-corrected, weighted standard deviation
`sigmaClipped_N`	$\lvert\mathcal{C}\rvert$ where $\mathcal{C} = \{i \lvert \overline{x} - 3s < x_i < \overline{x} + 3s\}$	Count of epochs within 3 sigma
`sigmaClipped_mean`	$\overline{x} = \frac{1}{N}\sum_{i\in\mathcal{C}} x_i$	Mean of sigma-clipped lightcurve
`sigmaClipped_std`	$s = \sqrt{\frac{1}{N-1}\sum_{i\in\mathcal{C}}(x_i - \overline{x})^2}$	Sample standard deviation of sigma-clipped lightcurve
`sigmaClipped_chi2`	$\chi^2_\nu = \sum_{i\in\mathcal{C}} \frac{(x_i - \overline{x})^2}{\sigma_i^2}$	Reduced chi-squared of sigma-clipped lightcurve
`sigmaClipped_wm`	$\overline{x} = \frac{\sum_{i\in\mathcal{C}}(x_i \sigma^{-2})}{\sum_{i\in\mathcal{C}} \sigma_i^{-2}}$	Weighted mean of sigma-clipped lightcurve
`sigmaClipped_wStdCorr`	$\sigma^2_{\overline{x}} = \frac{1}{\sum_{i\in\mathcal{C}} \sigma_i^{-2}} \times \frac{1}{(n-1)}\sum_{i\in\mathcal{C}} \frac{(x_i - \bar{x})^2}{\sigma_i^2}$	Dispersion-corrected, weighted standard deviation of sigma-clipped lightcurve
`annAvg_chi2_wm`	$\chi^2_9 = \sum_{j=1}^{10} \frac{(y_j - \overline{y})^2}{\sigma_j^2(y)}$	χ^2_ν of the annually averaged lightcurve where the average flux in in year j is the weighted mean $y_j = \frac{\sum_{i\in\mathcal{Y}}(x_i \sigma^{-2})}{\sum_{i\in\mathcal{Y}} \sigma_i^{-2}}$ where $\mathcal{Y} = \{i \lvert i \in j\text{th yr}\}$
`annAvg_chi2_clip_wm`	$\chi^2_9 = \sum_{j=1}^{10} \frac{(y_j - \overline{y})^2}{\sigma_j^2(y)}$	Same as `annAvg_chi2_wm` except each season is the 3 sigma-clipped weighted mean
`annAvg_e50_clip_wm`	$\text{median}(\sigma^2_{y_j})$	Median of the uncertainties on the annually averaged lightcurves: $\sigma^2_{y_j} = \frac{1}{\sum_{i\in\mathcal{Y}} \sigma_i^{-2}} \times \frac{1}{(\lvert\mathcal{Y}\rvert - 1)}\sum_{i\in\mathcal{Y}} \frac{(x_i - \bar{x})^2}{\sigma_i^2}$
`annAvg_rms_clip_wm`	$\sqrt{\frac{1}{9}\sum(y_j - \overline{y})^2}$	Sample standard deviation of the annually averaged lightcurve

4.4 Initial $z \sim 4$ Candidate Selection

We combine morphology, color, and variability selection methods to classify $z \sim 4$ quasars. In the following subsections we describe our selection procedure in more detail.

We start by filtering the 10 million sources using criteria chosen for retaining the complete sample of known $z \sim 4$ quasars. We track the 306 spectroscopically confirmed $3.75 < z < 4.5$ quasars (which can be split into two categories, previously identified quasars and quasars presented in chapter 3) through the criteria in survey area, morphology, color, and variability.

4.4.1 Uniform Survey Area

To limit the survey to a quantifiable, uniform volume, we required that objects fulfill the following criteria:

1. They must have i-band PSF fluxes brighter than the 22.5th magnitude, corrected for Galactic extinction.

2. They must lie between $321.5 < R.A. < 54.9$, $-1.26 < Dec. < +1.26$

3. They must not have a parent source brighter than $i < 17$. This criterion effectively masks regions around objects brighter than $i < 17$ at the catalog level. Model fluxes and forced photometry in these footprints were unreliable because of the challenges of deblending a stellar footprint. For point sources close to bright stars ($i < 17$), the deblender algorithm assigns some of the scattered flux to the child sources. This gives the child sources larger model-fluxes and `extendedness` > 0.1.

4. Finally, the candidates must not be in any of the four patches where the average SNR of $i \sim 22.5$ point sources was less than 5: patches (162,8); (163,8); (349,0); and (6,6) all had lower SNR because of substantial scattered light from bright stars and were inaccessible to the full i-band range of this survey. The SNR is measured

as the `psfFlux/psfFluxErr`, where psfFluxErr is the 1σ uncertainty of the PSF flux measured on the i-band co-add.

We treat survey depth as uniform because all unmasked area is deeper than the survey limit of $i \sim 22.1$. Figure 4.3 shows the 5σ depth as a function of survey area. All masked regions were removed from the survey area (see §5.4.1 for more details on the algorithm used to compute masks).

This criterion that the candidates not intersect a masked region removed the candidates that did not intersect a quantifiable, uniform survey area, which removed 35 of the 306 known quasars $3.75 < z < 4.5$ on Stripe 82. One that had been previously identified in DR9QSO was not detected in the $i-$ band co-add because it was in a footprint of an $i \sim 16$ star[5]. This quasar was not detected because our deblender was limited to 40 children, and the halo of this star contained more than 40 peaks. The 34 other known quasars were detected in the co-adds, but lay in the footprint of bright stars.

4.4.2 Morphology

To further cull the candidates, we selected only sources that appeared compact in the i-band co-add by requiring that `extendedness` < 0.1. Recall that `extendedness` is the ratio of model flux to the PSF flux. The model flux used was the multi-shapelet flux. If the multi-shapelet flux measurement failed for any reason, it was replaced with a 2-D Gaussian flux. We chose this `extendedness` threshold by analyzing a region of the sample that overlapped with archival HST data from the CLASHES survey (chapter 2). We found that a threshold of 0.1 minimizes the number of false classifications of extendedness or compactness.

Figure 4.4 shows the selection features of known $3.75 < z < 4.5$ quasars against the distribution of background contaminants and delineates the cuts used in this initial criteria. As seen in the upper left-hand panel, two quasars are eliminated because they have an `extendedness` greater than 0.1. The `extendedness` cut removed 73% of all sources, but

[5]RA: 322.61 Dec: 1.1585 z:3.8847

removed only two out of 364 of the known quasars $3.5 < z < 4.5$ on Stripe 82.

4.4.3 Color

Flux in u, g, r, i and z are converted to magnitudes and corrected for extinction using E(B-V) values from (Schlegel et al. 1998) and $Rv = 3.1$.[6]

For this dataset, the intrinsic scatter in colors is much greater than photometric uncertainties in the colors. Colors are derived from the medians of ~ 80 observations and are 2 magnitudes deeper than the limit of this survey. As measured by the width of the stellar locus in principal colors w and x (chapter 3), at the faint magnitude limits of this survey, $i \sim 22$ ($i \sim 22.5$), the photometric errors are ~ 0.02 (0.04). Because of the high SNR of measured colors, we may filter out objects with stellar colors using simple color criteria. While methods have been developed to take into account photometric uncertainties in noisy colors near the detection limit (Richards et al. 2001; Kirkpatrick et al. 2011; Bovy et al. 2012), these methods are unnecessary when high-SNR colors are available, as they are in this study.

We adopted the following color criteria:

- $(g - r) > 0.95$
- $(r - i) < 0.3(g - r) + 0.03$
- $-0.7 < i - z < 1$
- $-1 < (r - i) < 0.8$
- $u > 22.5$ (including negative or flagged flux measurement)

These color criteria were motivated by the empirically observed colors of spectroscopically confirmed $z \sim 4$ quasars and were confirmed as reasonable by model spectra from McGreer

[6] All magnitudes presented in this work have been extinction corrected.

et al. (2013). The upper middle and right panels in figure 4.4, show the color distribution of $z \sim 4$ quasars (points colored per z) against the log-scaled density of point sources (grayscale). Three quasars are too close to the stellar locus and do not satisfy the criteria that $r - i < 0.3(g - r) + 0.03$. Three quasars are too blue in $g - r$ and do not satisfy $g - r > 0.95$. Finally, there are two point sources with $u < 22.5$

4.4.4 Variability Pre-Cut

For preliminary variability selection, we computed a large number of variability metrics (all of the computationally cheaper metrics, which excludes the CAR(1) modeling) for all 10 million sources having $i < 23.5$, as described in section 4.3. As described in detail in section 4.5.2, there were two non-redundant variability features computed at this stage that had power to discriminate between quasars and non-quasars (figure 4.7): the annually averaged r-band, χ^2 (`annAvg_chi2_clip_wm_r`), and the χ^2 of the annually averaged gri-combined lightcurves (`chi2griClippedWMean`). The annually averaged χ^2 tests variability on yearly timescales, and it was therefore not surprising that it had the most discriminating power; while both stars and quasars are variable, quasars vary on longer timescales. In the range $3.7 < z < 4.4$, the r-band performed better than the other single-band lightcurves.

We chose the following cuts empirically to preserve known quasars:

- r-band annually averaged $\chi^2 > 1$
- `chi2griClippedWMean` > 2

As seen in middle bottom panel of figure 4.4, only two quasars do not satisfy the variability criteria that `annAvg_chi2_clip_wm_r`> 2.

Overall, these initial cuts in morphology, color, and χ^2 eliminate 7% of *previously* known quasars in our redshift range, but 99.8% of the contaminants. We emphasize that this percentage is the number cut over the *previously known* objects.[7]

[7]Stating that the objects selected based on these criteria conformed to these criteria would be meaningless

4.4.5 Computing CAR(1) Parameters

A subset of 19,304 candidates fulfilled the color, extendedness, and χ^2 criteria described above. For this subset, we computed a set of variability metrics that are more computationally expensive and from these built a classifier that returns a probability that an object is a quasar or a quasar with redshift greater than 3.5.

For the 19,304 candidates that fulfilled the color cuts, we fit CAR(1) processes to the g-, r-, and i-band lightcurves using the method of Kozłowski et al. (2010) (see appendix of Kozłowski et al. (2010) for computational details). The inferred CAR(1) parameters include the exponential damping timescale τ, the short-term variability amplitude $\hat{\sigma}$, and the likelihoods (L^{best}, L^{noise}, and L^{∞}) of the CAR(1) process that had the best-fit τ, white noise ($\tau = 0$), or infinite timescale ($\tau = \infty$), respectively.

For each lightcurve, we used jackknife resampling to estimate the standard errors of the best-fit parameters and to check the robustness of those parameters. Specifically, if a lightcurve had N datapoints, we fit this lightcurve N times, each time leaving out one epoch. Section 4.9.1.2 provides a validation of these uncertainty estimates.

For each lightcurve, we then calculated the following quantities for each of the three bands:

- `log_arith_mean_tau` $= \log[\frac{1}{N}\sum_{i=0}^{N}\tau_i]$
- `log_geo_mean_tau` $= \frac{1}{N}\sum_{i=0}^{N}\log\tau_i$
- `std_tau` $= \sqrt{\frac{1}{N}\sum_{i=0}^{N}(\tau_i - \mu_\tau)^2}$
- `log_arith_mean_sigma` $= \log[\frac{1}{N}\sum_{i=0}^{N}\sigma_i]$
- `log_geo_mean_sigma` $= \frac{1}{N}\sum_{i=0}^{N}\log\sigma_i$
- `std_sigma` $= \sqrt{\frac{1}{N}\sum_{i=0}^{N}(\sigma_i - \mu_\sigma)^2}$

---

and circular. This project was iterative; the first iteration used $r-$band co-adds from an earlier data-processing run. Specifically, the two objects with higher extendedness appeared as point sources in the Winter2012 $r-$band co-adds.

- $\texttt{meanPlikePnoise} = \frac{1}{N}\sum_{i=0}^{N}\log(L_i^{best}/L_i^{noise})$
- $\texttt{stdPlikePnoise} = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(\log(L_i^{best}/L_i^{noise}) - \texttt{meanPlikePnoise})^2}$

We combined these quantities by taking the mean of the three bands.

The CAR(1)-fitting failed for one object. Visual inspection revealed that the source was contaminated by a nearby $i \sim 16$ star. With this source removed, there remained a set of 19,303 candidates with CAR(1) features.

In the next sections, we analyze how the training set, the subset of features, and the classifiers affected how completely and efficiently $z \sim 4$ quasars were selected from this subsample of 19,303 objects. We define the choices of training set, features, and classifiers before summarizing their effect on performance.

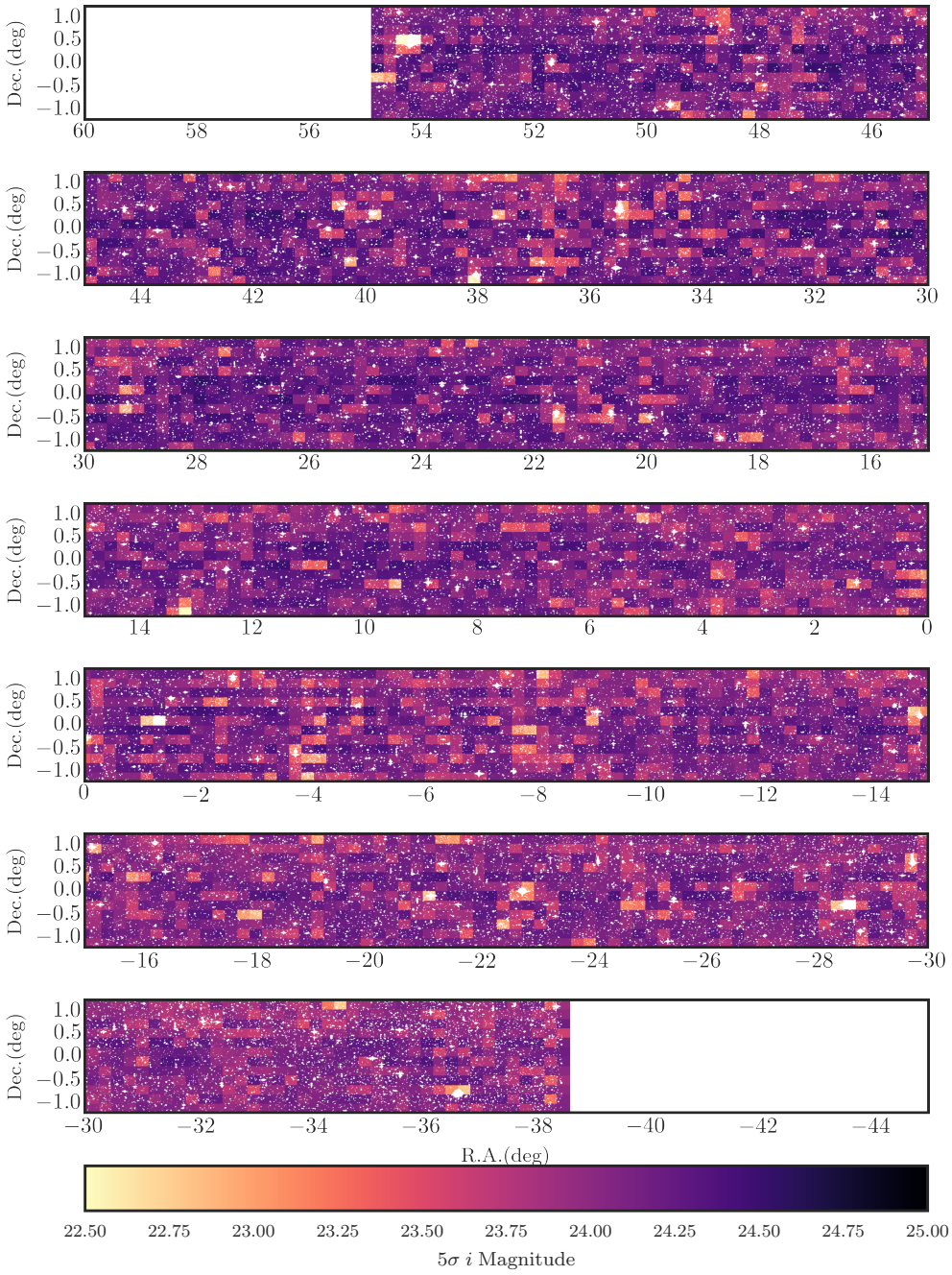

Figure 4.3 Survey area and 5σ depth: mean PSF magnitude of 5σ sources detected in our reprocessed i-band co-adds of Stripe 82. Colors represent the mean magnitude of sources having $4.9 < SNR < 5.1$, grouped by patch; this is an individual co-add image comprising the full mosaic. Masked regions in white are not accessible to the survey because of glints and bright stars. Because the 216 deg^2 unmasked area has a limiting magnitude deeper than $i = 22.5$, we treat the survey area as having a uniform depth.

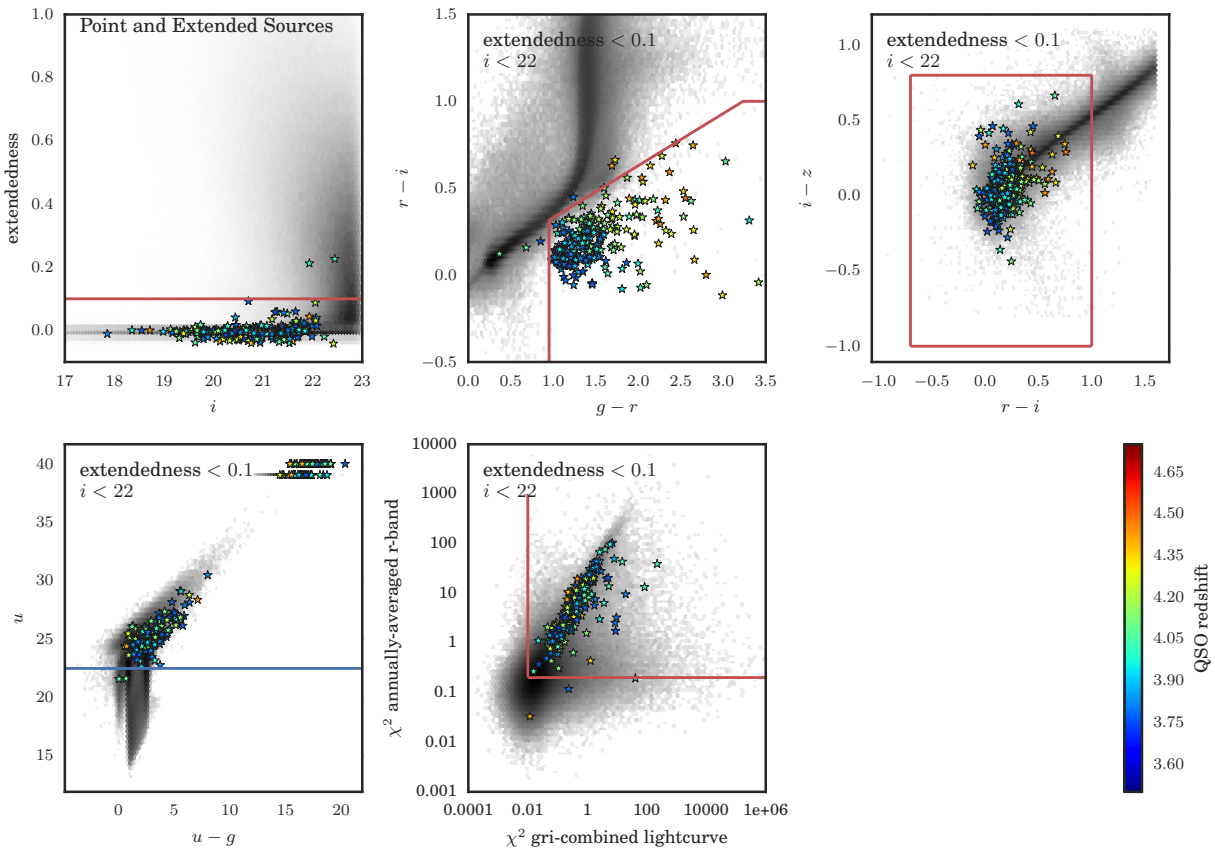

Figure 4.4 Colors, extendedness, and variability metrics for all spectroscopically confirmed quasars in survey footprint (shown as stars colored by redshift). The grayscale background density is the log-scaled density of point sources, except in the first panel, where it shows the linear-scaled extendedness density for all sources. While these initial cuts eliminate 7% of previously identified quasars in our redshift range, they also eliminate 99.8% of the contaminants.

4.5 Evaluation of Classification Factors

4.5.1 Training Set

To ensure that we had both ample data for training and training data that was similar to the data to be classified, we developed the training set in several steps. First, we compiled a catalog of spectroscopic redshifts. As explained in chapter 3, we compiled this catalog from both the spectroscopic observations presented in chapter 3 and published quasar catalogs that cover Stripe 82 and include quasars with redshifts greater than 3.5. These published catalogs included the SDSS I/II/III quasar catalogs from DR7, DR9, and DR12 (Schneider et al. 2010; Pâris et al. 2014); the $z \sim 6$ surveys of Jiang et al. (2006, 2008); the $z \sim 5$ survey of McGreer et al. (2013); the 2dF-SDSS LRG and QSO Survey (Croom et al. 2008, 2SLAQ); and the Stripe 82 variability survey of Palanque-Delabrouille et al. (2011). Then we cross-matched our point source catalog with the catalog of spectroscopic redshifts. The number counts of objects from these surveys that fell into this survey's footprint and redshift range are tabulated in table 4.2.

Some thought must also go into defining the labels (e.g. quasar vs. non-quasar, $z > 3.75$ vs. $z < 3.75$). There are two different approaches to compiling labeled lightcurves. One may assume that the number density of quasars is negligible compared to the number density of contaminants and use the entire point source catalog as the sample of contaminants. There are three orders of magnitude fewer $z \sim 4$ quasars than stars and galaxies per $\deg^2$, so this assumption is not unreasonable. However, we require instead that the contaminants in the training set also be spectroscopically confirmed. This limits the size of the available training set, because the number of spectroscopically confirmed stars and galaxies with these red colors are rare.

Redshifts for contaminants, such as stars and galaxies, were obtained from the SDSS DR12 spectroscopic catalog. We downloaded all 277,493 rows from the SkyServer `SpecObjAll` table overlapping our survey area with the following query:

```
SELECT specObjId, ra, dec, targetType, class,
```

```
                subclass, z_noqso, zErr_noqso,
                class_noqso, z, zErr
FROM dbo.SpecObjAll
WHERE dec between -1.26 and 1.26
 AND ((ra < 55) or (ra > 321.5))
 AND  TargetType == 'SCIENCE'
```

The table of contaminants was checked against the quasar catalog to ensure that no known quasars were in the contaminant set. Most stars in `specObjAll` were targeted as quasar candidates, and these can be erroneously labeled as stars by the automatic pipeline.

We cross-matched the 19,303 $z \sim 4$ quasar candidates to this quasar- and contaminant-combined spectroscopic catalog using a $1''$ matching radius. Of the 19,303 quasar candidates, 411 had spectroscopic redshifts (see table 4.2 for a summary). Of these redshifts, 46 were stars, galaxies, or low-redshift quasars ($z < 3.4$). 32 of these were from the `SpecObjAll` "SCIENCE" tables. These 32 redshifts from the automated SDSS pipeline were not reliable, so we visually inspected those 32 spectra. All were correctly identified except for two, J001220.48-010013.3 and J033709.3+002514.7; the SDSS pipeline classified these as $z \sim 1$ quasars, but they were in fact low-SNR galaxies. The remaining 364 spectra were quasars with $z > 3.4$ (360 having $z > 3.5$) 102 of these spectra were added by our spectroscopy program (chapter 3).

Because $N = 411$ is not sufficient to serve as a training set, we removed all initial selection criteria, including the variability cuts, except for two cuts, which we relaxed instead. We relaxed the `extendedness` threshold to 0.5, and relaxed the diagonal color cut in the $(g-r)$ and $(r-i)$ colors to $[(r-i) < 0.44(g-r)]$. Relaxing the cuts in this way provided a training set that added approximately 3300 stars, galaxies, and lower-redshift quasars, while also mitigating the biases that arise from making cuts in density distributions with steep gradients. For example, the quasar population decreases at higher redshift; therefore more bluer, lower-redshift quasars scatter redder to meet the $(g-r)$ criteria than vice versa. The

Table 4.2. Spectroscopic Redshifts for Final Cuts

Description	N	N: QSO	N: $z > 3.5$	Source
DR7QSO	97	97	97	Schneider et al. (2010)
DR9QSO	165	165	161	Pâris et al. (2012)
DR12Q	9	9	8	Pâris et al. (2014)
BOSS21	1	1	1	Palanque-Delabrouille et al. (2012)
JIANG6	3	3	3	Jiang et al. (2008) and Jiang et al. (2006)
SCIENCE	32	0	0	SDSS DR12
MMTO	74	68	66	This work
APO	17	13	13	This work
Gemini-North	13	11	11	This work
Total	411	367	360	

characteristics of the resulting training set are shown in figure 4.5 with the two cuts that were applied to the training set.

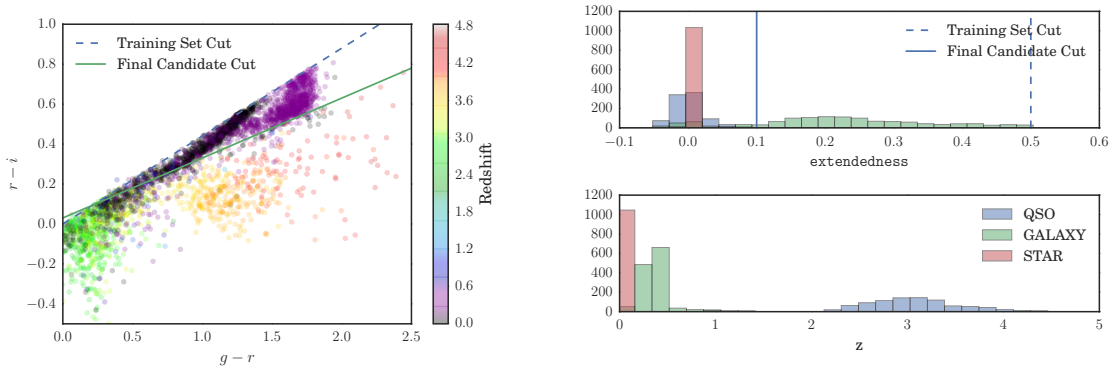

Figure 4.5 (left) Colors of the training set. The relaxed color cut allows into the training set some of the stellar locus (black points), some of the red sequence (purple points), and lower-redshift quasars. (right) Redshift and `extendedness` distribution of the three classes in the training set. By relaxing the `extendedness` cut, we allow in more galaxies.

We tested two different label definitions:

- `if z > 1`, which splits the training set into quasars and non-quasars, and
- `if z > 3.5`, which splits the training set into high-z quasars and everything else.

Because it is impossible to distinguish between high-z and low-z quasars without color information, we also investigate the use of a training set limited to rows with $g - r > 0.95$, a limitation intended to cut out the low-z quasars.

We further limited our training set to those objects with $g - r > 0.95$—the same $g - r$ cut applied to the candidates to eliminate lower redshift quasars. We did this because we found that most incorrect classifications were $z \sim 3.4$ quasars being classified as $z > 3.5$, and $z \sim 3.6$ quasars being classified as $z < 3.5$. In short, the scatter in the $(g - r)$ color, with which the redshift covaries, was the primary source of incorrect classification. Pre-cutting on color made the training data more similar to the data to be classified.

This resulted in 3 training set types which we call "$z > 1$," "$z > 3.5$," and "$z > 3.5; gr > 0.95$." Classifier performance is computed for all three scenarios.

4.5.2 Feature Selection

After building a training set with 3700 rows (quasars and contaminants) and 200 columns (features), we select the subset of features that provide the best classification performance. These feature selections need to be made for several reasons. First, many of the features are correlated and redundant. We calculated many of the same quantities using slightly different equations, with the intention of evaluating them at this feature selection stage. Note that empirical correlations can result from either intrinsic correlations of the sources (e.g. stars exhibit a correlation in $g-r$ and $r-i$ because of the stellar main sequence) or mathematical correlations of the metrics (e.g. the geometric and arithmetic means of jackknifed τ estimates, or the sigma-clipped and non-sigma-clipped χ^2). Second, we needed to model the variability metrics of known quasars in order to simulate variability metrics with which to estimate the selection function (Section 5.2.5.2). We only have 300 $3.75 < z < 4.5$ quasars upon which to model the empirical variability distributions. To model of the distribution in feature space, the number of data points must be much greater than the number of dimensions. Three hundred data points in a 200-dimensional space is an extremely sparse dataset, and modeling a 200-dimensional distribution is impossible with only 300 rows. Therefore, we reduced the number of features using feature selection methods.

The concept of feature selection is quite simple in theory, but in practice it comes up against computational limits. Thus, a wide body of machine learning literature focuses on tractable feature selection. The simplest brute-force feature selection algorithm relies on scoring each classifier (e.g. using cross-validation) and choosing the combination of features that score the best. To illustrate, consider a training set with 200 features, from which you want to choose the best 20 for a classifier. The brute-force method would compute the score for every 200 choose 20, $C(200, 20) \sim 10^{27}$, combination of features. Clearly, this is too computationally expensive. Some solve this problem by using greedy algorithms. For

example, the algorithm would score all 200 single-feature classifiers and choose the best. Next, it would score all 199 two-feature classifiers, locking in the feature selected from the first round, and choose the best. Next, it would score all 198 three-feature classifiers, etc. It would continue this iterative selection process until it was down to 20. In other words, you iteratively add one feature at a time; in this example, the entire process would require only 3810 trials. The same can also be done in reverse, by iteratively *removing* one feature at a time. These greedy algorithms, however, do not guarantee the selection of the best 20 features, and provide suboptimal solutions.

Another line of research has developed methods for assessing the discriminating power of features using information theory. These methods capture the discriminating power of each variable, but are classifier-independent, making them more efficient than brute-force scoring of classifiers. These information theory–based methods compare the mutual information between the features of the training set an the outcome variable (class or redshift in this application).

Information theory is based on the concept of entropy, which provides a measure of the uncertainty or impurity in a given distribution. Entropy of a distribution $p(x)$ is defined as

$$H(p) = -\sum_{x \in \Omega} p(x) \log_2 p(x), \tag{4.1}$$

where Ω is the discrete set of outcomes. Let X be a random variable with distribution $p(x)$. The conditional entropy $H(X|Y)$ is the entropy of X given that the value of another random variable Y is known. Conditional entropy of X given Y with a joint distribution p_{XY} is defined as

$$H(X|Y) = \sum_{y} p_Y(y) \sum_{x} H(p_{X|Y=y}). \tag{4.2}$$

The conditional entropy $H(X|Y)$ is always less than or equal to $H(X)$. Observing another variable, Y, decreases the uncertainty in X, and the amount of this decrease is called the mutual information between X and Y. The mutual information

$$I(X||Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{4.3}$$

is also symmetric. Mutual information is sometimes called "information gain" and denoted $IG(X, Y)$.[8] These definitions for discrete variables can be applied to continuous variables as well by replacing the summations with integrals. In practice, this integral is difficult to estimate; the estimation is usually done by binning the data. (There are other methods for estimating the continuous case, including the kNN estimation method (Ross 2014).) The problem with binning, however, is that the mutual information depends heavily on the bins chosen. One way around this is to try a variety of and bins and choose the binning that yields the maximum mutual information. This quantity is called the maximal information coefficient (MIC).

MIC can be used in the simplest application of information gain (IG) to feature selection. The simplest application is to compute the information gain from splitting the sample into two subsamples based on the value of a feature X. For example, this split is computed when adding each branch while building a decision tree. For a single split into two subsamples based on one feature, a.k.a. a "decision stump," the information gain, $IG(X, Y) = H(Y) - p_{Y=1} \sum_x H(p_{X|Y=1}) - p_{Y=0} \sum_x H(p_{X|Y=0})$, is the difference between the entropy of the parent sample and the the average entropy of the child samples of the outcome variable Y (e.g. class). For example, in one limiting case, if splitting the sample into two subsamples results in one subsample made up of all quasars and the other made up of all stars and galaxies, the entropy of the children is zero, and the information gain is equal to the entropy of the parent sample. If, on the other hand, the two subsamples each have the same fraction of quasars and non-quasars as the parent sample, then the information gain was zero.

As shown in figure 4.7, we computed the information gain for each feature, in this way, for each of the three training sets. For each feature X, we iterated through values on which to split the parent sample into two, and chose the split with the maximum IG, or the highest entropy reduction. In general, the single features with the highest information gain have the highest discriminating power and appear at the top of the charts. The features with the

[8]Nomenclature in the literature is not consistent on whether information gain *is* mutual information or a particular computation of mutual information performed by splitting the sample into two subsamples.

lowest information gain have the lowest discriminating power; features with zero information gain are not useful for classifying quasars.

The single features that perform the best (have the highest information gain) differ from training set to training set. For the training set with labels marked "true if $z > 1$" (equivalent to a quasar vs. non-quasar split), it is apparent that variability information is more important than color information; this is because quasars change colors depending on redshift, but maintain similar variability characteristics. The DRW τ parameter provides the best discriminating power of all the features tested. This finding confirms that the same finding at low redshift by MacLeod et al. (2011) also holds at high redshift, and justifies the expense of computing these DRW parameters. The fourth most powerful feature, not counting duplicates, is also from DRW fitting: `meanPlikePnoise`, the $\log(L^{best}/L^{noise})$. This feature, the relative likelihood that the lightcurve is DRW, is even more powerful in conjunction with the estimates of τ, as L^{best} can be high for even short τ.

Of the computationally cheaper metrics computed on the full 10 million lightcurves, two features stood out: the annually averaged r-band, χ^2 (`annAvg_chi2_clip_wm_r`), and the χ^2 of the annually averaged gri-combined lightcurves (`chi2griClippedWMean`). This finding motivated the initial variability cut (§4.4.4).

The g, r, i combined annually averaged lightcurve provides more information gain than the r, g, and i do individually in that order at $z \sim 4$. The u and z bands offer nearly no information gain. In general, quasars' variability amplitude is greater in the bluer bands, but at $z \sim 4$ the g-band flux is eaten away by the Lyα forest at this redshift. It comes as no surprise that r is more powerful than i because it probes a bluer part of the continuum; much of the r-band flux, however, is from the Lyα broad emission line. It is thought that the flux variation arises from the continuum (thermal radiation from the accretion disk) rather than the emission lines (from the broad-line region). Despite the stabilizing effect of the emission line, the r-band is still the most discriminating single band.

As seen in the upper right panel of figure 4.7, for the training set with labels marked "true if $z > 3.5$," color becomes important, because we want to distinguish not only between

quasars and non-quasars, but also between low-z quasars and high-z quasars. Colors, and not variability, are sensitive to redshift.

Interestingly, the best two features in this training set, $(u-g)$ and `sigma_intrinsic_u` ($\sigma_{0,u}$), sense the high-z quasars dropping out of the u-band imaging. The $\sigma_{0,u}$ is really a proxy for the high variance in these non-detections. The RMS of the forced photometry measurements at the positions of $z \sim 4$ quasars is $10x$ higher than would be expected based on the variance in the pixels. When we use a training sample that is pre-cut on $g - r > 0.95$ (like the color criteria applied to the $z \sim 4$ quasar candidates), these u-band drop-out features fall in importance (lower panel). Similarly, `std_tau_gri`, which is proportional to the uncertainty on the best-fit DRW damping timescales, is higher for $z \sim 4$ quasars. This also is not physical, but a result of a high scatter in best-fit timescales in the jackknife-resampled lightcurves for lightcurves with long timescales. These indirect features that result of measurement techniques pose an interesting problem: Should non-physical features be included in the classification? Although they offer discriminating power, they are difficult to model. For this science application we decided to include them because the selection function is based on the empirical distribution of observed features in known quasars (§5.2.5.3. Because we do not use a physical model, these features can be simulated like any other.

The information gain computed this way (figure 4.7) does not take into account correlations between the features. Since many features are different calculations of the roughly the same information, we do not want to take the (e.g.) 10 best *single* features; we want the set of 10 best *complementary* features. The mutual information *between feature variables* can be used to remove redundant features. Correlation coefficients offer a very limited measure of dependence between two variables. For example, if variable $X2$ has "U-shaped" dependence on $X1$, the correlation coefficient will be zero despite the fact that $X2$ is a deterministic function of $X1$. Mutual information provides a more robust measure of one variable's ability to predict another.

Figure 4.8 shows the mutual information for each feature pair and redshift. We compute

the mutual information between continuous variables using the kNN method introduced by Ross (2014). Features that were highly correlated also have high MI, but unlike correlation coefficients, the mutual information is always positive.

We performed feature selection by adding features one-by-one using a greedy algorithm. Each candidate feature was scored based on its mutual information with "class" (the label, or feature to predict) minus its average mutual information with the features that were previously selected (Mutual Information Feature Selection: Battiti 1994). With each iteration we selected the feature from the i remaining features which had the maximum $I(X_i; Y) - \beta \sum_j I(X_i; Y_j)$, where β is an input parameter weighting the mutual information of the already selected j features. The features selected for the $z > 1$ training set were:

1. `sum_log_geo_mean_tau_gri`
2. `ri`
3. `sum_meanPlikePnoise_gri`
4. `std_tau_gri`
5. `chi2griClippedWMean`
6. `annAvg_chi2_clip_wm_r`
7. `annAvg_chi2_clip_wm_g`
8. `log_geo_mean_sigma_gri`
9. `log_arith_mean_tau_gri`
10. `chi2griClippedWMeanDivideMean`
11. `sigmaClipped_chi2`
12. `gr`
13. `extendedness`
14. `meanPlikePnoise_gri`
15. `log_geo_mean_tau_gri`

in that order. The features selected for the $z > 3.5; g - r > 0.95$ training set were

1. `ri`
2. `sum_log_geo_mean_tau_gri`
3. `std_tau_gri`
4. `iz`
5. `ug`
6. `sum_meanPlikePnoise_gri`
7. `sigma_intrinsic_u`
8. `chi2griWMean`

9. log_arith_mean_tau_gri
10. extendedness
11. annAvg_chi2_clip_wm_r
12. log_geo_mean_sigma_gri
13. chi2griClippedWMeanDivideMean
14. sigmaClipped_chi2
15. log_geo_mean_tau_gri

These lists show that features were being selected that were known to be redundant to previously selected features. Ultimately, we used our domain knowledge of the problem and how features were computed to manually remove these redundant features. We replaced the redundant measures of τ (which appeared three times) and χ^2_{gri} (twice) with sigma_intrinsic_g and sigma_intrinsic_i, and $g - r$. We intentionally left in, despite their correlation, both of the two annually averaged χ^2 metrics (chi2griClippedWMean and sigmaClipped_chi2), because we needed to simulate these quantities to estimate the selection function's response to the cuts made previously.

We defined the fiducial 15-feature set as

1. gr
2. ug
3. ri
4. iz
5. sigma_intrinsic_u
6. sigma_intrinsic_g
7. sigma_intrinsic_i
8. chi2griClippedWMean
9. sigmaClipped_chi2
10. annAvg_chi2_clip_wm_r
11. log_geo_mean_tau_gri
12. log_geo_mean_sigma_gri
13. std_tau_gri
14. meanPlikePnoise_gri
15. extendedness

We defined reduced 9-feature set as

1. gr

2. `ri`
3. `iz`
4. `sigma_intrinsic_g`
5. `chi2griClippedWMean`
6. `annAvg_chi2_clip_wm_r`
7. `log_geo_mean_tau_gri`
8. `log_geo_mean_sigma_gri`
9. `meanPlikePnoise_gri`

4.5.3 Classification Algorithms

With the training set pared down to 15 features, we trained and cross-validated two ensemble classifiers: Random Forest and Adaboost. We used implementations from `Scikit-learn` (Pedregosa et al. 2011). We configured Adaboost to iterate until convergence or a maximum of 300 iterations. We configured the Random Forest classifier to average the votes of 50 decision trees. In building each tree, at each node $\sqrt{N_{features}}$ were tested, using information gain to measure the quality of a split. Nodes were added until the leaves were pure.

To assess the performance, we cross-validated both classification algorithms on each combination of training-set label definitions and each of the 15- and 9- feature set. We trained each classifier 50 times, holding back a random 30% of the training each time to assess the accuracy.

The classifier performance is easily visualized using the receiver operating characteristic (ROC) curve. The ROC curve maps the true positive rate as a function of the false positive rate, stepping through the probability threshold as the classification boundary; a perfect classifier will have a true positive rate of 1 and a false positive rate of 0. The ROC curve can be reduced to a single value representing the classifier's power: the area under this curve

(AUC). A perfect classifier would have an AUC of 1 and a completely random classifier would have an AUC of 0.5. If the classifier were any worse than random, simply flipping the predictions would give better than random predictions.

The AUC provides a more informative measure of a classifier's performance than a single value for completeness and purity. Completeness is achieved at the expense of purity. The requirements for completeness and purity depend on the details of the science application.

Figure 4.9 shows the ROC curves of the cross-validation of the 15-feature classifier training set. Each line is the mean of the 50 trials, and the shaded regions show the minimum and maximum of the 50 trials. The darker shading shows the range between the 32nd and 68th percentile, but is almost too narrow to see. At a 1% false positive rate, we achieve an average completeness of 98%.

As seen in figure 4.9, Random Forest and Adaboost classifiers perform similarly for the 15-feature set. The 15-feature Adaboost (Random Forest) classifier produces a mean AUC of 0.9978 (0.9977). The 9-feature Adaboost (Random Forest) classifier produces a mean AUC of 0.9972 (0.9977). Interestingly, the Random Forest classifier did not degrade in performance with fewer features. For this reason, and because Random Forest is already a familiar classifier in the astronomy community, we chose the Random Forest classifier for the selection of the final luminosity function sample.

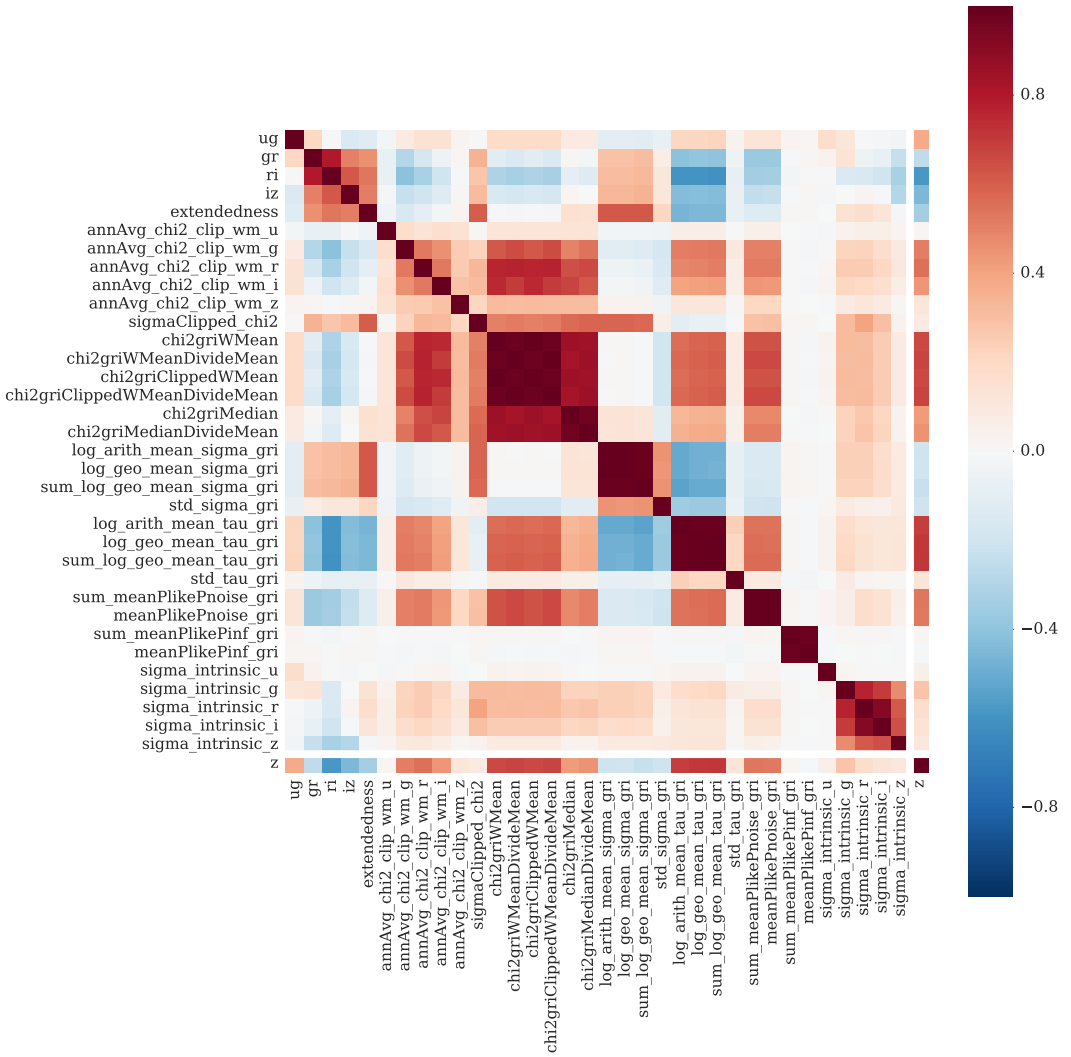

Figure 4.6 Pairwise correlation of computed features for the 3700-row training set of quasars, stars, and galaxies. Features are correlated because of similarity of metrics (e.g. means vs. medians in `chi2griWMean` and `chi2griMedian`) and because of intrinsic correlations in the distribution of sources (e.g. tight correlation of $(g-r)$ and $(r-i)$ in the stellar locus). Blocks of redundant metrics are apparent in the `chi2gri[*]` metrics (the χ^2 of the gri-combined lightcurves differs based on how the lightcurves were averaged) and the `[*]sigma_gri`, and `[*]tau_gri` metrics (these differ in how the jackknife estimates of $\hat{\sigma}_{DRW}$ and τ were combined and whether the sum or average was taken of the parameter estimates on the g, r, and i lightcurves). The correlation coefficients with redshift z are also shown as a crude summary of discriminating power, as redshift is a proxy for class.

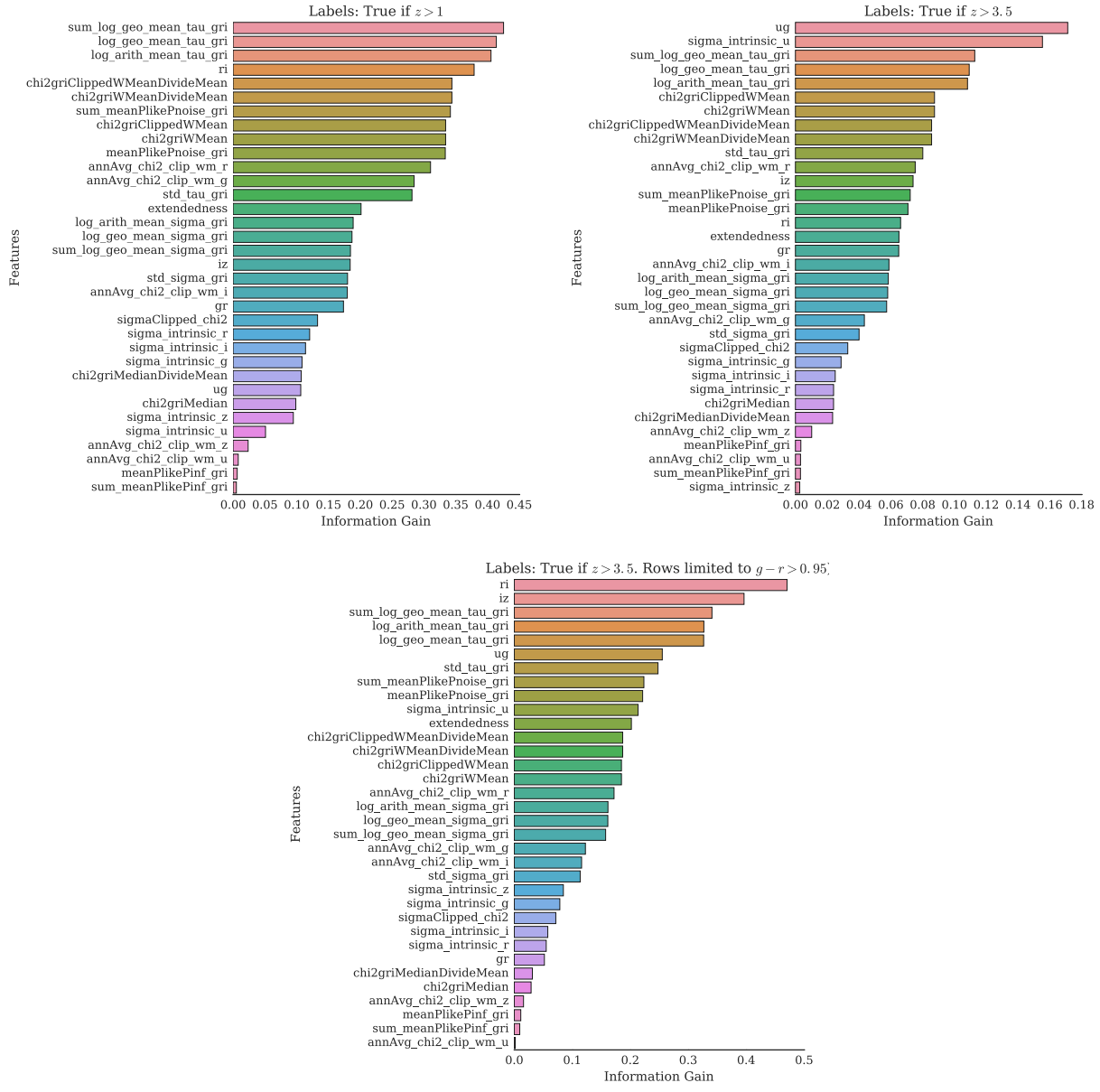

Figure 4.7 Independent information gain of selected features for three training set label definitions. Features with the highest information gain have the highest discrimination power. For discriminating between quasars and non-quasars (the upper left panel), the best feature is the τ from the DRW fits. When differentiating between low redshift and high redshift, color information becomes important.

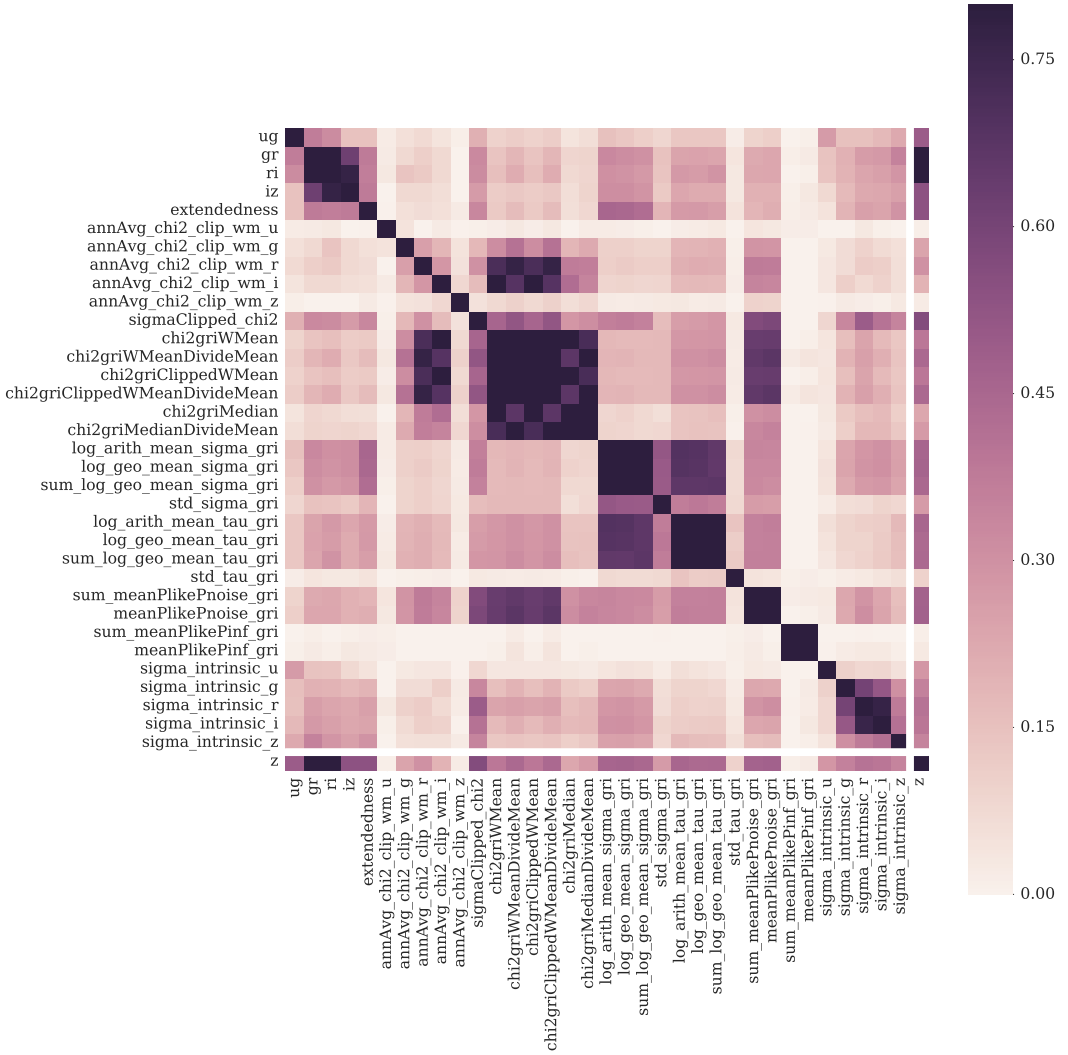

Figure 4.8 Pairwise mutual information of computed features.

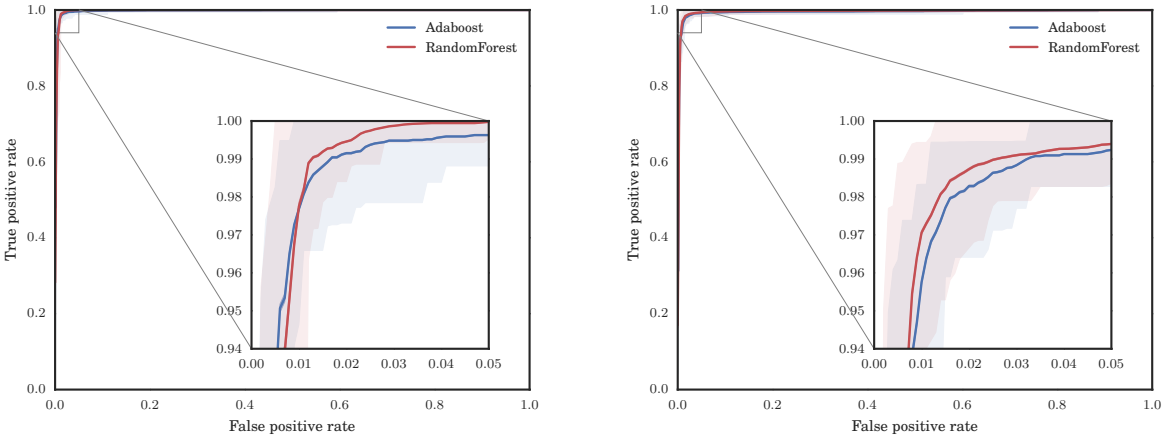

Figure 4.9 The average ROC curve for the cross-validation of the Adaboost and Random Forest 15-feature classifiers (left) and 9-feature classifiers (right). Dark lines show the mean of 50 trials, in which we trained on 70% of the training set and tested on 30%. The lightly shaded region shows the the range between the minimum and the maximum of the trials. These almost match the performance of an ideal classifier, which would have an AUC of 1. The 15-feature Adaboost (Random Forest) classifier produces a mean AUC of 0.9978 (0.9977). The 9-feature Adaboost (Random Forest) classifier produces a mean AUC of 0.9972 (0.9977).

4.6 Results

4.6.1 Summary of Classification Performance

Figure 4.10 presents the summary of the AUC performance metric for the combinations of training set, feature sets, and algorithms described above (§4.5). For all combinations, using both colors and variability performed better than using just colors alone or variability alone.

Both training set label definition and feature selection are important and inter-dependent. When selecting between quasars and non-quasars (left hand column), variability performed better than colors alone. When selecting only high-z quasars with the same training set (middle panel), colors alone performed better than variability alone (the AUC= 0.967 ± 0.001 was plotted, but it was below the y-limit), because colors carry the all redshift information. The intrinsic scatter in quasar colors at $z \sim 4$ is large, and most misclassifications for the $z > 3.5$ labeled classes are $z \sim 3.4$ false positives and $z \sim 3.6$ false negatives.

The AUC did not decrease when reducing the number of features from 34 to 15, indicating that the 15 selected features captured nearly all information available. Reducing the number of features from 15 to 9 increased the advantage of Random Forest over Adaboost. For the 34/15 feature sets, the performance of the two classification algorithms was comparable within the error bars when both colors and variability were used. This suggests that Random Forest performs better in scenarios with insufficient feature sets or inappropriate training sets.

4.6.2 Classification for $z \sim 4$ Luminosity Function

As a result of the analysis presented, we used the following classification factors for the luminosity function (chapter 5):

1. 15-feature set with variability metrics, colors, and extendedness

2. Training set with initial cut on $g - r > 0.95$ and labeled "true" if $z > 3.5$.

3. Random Forest algorithm

After pre-selecting the 10 million lightcurves on color, morphology and simple annually averaged χ^2 criteria, we classify the 19,303 remaining lighcurves, yielding a statistical sample of $z \sim 4$ quasars. The completeness analysis is presented in Chapter 5.

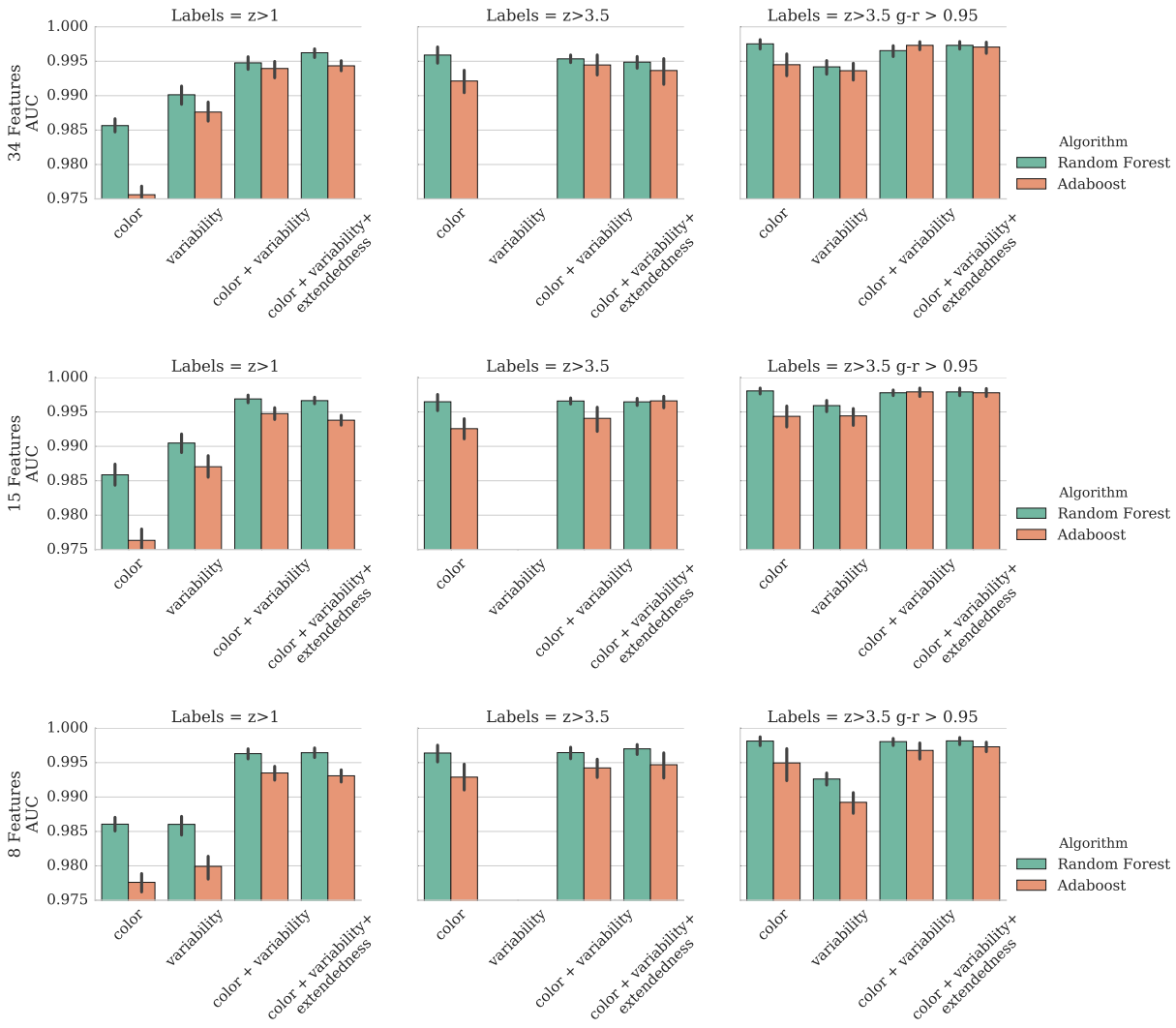

Figure 4.10 Summary of AUC results for the cross validated training sets, feature subsets, and classification algorithms. Classifiers using colors and variability together are more accurate than one or the other in all cases. Color is necessary to separate high-z from low-z quasars; therefore, variability-only classifiers performed poorly unless the low-z quasars were removed from the training set. The fiducial classier (in the middle row on the far right) uses the 15 best features, including color, variability, and extendedness, and a training set defining labels as true if $z > 3.5$, and training only on rows with $g - r > 0.95$. For this best combination, the classification algorithms have identical accuracy.

4.7 Discussion

Most of the effort in this study was spent on feature engineering, which will likely always involve some manual process. While there are constant claims that new parts of the machine learning classification pipeline have been automated, feature engineering will likely never be fully automated, because transforming observations into a series of discriminating metrics requires domain knowledge. For example, because we know that quasars vary on month-long timescales, we also know that it is worth computing a metric that measures lightcurve variance on month-long timescales or measures the characteristic timescale of variation.

Similarly, human intervention will likely always be needed in the selection of statistical samples of astronomical objects; ensemble classifiers will, like feature engineering, probably never be fully automated, because a deep understanding of the selection function is necessary. However, ensemble classifiers are becoming commonplace in the astronomical community, and applying them is as simple as `import sklearn`—sometimes without the understanding needed to apply them well. But astronomers are not trying to identify trends or generate recommendations (tasks for which fully-automated ensemble classifiers can be used successfully); instead, we are trying to infer real physical quantities and the uncertainties of those inferences. Good classifiers are of no use without a clear estimate of the selection functions. Understanding the completeness of the selection method is more important than high completeness in itself.

Ideally, we would estimate completeness using models of the phenomena we are trying to observe. In quasar selection, we have good models of quasar colors from which we can simulate observations and "classify," but we do not yet have good models of variability against which to check selection. A definitive explanation for why quasars vary still eludes us. Looking forward, estimating the selection function of variability-selected objects in LSST would benefit from high-fidelity simulated *time-varying spectra*. Creating these simulated lightcurves would require understanding all known correlations between emission lines, slopes, and broad absorption lines, as they vary.

In the meantime, we have worked around the lack of understanding with the collection of more data. Whereas one sample is not sufficient to estimate classification completeness, two or more samples, obtained by orthogonal selection methods, can be used to validate each other. This is the approach we take here. By combining quasars selected from multiple selection methods used in previous surveys, we can assume that the observed distribution of quasars in feature space is an accurate representation of the true distribution and feature space.

4.8 Conclusion

We have used the new Stripe 82 lightcurves to build a catalog of colors and u, g, r, i, z, gri variability metrics for 10 million sources $i < 23.5$. We used this resulting catalog, in combination with new and existing spectroscopy, to develop a $z \sim 4$ quasar classifier. We find that

- simultaneous classification based on colors and variability performs better than classification based on either color or variability alone,
- at a 1% false positive rate, we achieve a sample completeness of 98%, sufficient for a estimating a photometric luminosity function, and
- fainter $z \sim 4$ quasars follow the same variability trends as lower-redshift and brighter quasars do.

Ultimately, our classification method produces a sufficiently pure sample with which to estimate statistical quantities such as luminosity functions and clustering measurements. However, the selection method alone is not useful without a detailed estimate of the selection function. The next chapter presents methods for estimating the selection function for variability-based classifiers. We quantify the selection function and discuss the sample completeness therein.

4.9 Supplementary Information

4.9.1 Validating Variability Estimators

Before computing the variability metrics, we evaluated estimators for variability metrics using simulated lightcurves, which were created using two different models. The first model was a simple normal distribution in which, at each epoch, the observation was drawn from a normal distribution with a mean and an intrinsic sigma, with added realistic noise. The second model is a CAR(1), or damped random walk (DRW), parameterized by the exponential damping timescale τ and the short-term variability amplitude $\hat{\sigma}$ (figure 4.11).

We added noise to the simulated lightcurves by drawing from a model of the flux uncertainties apparent in the forced photometry. Flux uncertainties, σ_e, are flux dependent. Using the empirical distribution of measurement uncertainties from the forced photometry, we modeled the relationship between the measured magnitude, r (in this example, we call the magnitude r, but the band does not matter in simulated lightcurves), the expected magnitude uncertainty, $\mathrm{E}\sigma_e(r)$, and the variance in magnitude uncertainty, $\mathrm{Var}[\sigma_e](r)$. For each flux bin, we took the median of the r-band measurement errors as an estimate for $\mathrm{E}\sigma_e(r)$ and σ_G^2 as an estimate for $\mathrm{Var}[\sigma_e](r)$, as shown in figure 4.12. To evaluate the expected flux uncertainty at a particular magnitude r, the median(σ_r) and $\sigma_G(\sigma_r)$ were interpolated to find a distribution of uncertainty at that magnitude. To simulate measurement error for each simulated point in the lightcurve, we then drew from the interpolated $\sigma_{ei} \sim \mathcal{N}(\mathrm{median}(\sigma_e), \sigma_G^2(\sigma_e))$. The simulated datapoint was then perturbed by this uncertainty drawn from $\mathcal{N}(0, \sigma_{ei})$.

4.9.1.1 Intrinsic Sigma

Variance in the flux measurements of a lightcurve arise from two sources, the intrinsic variability of the source σ_0^2 and the variance in the measurement, σ_e^2. The total variance is a sum of the two:

$$\sigma_{tot}^2 = \sigma_0^2 + \sigma_e^2 \tag{4.4}$$

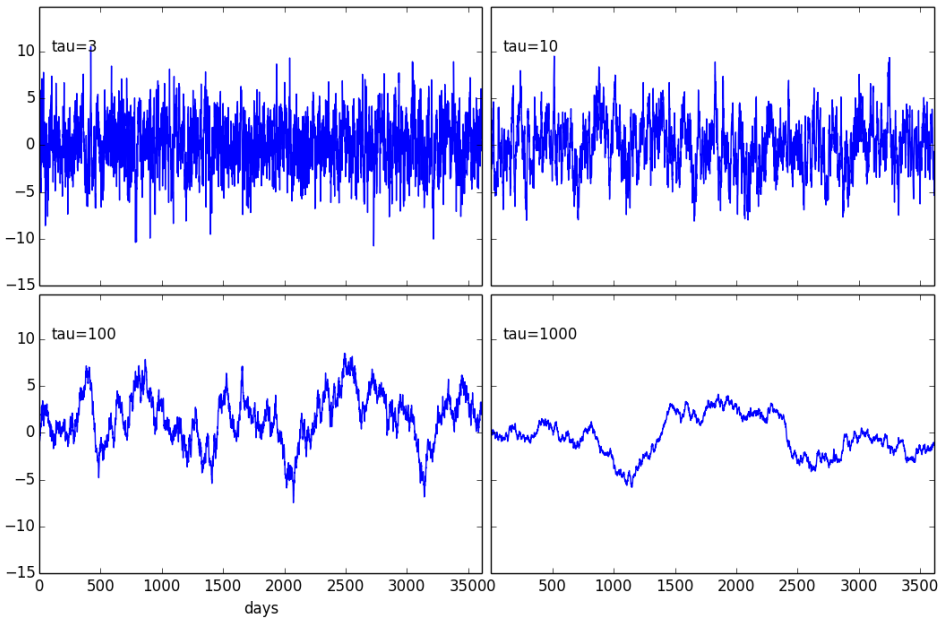

Figure 4.11 Example well-sampled lightcurves simulated using a CAR(1), or damped random walk, model that parameterized by the exponential damping timescale τ and the short-term variability amplitude $\hat{\sigma}$. The panels show varying τ while fixing overall standard deviation in the flux $\sigma = 1$. The long-term σ is related the the short-term $\hat{\sigma}$ by $\hat{\sigma} = \sqrt{2\sigma^2/\tau}$. These lightcurves do not include measurement noise.

σ_0^2 is a property of the source that we would like to estimate. While $\sigma_0^2 = \sigma_{tot}^2 - \sigma_e^2$, σ_{tot}^2 and σ_e^2 are not known exactly and themselves need to be estimated. The presence of outliers requires the use of robust estimators. We evaluated the following five estimators for σ_0:

1. s (simple rms of observations)
2. $\sqrt{\zeta^2\sigma_G^2 - e_{50}^2}$ (p210 of Ivezić et al. (2014))
3. $\sqrt{\sigma_G^2 - e_{50}^2}$ (used in this study)
4. $\sqrt{s^2 - \langle\xi(m)\rangle^2}$ (Sesar et al. 2007)

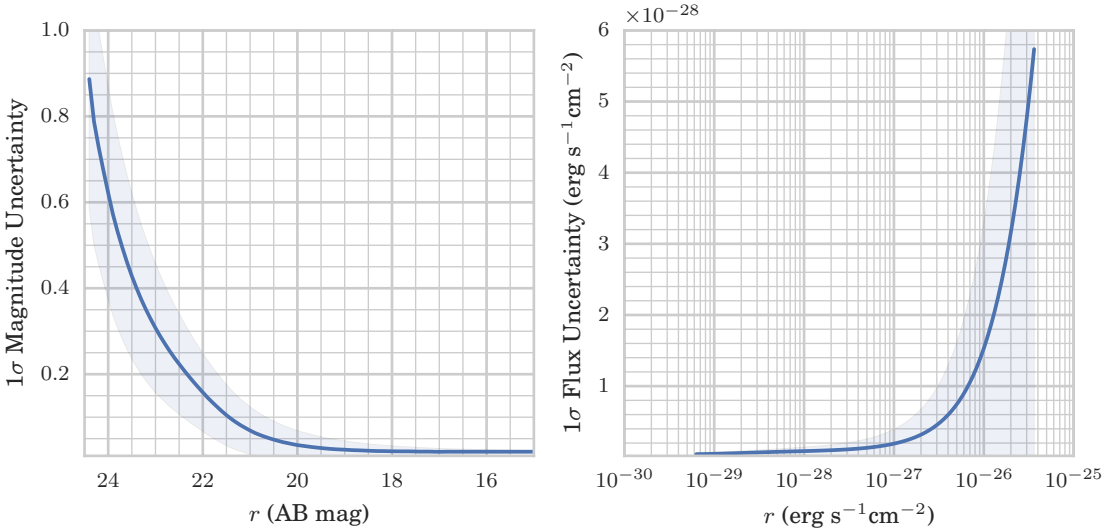

Figure 4.12 Flux uncertainty model for the forced PSF photometry measurements.

5. $\sqrt{\Sigma_3^2 - e_{50}^2}$,

where Σ represents a sigma-clipped RMS (wherein only points within $3\sigma_G$ of the median were used), $\zeta^2 = \text{median}(\tilde{\sigma})/\text{mean}(\tilde{\sigma})$ where $\tilde{\sigma}^2 = \sigma_G + \sigma_e^2 - e_{50}^2$, $e_{50} = \text{median}(\sigma_e)$, and $\sigma_G^2 = 0.7413(q_{75} - q_{25})$.

To validate intrinsic sigma estimators, we simulate 100 Gaussian lightcurves, each with 70 epochs, at four mean magnitudes, $r = \{21,\ 22,\ 23,\ 24\}$, at 19 intrinsic sigmas, $0.005 < \sigma_0 < 0.5$. Figure 4.14 shows the degradation of these estimators with increasing magnitude. The estimator $\hat{\sigma_0} = \sqrt{\sigma_G^2 - e_{50}^2}$ performs the best in terms of accuracy and robustness for all magnitudes and σ_0s tested, and it was therefore chosen as the estimator for intrinsic sigma. $\hat{\sigma_0} = \sqrt{\Sigma_3^2 - e_{50}^2}$ performed well in terms of accuracy, but was very sensitive to the sigma-clipping threshold chosen. When a higher threshold was chosen, the estimate moved higher; when a lower threshold was chosen, it moved lower. For reference, $z > 3.5$ quasars have a mean $\sigma_0(r) = 0.11$ (figure 4.13), with a scatter of 0.06, which means that the measurement error surpasses the intrinsic sigma for $r > 21.6$.

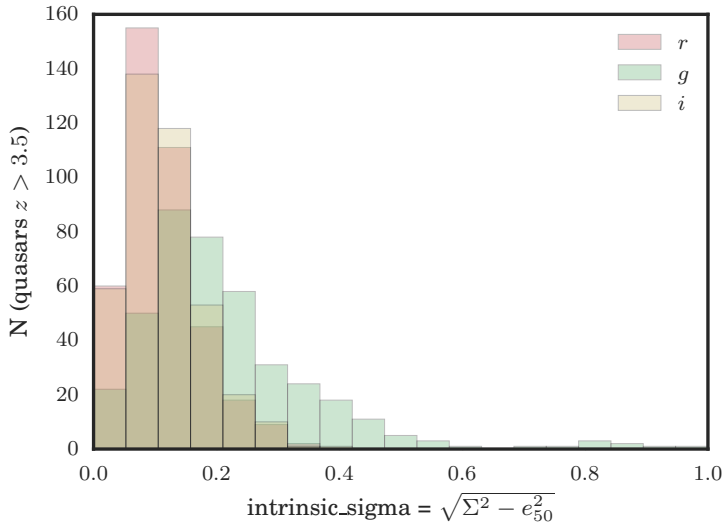

Figure 4.13 Estimated σ_0 for known $z \sim 4$ quasars in g, r, and i to guide interpretation of figure 4.14

4.9.1.2 CAR(1) Parameters

As introduced in section 4.4.5, we fitted CAR(1) processes to the g-, r-, and i-band lightcurves using the method of Kozłowski et al. (2010) (this method is based on the approach of Press et al. (1992); see the appendix of Kozłowski et al. (2010) for computational details). The inferred CAR(1) parameters include the exponential damping timescale τ, the short-term variability amplitude $\hat{\sigma}$, and the likelihoods that the lightcurve is a CAR(1) process, white noise, or has an infinite timescale.

For each lightcurve, we used jackknife resampling to estimate the standard errors of the best-fit parameters and to check the robustness of those parameters. Specifically, if a lightcurve had N datapoints, we fit this lightcurve N times, each time leaving out one epoch.

To assess the method of calculating uncertainties on these parameters, we fit simulated CAR(1) lightcurves with known power spectral densities. The simulated lightcurves were generated as follows. For each combination of mean-magnitude $r \in \{20,\ 22,\}$, $\tau \in$

$\{10,\ 30,\ 100\}$, and $\sigma \in \{0.1,\ 0.3,\ 1.0\}$ we generated 100 CAR(1) lightcurves in magnitude space sampled with a uniform cadence of cad $\in \{1,\ 3,\ 5,\ 10,\ 30\}$ days. We then fit the lightcurves with a CAR(1) model, estimating the τ and $\hat{\sigma}$ using the method of Kozłowski et al. (2010).

To test if the standard errors estimated from the jackknife resampling were accurate, we compared the estimated uncertainties from the jackknife resampling to the observed RMS in the 100 instances of simulated lightcurve parameters. The parameters estimated for each of the 100 lightcurves theoretically should be normally distributed, with a mean of zero and standard deviation equal to the mean estimate of standard error.

We found that the accuracy of the estimate depends on both the cadence versus τ and the magnitude (a proxy for SNR) versus $\hat{\sigma}$. Figure 4.15 shows the recovery of timescales from these simulated lightcurves for an input $\tau = 30$. Plots show the histograminfer r, and that jackknifes of the difference between the input timescales and estimated timescales for flux uncertainties that are characteristic for lightcurves at 20th and 22nd magnitude. Panels from left to right have increasing spacing between observations. For example, the "cad=30" panel shows the trial in which the lightcurve was observed every 30 days. As expected, it is more difficult to measure τ's shorter than the cadence.

Panels from top to bottom show increasing SNR. At 20th magnitude, the estimated uncertainty in measured τ (numbers labeled "mean jk std" (days)) is consistent with the observed scatter in residuals (numbers labeled "hist rms" (days)) for even a $\hat{\sigma} = 0.1$. This confirms that at high SNR, the fitting procedure can accurately infer τ, and that jackknife resampling accurately estimates the uncertainty on those estimates. In the 22nd mag panel, at low SNR (0.634 for a low amplitude of variability, $sig = 0.1$), the estimates on τ break down. The estimated uncertainties are an order of magnitude lower than the observed scatter (due to catastrophic outliers).

This is not surprising, because quasar lightcurves are stochastic. To distinguish a stochastic process from white noise requires sufficient SNR. Consequently, we limited the $z \sim 4$ survey to $i < 22.5$, approximately the average single-epoch 5σ limit.

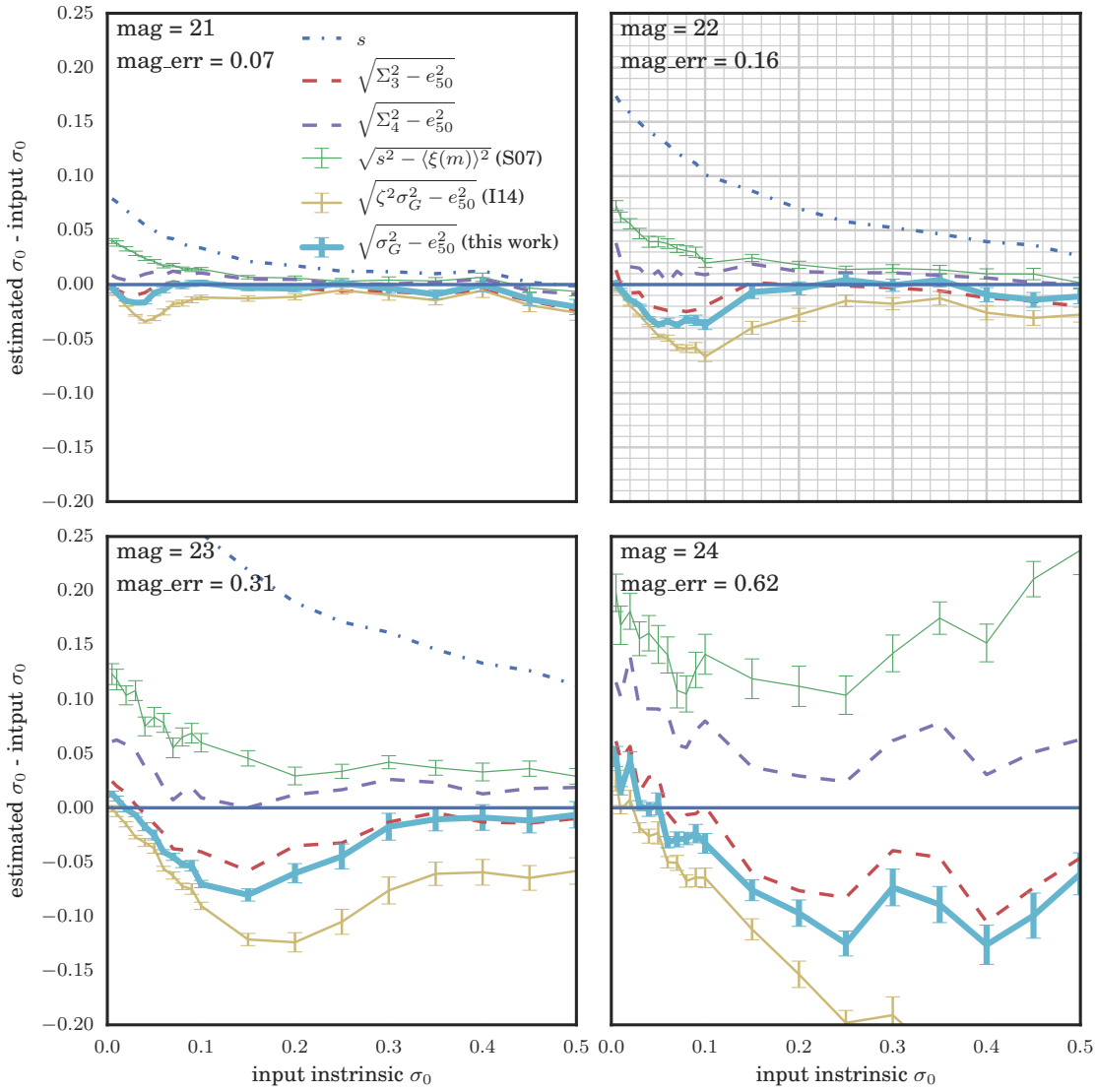

Figure 4.14 Difference in input and estimated intrinsic sigma σ_0 from simulated lightcurves with 4 mean magnitudes, $r = \{21,\ 22,\ 23,\ 24\}$. The recovery of the intrinsic sigma worsens for fainter magnitudes, which have larger average measurement errors. At all magnitudes, the estimator $\sqrt{\sigma_G^2 - e_{50}^2}$ performs the best in terms of accuracy and robustness and thus was computed as the estimator for intrinsic sigma. While $\sqrt{\Sigma_3^2 - e_{50}^2}$ was comparable in accuracy, we found that it was sensitive to the sigma-clipping threshold chosen, and did not use it for that reason.

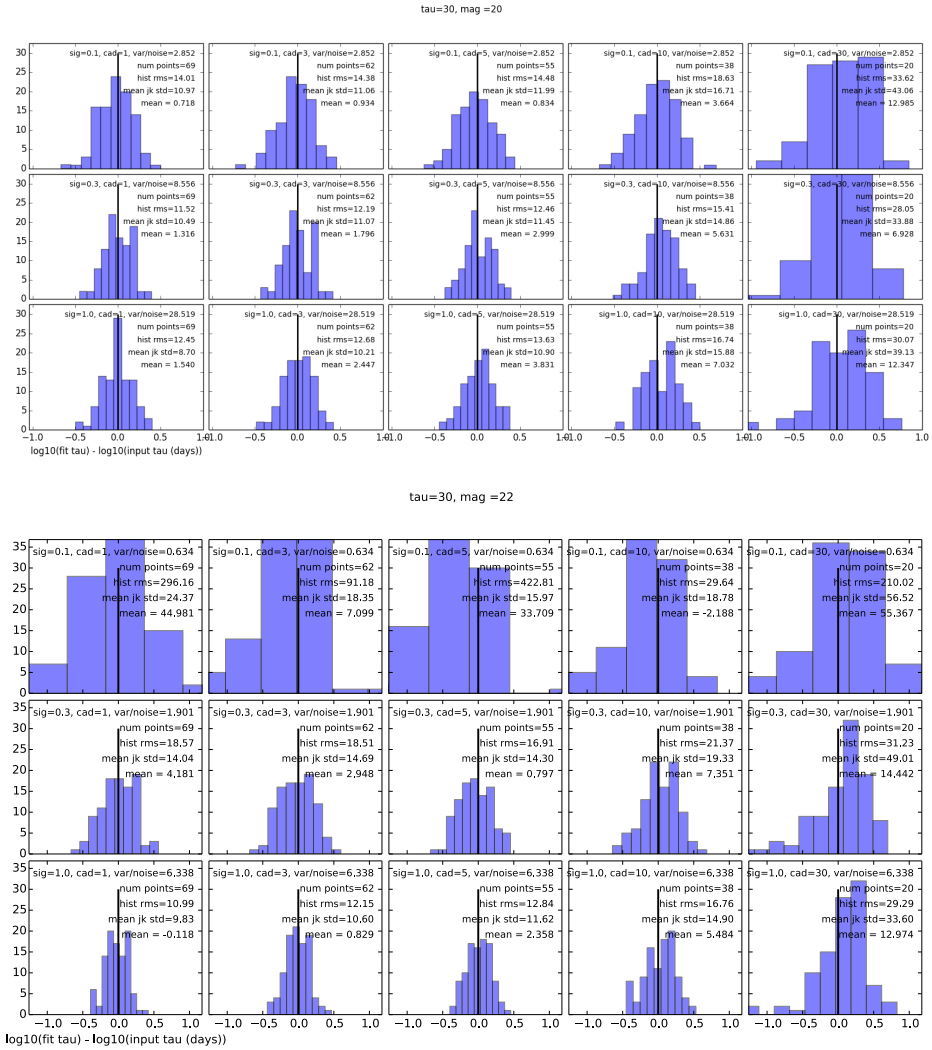

Figure 4.15 Recovery of τ from simulated CAR(1) lightcurves.

Chapter 5

THE REDSHIFT 4 QUASAR LUMINOSITY FUNCTION

5.1 Introduction

In Chapter 4 we developed a classifier for selecting $z \sim 4$ quasars from multi-epoch imaging data. We applied this classifier to the lightcurves extracted from two regions of the Stripe 82 imaging. The first region is the $238\,\mathrm{deg}^2$ ($216\,\mathrm{deg}^2$ unmasked) processed by LSST Summer 13 Data Challenge (Chapter 2). Spectroscopic redshifts are not available for all the quasars in this sample; we call the resulting sample the *photometric sample*. The second region is a $62\,\mathrm{deg}^2$ ($54.6\,\mathrm{deg}^2$ unmasked) sub-region of Stripe 82 on which we achieved 90% spectroscopic completeness; that is, we obtained a spectroscopic redshift (Chapter 3) for nearly all of the candidates in this region for $i < 21.5$. We call the resulting sample the *spectroscopic sample*. The spectroscopic sample is a subset of the Photometric Sample because we used identical selection techniques for both samples and the $54.6\,\mathrm{deg}^2$ survey area is a sub-region of the $216\,\mathrm{deg}^2$ survey area. The only difference is that the Photometric Sample uses photometric redshifts when spectroscopic redshifts are not available.

In this chapter we describe the two samples in more detail and and explain how we estimated the selection function using simulated quasars. Finally, we present the two LF measurements and discuss implications for the evolution of SMBHs and reionization.

5.2 Sample and Completeness

To measure the QLF, we must define uniformly selected samples and quantify the selection function—the probability, $P(z, M)$, that the classifier selects a quasar of a given redshift z and absolute magnitude M. In this section, we describe both the resulting spectroscopic and photometric QLF samples and analyze all sources of incompleteness for both samples,

including the spectroscopic completeness and the selection function.

5.2.1 Photometric Sample

The final photometric sample covers the full, uniform 216 $\deg^2$ survey area, shown in figure 4.3. It covers the 235 $\deg^2$ rectangle $-38.5 < R.A. < +55$, $-1.256 < Dec. < +1.256$. Masked regions comprised 19 $\deg^2$ that were inaccessible to the survey because of bright sources and scattered light.

The sample consists of all quasars that met the selection criteria and for which the classifier gave a probability greater than 72%. We confirmed that the resulting luminosity function was not sensitive to the probability thresholds between 50% and 80%. We chose a threshold of 72% to minimize false positives. Missed quasars are accounted for by the selection function. The final sample size is 625 quasars.

Approximately half ($N = 348$) of the photometric sample ($N = 625$) has spectroscopic redshifts. Note that this count includes lower-redshift $3.4 < z < 3.75$ quasars that will not make it into our luminosity function bin. For the 277 quasar candidates that do not have spectroscopic redshifts, we impute these missing redshifts with photometric redshifts.

5.2.1.1 Photometric Redshift Estimation

For the photometric sample, we estimate missing redshifts using colors. Redshift correlates very strongly with $(g - r)$ (§4.3), and regressing by $(g - r)$ is an efficient way to estimate redshift. We fit a multiple regression model to the quasars for which spectroscopic redshifts were available. We tested several models and found that i-mag and interaction terms (including between $(g-r)$, $(r-i)$), were not significant variables. We fit a robust linear model with variables $(g - r)^2$, $(g - r)$, and $(r - i)$ using a Huber loss function.

Figure 5.1 shows the spectroscopic versus fit photometric redshifts for the fiducial robust photo-z model (left) and a weighted least squares model (right) for comparison. Points are colored by i-mag. Both perform comparably, though the RMS of residuals for $z > 3.7$ quasars for the fiducial model was 0.118 vs. 0.127 for the non-robust, weighted least squares plotted.

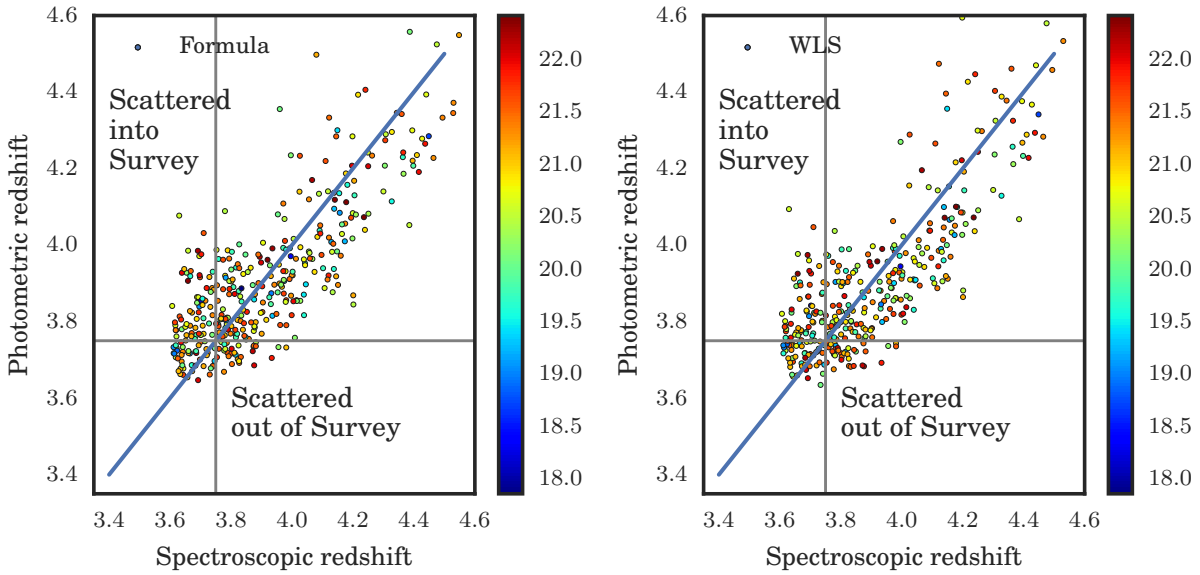

Figure 5.1 Photometric redshifts vs. spectroscopic redshifts. Left shows the multiple regression formula used for the photometric luminosity function. Right shows a robust weighted least squares for comparison. Because number counts drop with redshift, photometric redshift uncertainties mainly affect the lower edge of the redshift bin.

The RMS of ~ 0.12 comes primarily from the intrinsic scatter. With the information available, optical colors, we cannot achieve more accurate photometric redshifts. However, our redshift bin is wide ($3.75 < z < 4.5$) and the effects of uncertainties in photometric redshifts minimal.

The main effect of this scatter is to move a quasar into or out of the redshift bin from the low-redshift side. More precisely, it will underestimate the redshift of a handful of quasars with $z > 3.75$ which will be erroneously excluded from the sample. Likewise it will overestimate the redshift of a handful of quasars with $z < 3.75$, which will be erroneously included in the sample. Therefore, it is important that the residuals are randomly distributed so as not to introduce any bias in the LF. Gray lines in figure 5.1 show the edges of the LF survey bin, and the blue line shows the one-to-one line. Quasars that would be scattered

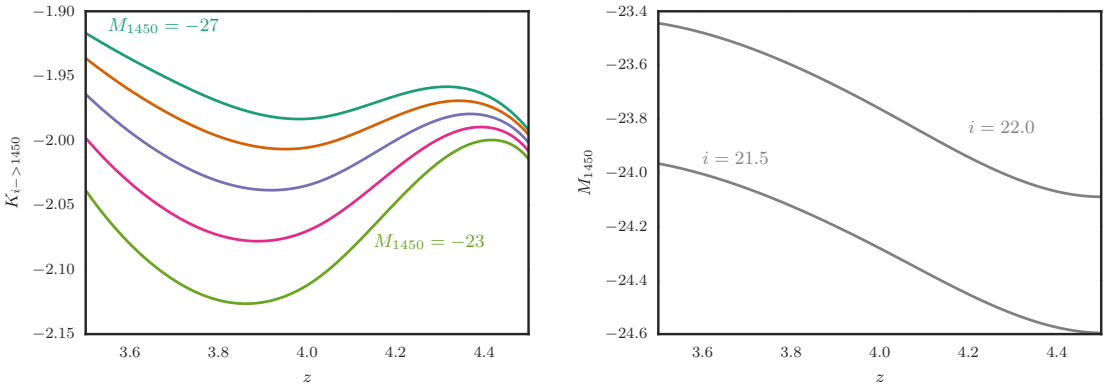

Figure 5.2 Left: Luminosity-dependent k-correction as determined from the quasar spectral simulations. The lines show the k-correction from observed i-band magnitude to M_{1450} for $M_{1450} = -27$ (top) to $M_{1450} = -23$ (bottom) in steps of $\Delta M = 1$. Right: Estimated luminosity as a function of redshift for $i = 21.5$ and $i = 22.0$.

into or out of the survey (if all redshifts were replaced with photometric redshifts) fall into the marked regions. Ideally, these regions would contain equal numbers of quasars.

A secondary effect of redshift error on the LF is the translation to an error in luminosity. Within the sample, the photometric redshift scatter will manifest in the LF by affecting which luminosity bin a quasar falls into. However, at $z = 4$, a 0.15 uncertainty in redshift is only a 0.1 uncertainty in luminosity (figure 5.2), so the resulting errors in luminosity are minimal, especially given that the bins in the binned QLF are 0.5 mag wide.

5.2.2 Spectroscopic Sample

The spectroscopic survey area included the patches covered by the North runs in the R.A. range of $-55 <$ R.A $< +10$.[1] We chose this particular region to take advantage of both our observing time and the native organization of the data. We had full nights at the telescope, and we wanted a region small enough that we could achieve spectroscopic completeness in

[1]Patches with odd indices in the declination direction correspond to the North strip.

that time. We also wanted accessible targets in a wide range of right ascensions, and we wanted clean borders (edges), which are provided by the inner patch boundaries.

The LF derived from the spectroscopic sample only contains spectroscopic redshifts. Thus, an additional source of incompleteness that applies only to this sample is the spectroscopic completeness——the fraction of quasar candidates in the $55\,\mathrm{deg}^2$ region spectroscopic redshifts.

We compute the spectroscopic completeness in order to account for the missing spectra in the spectroscopic QLF calculation. Figure 5.3 shows a heatmap of the spectroscopic completeness of the $55\,\mathrm{deg}^2$ spectroscopic survey as a function of apparent i-mag and color $(g-r)$. We included $(g-r)$ colors as a dependent variable because of the possibility that we preferentially targeted redder sources at the telescope, knowing that they were more likely to be higher redshift (in our sample at $z > 3.75$). The survey is spectroscopically complete for brighter candidates $i < 21$ and drops to $\sim 50\%$ at $i = 22$. The $(g-r)$ dependence is only apparent in the faintest bin. Of the 150 candidates in the spectroscopic survey with $P(z > 3.5) > 75\%$, we are missing spectra for 32 candidates, 25 of which are $21.5 < i < 22$.

5.2.3 Luminosity Estimation

We estimate the luminosity from the i-band magnitude and best available redshift. Luminosity can be estimated directly from spectra, but doing so requires high-SNR, photometrically calibrated spectra. Our spectra are low SNR, because we focused on getting many redshifts and because we did not undertake an accurate flux calibration of our spectra.

In the redshift range $3.7 < z < 4.4$, the i-band is the most accurate broad band with which to estimate the monochromatic luminosity at rest-frame wavelength of 1450 Å (M_{1450}), because it directly covers the rest-frame wavelength range 1300–1700 Å. The relationship between M_{1450} and i (i.e. the k-correction) is computed using model SEDs of quasars. Because the i-band directly measures the rest-frame 1450 Å, the estimate of M_{1450} is less sensitive to the SED models. We derive the average correction for the observed i-band flux to M_{1450} for a large number of simulated quasars—derived from our quasar SED model

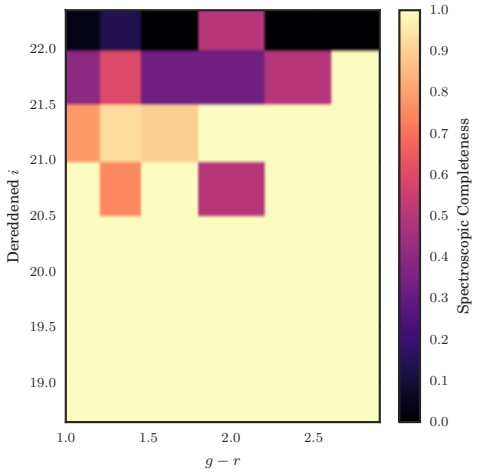

Figure 5.3 Spectroscopic completeness for the 55 deg^2 spectroscopic sample. Candidates are binned by color $(g-r)$ and apparent magnitude (i) with nearest-neighbor interpolation. The survey is spectroscopically complete at $i < 21$, dropping to $\sim 50\%$ at $i = 22$. The $(g-r)$ dependence is only apparent in the faintest bin.

(§5.2.5)—in narrow bins of (M, z), then interpolate this grid to derive an individual quasar correction.

The distribution of intrinsic quasar SEDs at a given luminosity and redshift is quite broad, introducing uncertainty into our luminosity estimation.

5.2.4 Selection Function

To measure the selection function, we simulate all features used for selection, which can be separated into two categories. In the first category are the features that can be sampled from our quasar SED model (§5.2.5): M_{1450}, z, and colors $(u-g)$, $(g-r)$, $(r-i)$, and $(i-z)$. In the second category are the features that *cannot* be sampled from the quasar model: variability metrics and extendedness. In the following two subsections we describe how we produced a full-featured set of simulated quasars, starting with the quasar model

and then sampling the variability metrics and extendedness from the empirical distributions of known quasars.

5.2.5 Simulations

5.2.5.1 Simulating Colors

We constructed a model for the observed spectroscopic and photometric properties of quasars. From this model we generate a large sample of simulated quasars. We can use the observed properties of the model to estimate luminosity (M_{1450}), and to validate photometric redshift estimates. Most importantly, we can test the classifier on the simulated quasars to estimate the selection function.

We use the quasar model developed by McGreer et al. (2013). The quasar spectral model consists of a broken power law with a break at 1100 Å. On the blue side, the power law are indexes drawn from a Gaussian distribution $\mathcal{N}(-1.7, 0.3)$, and on the red side, from $\mathcal{N}(-0.5, 0.25)$. Emission lines are generated from composite spectra of BOSS quasars binned by luminosity, reproducing trends between emission line properties and continuum luminosity (e.g. Baldwin 1977). It also takes advantage of the many improvements in recent years for understanding quasar spectral properties. These include the improved measurements of absorption due to the Lyα forest (Worseck et al. 2011), and to new templates for iron emission which affect quasar colors (Vestergaard & Wilkes 2001).

We simulated colors using the quasar models by first constructing a grid of model quasars in (M_{1450}, z) space. From each bin, in increments of 0.1 in M_{1450} and 0.05 in z, we sampled 100 M_{1450}, z pairs. For each quasar, we generated a simulated spectrum by randomly sampling from the distribution of spectral features of the luminosity- and redshift-dependent model.

After sampling a spectrum from this quasar model, we convolve the spectrum with the SDSS filter curves and add simulated error, drawn from a model of the photometric flux uncertainty. The flux uncertainties were modeled from the empirical uncertainties on the median forced photometry in $u, g, r, i, z.$.

5.2.5.2 Simulating Variability Metrics

Because we jointly selected quasars based on their colors and their variability, only simulating colors is not sufficient; we also need to simulate the variability metrics. However, the study of quasar variability is still relatively new. Because we do not know definitively what drives quasar variability, we cannot produce model lightcurves from intrinsic models. Doing so would require a detailed understanding of the correlations between color *and variability* characteristics. Some of these correlations are well known. For example, individual quasars get bluer when they are brighter (Schmidt et al. 2012); statistically, intrinsically fainter quasars are more variable. Many other correlations, however, remain unknown. Therefore, we simulate quasar variability metrics based on empirical relations observed in our sample.

We simulate variability metrics by fitting a Gaussian mixture model (GMM) to the observed values—z, M_{1450}, 4 colors, and 11 variability metrics—from the training sample using the following procedure.

1. **Build the variability model:** Transform each of the 17 variables to follow a Gaussian distribution using a Box-Cox transformation (Box & Cox 1964); this transformation is necessary to make the distribution well fit by a GMM. Model the 17-dimensional transformed distribution as a GMM, using five 17-dimensional Gaussians. Five was chosen because it minimized the Akaike information criterion (Akaike 1974).

2. **Sample the remaining 11 features[2] from this variability model for each simulated quasar.** For each of the quasars simulated from the spectral models described above, z, M_{1450} and the four colors are given and the independent variables in this step.

    - Transform these six independent variables using the same λ used to Box-Cox-transform the training set on which the model was built.

[2]These features include `extendedness` and variability metrics, but we will refer to them as variability metrics for simplicity in this demonstration.

- Condition the 17-dimensional GMM on these six variables to produce an 11-dimensional GMM of variability characteristics for that particular simulated quasar.
- Draw 11 variability metrics from this 11-dimensional conditional GMM.
- Box-Cox transform these simulated variability metrics back to observed units.

The resulting simulated samples reproduce all correlations between luminosity, redshift, colors and variability apparent in the empirical sample.

5.2.5.3 Result: Selection Function

We compute the selection function by passing the simulated sample of quasars through the selection procedure (chapter 4), recording the fraction of quasars selected. Figure 5.4 shows the resulting selection function for the $z \sim 4$ classifier as a function of redshift and luminosity. Lighter colors indicate higher completeness. The sharp cutoffs at $z = 3.6$ and $z = 4.4$ are due to the criteria that $(g - r) > 0.95$ and $(r - i) < 0.3 \times (g - r) + 0.03$, respectively. These criteria narrow the selection to this particular redshift range. Completeness drops off for luminosities fainter than $M_{1450} = -23.5$ because of the requirement that $i < 22.0$. In the luminosity and redshift range of interest, we reach nearly uniform $\sim 90\%$ completeness.

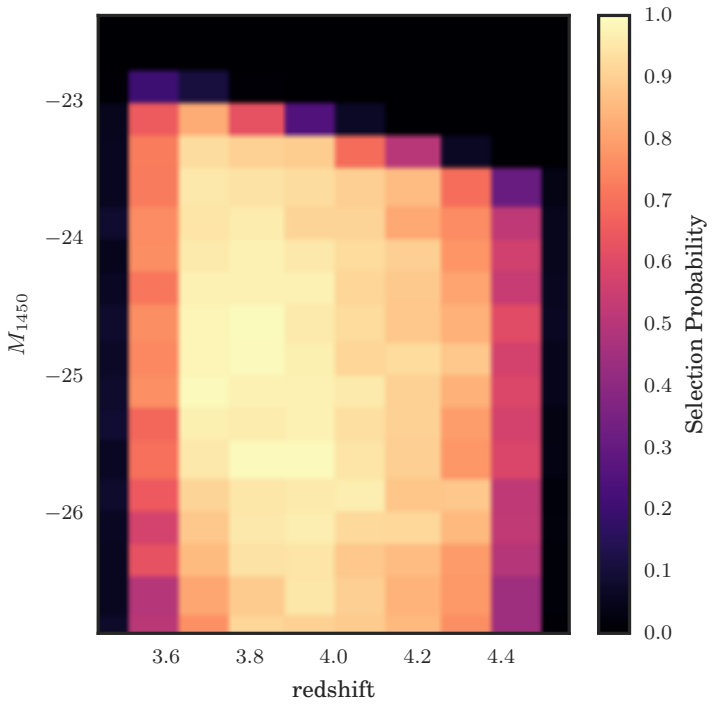

Figure 5.4 The selection function for our $z \sim 4$ survey. The selection probability is calculated in each bin by counting the fraction of simulated quasars that pass the selection criteria. The sharp cutoffs at $z = 3.6$ and $z = 4.4$ are because of the criteria that $(g - r) > 0.95$ and $(r - i) < 0.3 \times (g - r) + 0.03$, respectively. The diagonal cutoff is because of the criterion that $i < 22.5$. Nearest-neighbor interpolation was applied.

5.3 Luminosity Function

We compute the luminosity function for both the spectroscopic and photometric samples after correcting for incompleteness from the selection function, as described in section 5.2.

5.3.1 Binned Luminosity Function

First, we calculate the binned luminosity function for the spectroscopic and photometric samples. We restrict the sample to one redshift bin, $3.75 < z < 4.4$, and ignore any evolution

of the QLF parameters over the width of the bin. We use nine luminosity bins, each 0.5 mag wide, in the range $-27.5 < M_{1450} < 23.5$. We use the $1/V_a$ method (Schmidt 1968; Avni & Bahcall 1980), including the correction from Page & Carrera (2000), which shrinks the volume in the faintest bins if the contents are only partially observable within the limiting magnitude of the survey. We estimate ϕ, the number density for a given (M_{1450}, z) bin per unit volume, as

$$\hat{\Phi} = \frac{N_{cor}}{\int_{L_{\rm min}}^{L_{\rm max}} \int_{z_{\rm min}}^{z_{\rm max}(L)} (dV/dz)\ dz\ dL}. \tag{5.1}$$

The number count, N, is corrected by the selection function and divided by the effective volume element dV of that bin. The effective volume is calculated assuming the WMAP9 (Hinshaw et al. 2013) cosmology. The survey area for the uniform spectroscopy and photometric samples are 54.6 and 216 $\deg^2$, respectively. The uncertainty on $\hat{\Phi}$, given by $\delta\hat{\Phi}$, is estimated by

$$\delta\hat{\Phi} = \frac{\sqrt{N_{cor}}}{\int_{L_{\rm min}}^{L_{\rm max}} \cdot \int_{z_{\rm min}}^{z_{\rm max}(L)} (dV/dz)\ dz\ dL} \tag{5.2}$$

Table 5.1 presents the binned luminosity function for both the spectroscopic and photometric samples. In each magnitude bin, we include the number counts, the corrected number counts after accounting for all sources of incompleteness, and the QLF estimates $\hat{\phi}$ and $\delta\hat{\phi}$.

Figure 5.5 shows the binned luminosity functions for the spectroscopic (green) and photometric (orange) samples, with previous measurements overplotted in gray. The LFs are consistent in the photometric and spectroscopic samples, thus validating the purely photometric classification method.

The binned luminosity function agrees with bright-end measurements derived from the wide area SDSS DR7 (Shen & Kelly 2012), and goes two magnitudes deeper, revealing the break in the power law. The LF measurements agree with the faint-end LF measured by Masters et al. (2012) and Ikeda et al. (2011) using COSMOS data.

Table 5.1. Binned QLF $3.75 < z < 4.5$

M^*_{1450}	N	$N_{\rm cor}$	$\Phi^{\rm a}$	$\sigma_\Phi{}^{\rm b}$
		Spectroscopic		
-27.25	1	1.1	0.5	0.5
-	0	-	-	-
-26.25	6	7.5	3.2	1.3
-25.75	9	10.2	4.3	1.5
-25.25	9	11.8	5.0	1.8
-24.75	22	26.5	11.2	2.4
-24.25	25	41.5	17.5	4.0
-23.75	7	44.5	19.8	12.1
-	0	-	-	-
		Photometric		
-27.25	2	3.3	0.4	0.3
-26.75	10	11.5	1.2	0.4
-26.25	22	25.1	2.7	0.6
-25.75	54	58.9	6.3	0.9
-25.25	64	71.6	7.6	1.0
-24.75	96	103.1	11.0	1.1
-24.25	136	146.0	15.5	1.3
-23.75	186	225.1	25.3	2.0
-	0	-	-	-

[a] Φ is in units of 10^{-8} Mpc^{-3} mag^{-1}.

[b] σ_Φ is in units of 10^{-8} Mpc^{-3} mag^{-1} and includes statistical uncertainty.

5.3.2 Parameter Estimate from Maximum Likelihood

Over a wide redshift range, the QLF is well fit by a double power law (Boyle et al. 1988) of the form $\Phi(L) = \phi^*[(L/L^*)^\alpha + (L/L^*)^\beta]^{-1}$, with a characteristic, or break, luminosity L^*. Converting this parameterization to absolute magnitude gives the number density of quasars per unit magnitude:

$$\Phi(M, z) = \frac{\phi(M^*)}{10^{0.4(\alpha+1)[M-M^*(z)]} + 10^{0.4(\beta+1)[M-M^*(z)]}}, \tag{5.3}$$

where $\Phi(M^*)$ is the normalization and M^*_{1450} is the break luminosity separating the bright and faint ends. We follow the convention that α describes the faint-end QLF slope and β the bright-end slope. Evolution of the QLF is encoded in the redshift dependence on the break luminosity M^*_{1450}, ϕ_*, and on the slopes α and β.

The model fits are presented in table 5.2. Figure 5.6 shows the likelihood contours for α, β, Φ, and M_{1450} fits for the photometric sample. The break luminosity M_{1450} is well constrained, which enables study of the evolution of the QLF through $z \sim 4$. The density at the break luminosity Φ^* covaries with M^*_{1450} because density is a steep function of luminosity. The faint-end slope is also well constrained, which enables constraints on quasars' contribution to reionization. The bright-end slope β is poorly constrained, because there are only 11 quasars brighter than $M_{1450} < -26.5$, and 2 brighter than $M_{1450} < -27.0$, in the photometric sample. Better constraints on β could be achieved by combining the sample with wide-area surveys.

Like the COSMOS measurements, the QLF normalization is a factor of three lower than that of Glikman et al. (2011) and (Giallongo et al. 2015). The source of this discrepancy is unclear. Masters et al. (2012) suggest it may be due to sample contamination by dwarf stars and high-redshift galaxies in Glikman et al. (2011) at the faintest bins in study. The number of contaminants in broadband quasar searches grows rapidly at the faintest magnitudes, where spectroscopic follow-up is the most challenging and contamination is less well constrained.

Table 5.2. MLE fit parameters

Sample	$\log \Phi^*$	M^*_{1450}	α	β
Photometric	$-7.56^{+0.02}_{-0.02}$	$-26.77^{+0.05}_{-0.05}$	$-1.79^{+0.02}_{-0.02}$	$-4.56^{+0.7}_{-0.9}$
Spectroscopic	$-7.59^{+0.05}_{-0.06}$	$-26.56^{+0.09}_{-0.09}$	$-1.76^{+0.06}_{-0.06}$	-4.5[b]

[a]Φ is in units of Mpc^{-3} mag^{-1}.

[b]Spectroscopic sample β was under-constrained and held fixed.

5.3.3 Evolution of the QLF

Phenomenological models have been proposed to describe how the four double power-law parameters evolve with redshift. In pure luminosity evolution (PLE), only the break luminosity, $M^*(z)$, evolves; the overall number density remains constant while the luminosity changes. In pure density evolution (PDE), only the normalization, $\Phi^*(z)$, evolves; the shape of the QLF remains constant while the number density changes.

Various hybrid models allow both to vary, but hold the bright- and faint-end slopes fixed. In luminosity evolution and density evolution (LEDE), $M^*(z)$ and $\Phi^*(z)$ both evolve independently. In luminosity dependent density evolution (LDDE), the evolution of $\Phi^*(z, M^*)$ is related to that of $M^*(z)$. Finally, extensions to these models allow the power law slopes α and β to evolve as well.

Only by observing the break in the QLF can we constrain models for quasar demographics. With a single power law (i.e. the bright or faint end alone), it is impossible to distinguish between luminosity or density evolution because both would manifest in identically in the QLF.

Figure 5.7 shows the evolution in shape parameters by comparing our $z \sim 4$ best-fit

parameters with those at higher and lower redshifts. At higher redshift, we compared to the $z \sim 5$ QLF measurement reported by McGreer et al. (2013) and the $z \sim 6$ measurement reported by (Willott et al. 2010). At lower redshift, we compared to the $2.2 < z < 3.5$ QLF measurement reported by Ross et al. (2013).

The dashed line in figure 5.7 shows an extrapolation of the LEDE model from Ross et al. (2013). While the measurement of M^*_{1450} at $z \sim 4$ is only slightly discrepant with the model, the model over-predicts the normalization $\Phi*$ for $z > 3.5$. There appears to be a possible turnover in the parameters at $z = 5$. More data would allow for finer redshift bins and a better understanding of this potential effect.

5.3.4 Contribution to Ionizing Background

We compute the ionizing emissivity due to quasars at $z \sim 4$ by integrating the QLF. By comparing this emissivity to theoretical predictions of the number of UV photons required to ionize the IGM, we can estimate the fractional contribution of quasars to reionization.

We integrate the QLF parameters to derive the ionizing emissivity due to quasars at $z \sim 4$. The number of HI and HeII ionizing photons produced per second per unit comoving volume at a given redshift is given by

$$\dot{\mathcal{N}}_{\rm HI}(z) = \int \Phi(M, z)\dot{n}_{\rm HI}(M)\, dM \tag{5.4}$$

and

$$\dot{\mathcal{N}}_{\rm HeII}(z) = \int \Phi(M, z)\dot{n}_{\rm HeII}(M)\, dM, \tag{5.5}$$

where $\dot{n}(M)$ is the number of ionizing photons emitted per second by a single quasar of absolute magnitude M_{1450}, and $\dot{n}(M)$ is simply the quasar SED integrated over ionizing wavelengths. For hydrogen ionization,

$$\dot{n}_{\rm HI}(M_{1450}) = \int_{13.6{\rm eV}}^{\infty} \frac{L(M_{1450}, \nu)}{h\nu}\, d\nu. \tag{5.6}$$

Similarly, the rate of emission of HeII ionizing photons is given by

$$\dot{n}_{\rm HeII}(M_{1450}) = \int_{54.4{\rm eV}}^{\infty} \frac{L(M_{1450}, \nu)}{h\nu}\, d\nu. \tag{5.7}$$

We calculate the ionizing photon output for each quasar by assuming a broken power-law SED with spectral slopes:

$$L(\nu) \propto \begin{cases} \alpha_\nu^{-0.5} & (\lambda > 1100 \text{ Å}) \text{ (Vanden Berk et al. 2001)} \\ \alpha_\nu^{-1.7} & (\lambda < 1100 \text{ Å}) \text{ (Telfer et al. 2002) .} \end{cases} \tag{5.8}$$

5.3.4.1 H I Reionization

We use the best-fit QLF parameters to compute the ionizing photon production and emissivity due to quasars at $z \sim 4$. We integrate the QLF model down to $M_{1450} = -20$ to find the emissivity at 1450 Å and use a conversion factor of $\epsilon_{912}/\epsilon_{1450} = [(1450/1100)^{-0.5} \times (1100/912)^{-1.7}]^{-1} = 1.58$ to convert to the corresponding emissivity at 912 Å. We find $\epsilon_{912} = 1.6 \times 10^{24} \mathrm{erg\,s^{-1}\,Mpc^{-3}\,Hz^{-1}}$ and $\epsilon_{1450} = 2.3 \times 10^{24}$ ergs s^{-1} Mpc^{-3} Hz^{-1}. We integrated the emissivity from 1 Ryd to 4 Ryd, to find the number of ionizing photons $\dot{\mathcal{N}}_{\mathrm{HI,Q}} \sim 1.3 \times 10^{50}$ s^{-1} Mpc^{-3}.

These numbers are sensitive to both the faint limit of integration and the faint-end slope of the QLF.

The number of ionizing photons produced by quasars at $z \sim 4$ can be compared to the critical photon rate per unit comoving volume, $\dot{N}_{\mathrm{HI}}$, at which the new ionization events balance the recombinations. The rate of (case B) recombinations is given by $\alpha_B(T) n_e n_H C$, where C is a clumping factor, n_e the number density of electrons, n_H the number density of protons, and $\alpha_B(T)$ the temperature-dependent hydrogen recombination rate. Previous studies (Glikman et al. 2011; Masters et al. 2012) compared their estimated rates of ionizing photons to the critical $\dot{n}_{\mathrm{HI}}$ from equation 26 in Madau et al. (1999) and assume the fiducial clumping factor of $C = 30$. New estimates of the clumping factor H I $\rightarrow$ H II indicate that $2 < C < 3$ McQuinn et al. (2011). This update indicates that 10$\times$ fewer ionizing photons are needed to counterbalance recombination and keep the hydrogen ionized. We estimate that the number of photons required to maintain full ionization at $z = 4$ ($z = 6$) is $\dot{N}_{\mathrm{HI,rec.,z=4}} = 1.3 \times 10^{50} (C/3)\,\mathrm{s^{-1}\,Mpc^{-3}}$ ($\dot{N}_{\mathrm{HI,rec.,z=6}} = 3.5 \times 10^{50} (C/3)\,\mathrm{s^{-1}\,Mpc^{-3}}$).

Comparing the recombination rate with the observed UV photons from quasars suggests

that at $z \sim 4$ there were just sufficient numbers of quasars to maintain hydrogen reionization alone. Whereas we did not directly measure the ionizing UV flux at $z = 6$, the UV flux from quasars at $z = 4$ is less than a third of the recombination rate at $z = 6$. The quasar emissivity *decreases* from from $z = 4 \rightarrow 6$, by a factor of 4 (Giallongo et al. 2015) to 12 (McGreer et al. 2013; Willott et al. 2010). This suggests that type I quasars were unlikely contributors (3–10%) to hydrogen reionization at $z = 6$. Futhermore, the estimated $\dot{\mathcal{N}}_{\rm HI,Q}$ assumes an escape fraction $f_{\rm esc,Q}$ of unity, and therefore these estimated ionizing rates are *upper limits* on the photons available to ionize the IGM.

5.3.4.2 He II Reionization

Observations of the He II Lyα forest suggest that He II reionization ended at around $z \sim 3$ (McQuinn 2009; Shull et al. 2010; Syphers et al. 2011; Worseck et al. 2011). Quasars have sufficient number counts at this epoch and a sufficiently hard spectrum to ionize He II at this epoch. Thus, He II reionization likely occurred over the course of $3 < z < 4$ and was caused by quasars.

At $z \sim 4$, we estimate that the number of He II-ionizing photons is $\dot{\mathcal{N}}_{\rm HeII,Q} = 1.2 \times 10^{49}\,{\rm s^{-1}\,Mpc^{-3}}$ and the monochromatic emissivity at 4 Ryd is $\epsilon_{228} = 1.9{\times}10^{23}\,{\rm erg\,s^{-1}\,Mpc^{-3}\,Hz^{-1}}$ using a conversion of $\epsilon_{228}/\epsilon_{1450} = [(1450/1100)^{-0.5} \times (1100/228)^{-1.7}]^{-1} = 16.6$ to convert to the corresponding emissivity at 228 Å.

Because He II reionization is ongoing at $z \sim 4$, comparing this $\dot{N}_{ion}$ value with the rate of He III to He II recombinations is of little value because the low comoving n_{HeIII}. The recombination rate can be expressed as

$$\dot{N}_{recomb}^{\rm HeIII\rightarrow HeII} = C_{He}\alpha_B^{\rm HeIII\rightarrow HeII}(T) n_{\rm HeIII} n_e (1+z)^3, \tag{5.9}$$

where $n_e = n_H + n_{He}(1 + X_{HeIII})$, assuming that H I and He I are fully ionized and that He II is fractionally ionized. The fraction of doubly ionized helium X_{HeIII} is the unknown

variable we are computing. The fraction, X_{HeIII}, is given by,

$$H(z)(1+z)n_{He}\frac{\mathrm{d}X_{HeIII}}{\mathrm{d}z} = \int_{4\mathrm{Ryd}}^{\infty} \dot{n}_{\nu}\mathrm{d}\nu - \dot{N}_{\mathrm{recomb}}. \tag{5.10}$$

Integrating equation 5.10 from $z = 5 \rightarrow 3$ (figure 5.8) gives the expected evolution of X_{HeII} with z, given the observed number counts of quasars. We use $C_{He} \sim 1 - 2$, $n_H = 1.8 \times 10^{-7}\,\mathrm{cm}^{-3}$, and a temperature increasing linearly from $T(z = 5) = 8000\mathrm{K}$ to $T(z = 3) = 15,000\mathrm{K}$ (Becker et al. 2011). The IGM temperature increase at this redshift directly results from He II ionization, and can also be computed along with the ionization fraction(Sanderbeck et al. 2015). We use the observed temperature as an input rather than output. The He III ionizing photon rate was interpolated between McGreer et al. (2013) at $z = 5$, to this work at $z = 4$, to Masters et al. (2012) at $z = 3.2$.

From figure 5.8, we see that quasars at $z \sim 4$ have sufficient number density to doubly ionize helium by $z \sim 3$. Sources of uncertainty include the spectral slope from the UV to X-Ray (typical values fall in the range $\alpha = -1.76$ to -1.5), the clumping factor $C_{He} = 1$–2, and the escape fraction. We assume an escape fraction of unity.

D'Aloisio et al. (2016) recently analyzed the effects of ionizing emissivity from quasars to heat the IGM. They computed the heating due to four different estimates of the quasar luminosity function: higher estimates based on the QLF measurements by Glikman et al. (2011); Giallongo et al. (2015), and lower estimates based on the QLF measurements by McGreer et al. (2013); Masters et al. (2012); Willott et al. (2010). Our estimates are consistent with the lowest QLF model, which D'Aloisio et al. (2016) find best reproduces the temperature observed in the IGM (Becker et al. 2011) over $z = 5 \rightarrow 3$. This work offers independent evidence for the lower estimate of UV photon build up due to quasars at $z = 4$.

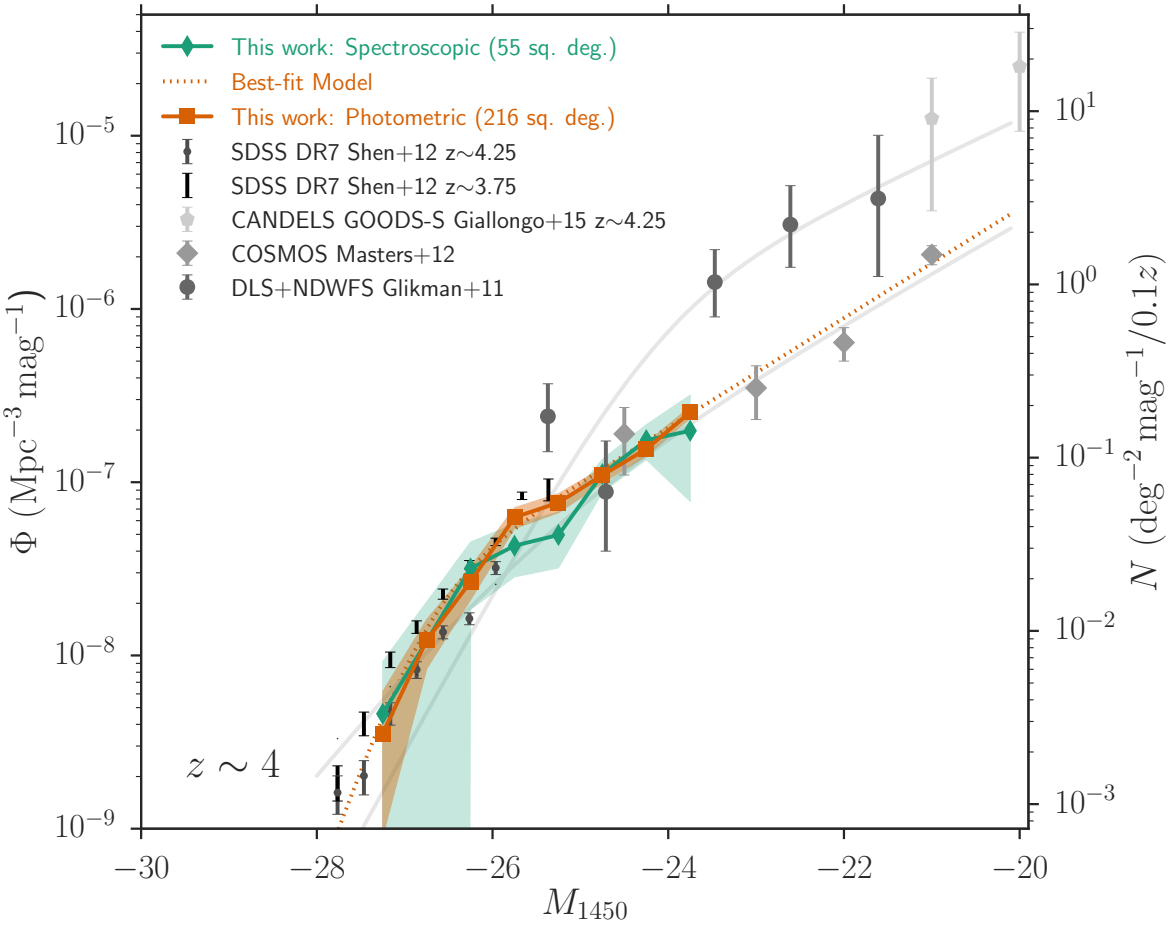

Figure 5.5 The binned quasar luminosity function at $3.75 < z < 4$ for the photometric (orange squares) and the spectroscopic (green diamonds) samples. Points are plotted at the bin centers; uncertainties are shown by the shaded regions. The dotted orange line is the best fit to the double power law for the photometric sample. Previous measurements (grey points) at faint absolute magnitudes are discrepant; the two faint gray lines show the models from (Masters et al. 2012) and (Glikman et al. 2011). We fill the gap between bright measurements in large-area surveys and faint measurements in narrow-field surveys with a variability-selected sample over Stripe 82. The spectroscopically complete sample over 55 deg^2 (green) served as the training set for the full sample (orange) over an effective 216 deg^2.

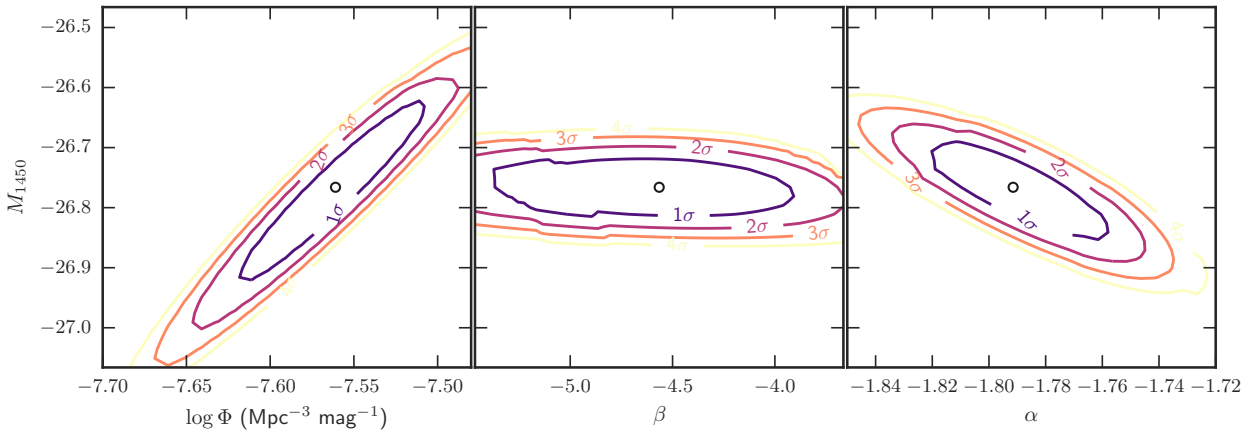

Figure 5.6 Likelihood contours for α, β, and Φ^* vs. M^*_{1450} for the photometric sample.

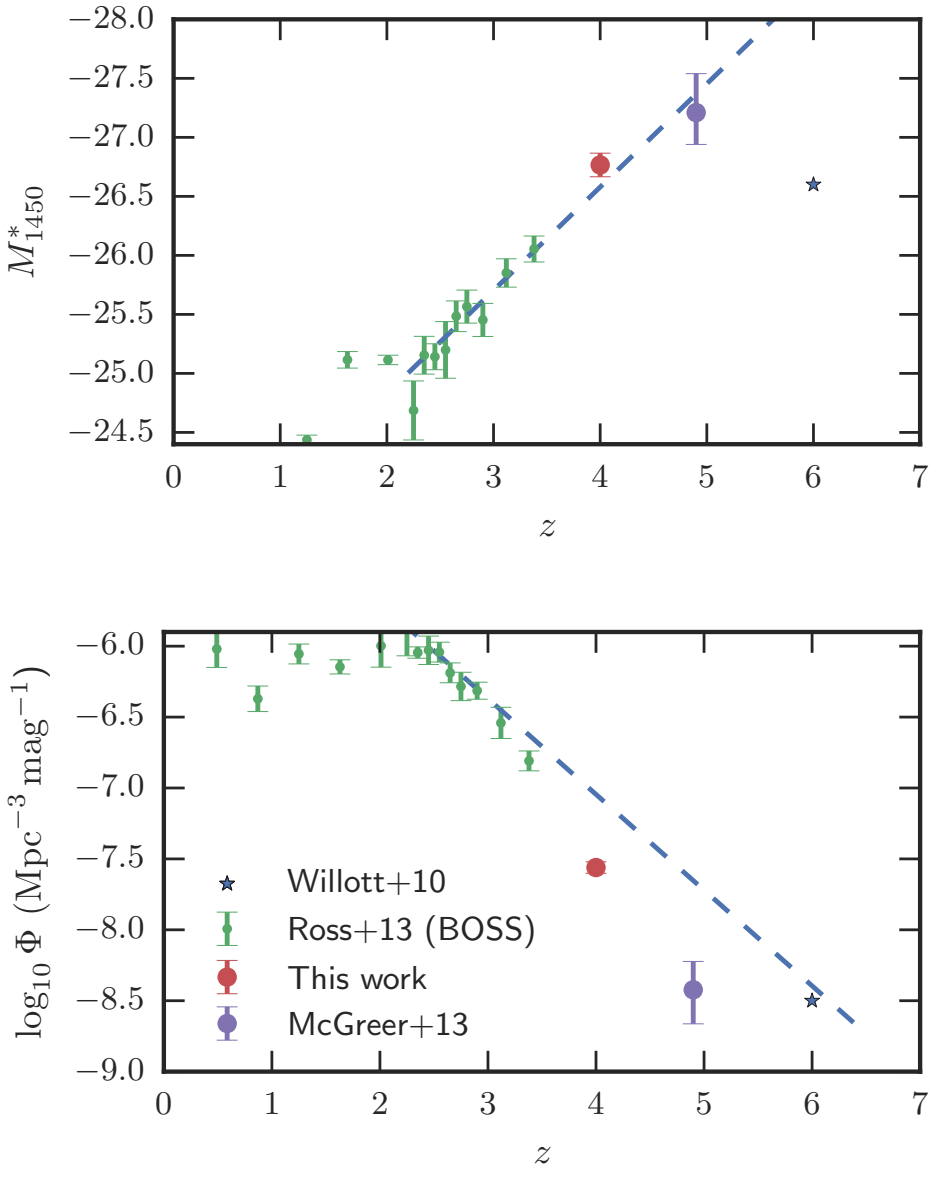

Figure 5.7 Evolution of the QLF normalization (Φ^*) and break luminosity (M^*_{1450}) between $z \sim 2$ and $z \sim 5$. The dashed line shows an extrapolation of the LEDE model from the BOSS DR9 QLF (Ross et al. 2013). The point at $z = 4.9$ is from the best-fit model from McGreer et al. (2013). The point at $z = 4$ is from the best-fit double model in table 5.2.

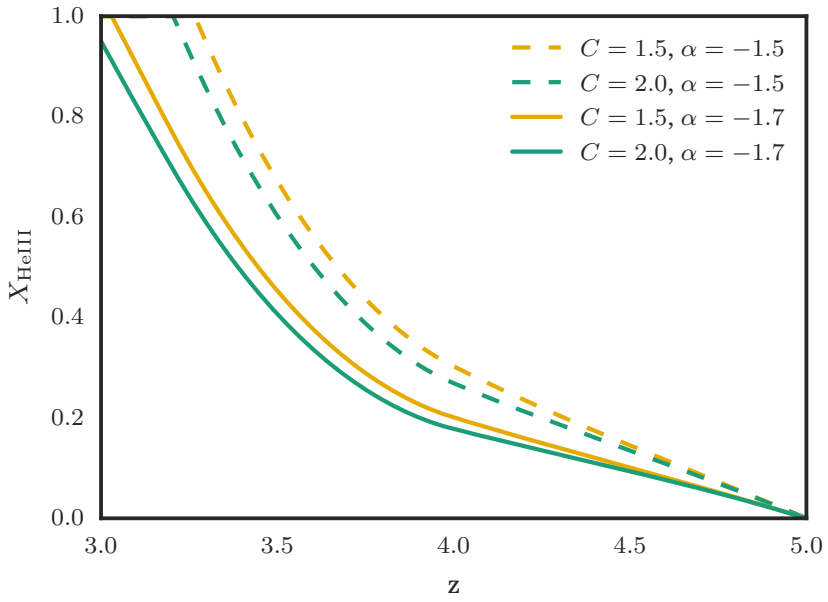

Figure 5.8 Evolution of the He III fraction, $X_{\rm HeIII}$. The He III ionizing photon rate was interpolated from McGreer et al. (2013) at $z = 5$, to this work at $z = 4$, to Masters et al. (2012) at $z = 3.2$. Quasars at $z \sim 4$ have sufficient number density to doubly ionize helium by $z \sim 3$.

5.4 Supplementary Information

5.4.1 Measuring Survey Area

An accurate measurement of survey area is necessary for the overall normalization of the luminosity function. This normalization just requires a single number in $\deg^2$. However, for a clustering measurement, the spatial distribution of candidates is also important. Therefore, we need an accurate map of regions accessible to the survey (for producing random draws when simulating random positions). This section provides a description of problem/assumptions, algorithm, validation, and resulting survey area maps.

Generally, survey area is a function of magnitude, because the limiting depth varies spatially. Even when detecting on a co-add, the number of images going into each patch varies across the sky. However, the SNR requirements for the lightcurve metrics are stricter than those for the colors. Our co-add goes down to an average 5σ depth of 24.5 (figure 4.3). Because we only select objects to $i = 22.5$, 2 mag brighter than the 5σ depth, we treat the survey area that is uncontaminated by bright stars and glints as having a uniform depth.

We mask roughly 10% of the survey area that is inaccessible due to bright stars and glints. On a catalog level, this mask results from filtering sources that are deblender children of footprints brighter than 17th mag. We limit our quasar selection to sources that are not children of sources brighter than $i < 17$.

We compute the survey area masks at the catalog-level as well. We generate the masks at the catalog level using a void-finding algorithm, because bright "source" footprints are empty of $i > 17$ sources. As seen in the left panel in figure 5.9, the bright footprints are uncovered by detected sources (white circles).

To compute the masked polygons, we use an algorithm sometimes referred to as an alpha-hull algorithm. We first compute the Delauney triangulation of the $> 3\sigma$ sources that are not peaks of footprints brighter than $i > 17$; this is seen in the right panel of figure 5.9. Every edge between two adjacent sources contributes to exactly two triangles. Triangles with a circumference larger than a tunable threshold (red) are *pruned* from the full set of

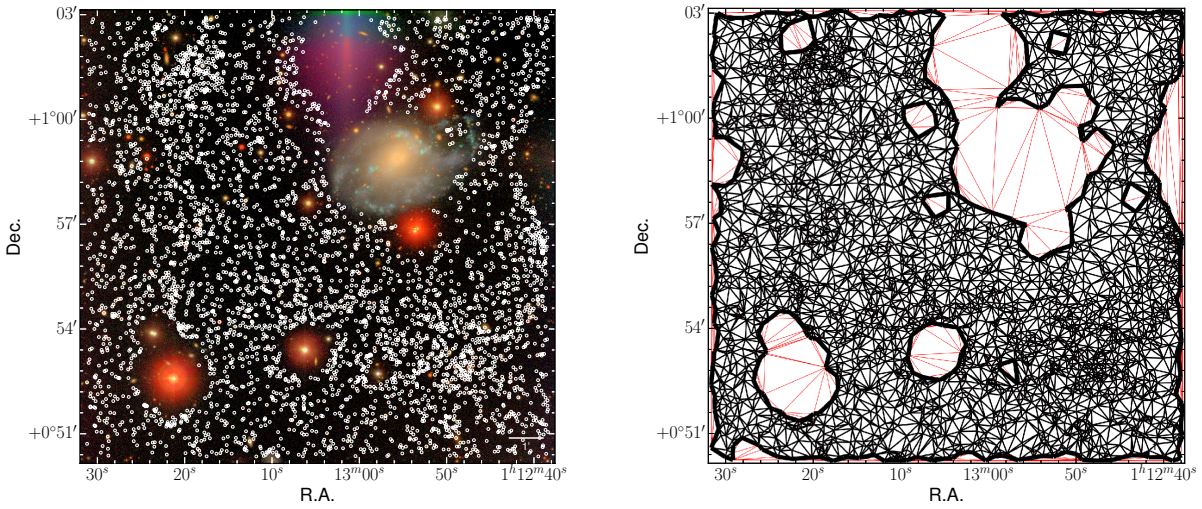

Figure 5.9 Demonstration of survey area measurement on a pathological patch "326,10" on tract "0." Left: Image with RGB = (irg). Overplotted are $> 3\sigma$ sources that are not peaks of footprints brighter than $i > 17$, or within 30 pixels of the edge. Neighboring patches overlap by 60 pixels. Large voids around bright stars, NGC 428, and glints exist. Right: A void-finding algorithm is used to mask these glints and bright objects. Triangles with a circumference larger than a threshold (red) are pruned from a Delauney triangulation of the sources. The remaining edges that appear in the triangulation only once form the edges of the survey area.

triangles. The remaining edges (black only) that appear in the triangulation only once form the boundary of the survey area (thick black lines).

The vertices of the holes (edges that only appear once in the set) comprise mask polygons. We compute both the area of these polygons and the remaining survey area using the Python package `Shapely` (Gillies 2007). We use vertices in R.A. and Dec. coordinates, because Stripe 82 is on the equator and cos(dec) terms are negligible. Note that this algorithm can be used on any part of the sky, if coordinates are first converted to an equa- area projection.

Chapter 6

CONCLUSION

We have used the 238 $\deg^2$ of repeated Stripe 82 imaging, reprocessed as part of the LSST Data Management Software efforts, to conduct a survey of quasars at $z \sim 4$, a redshift where previous constraints of the quasar luminosity function were relatively weak. We extracted lightcurves from the imaging using forced photometry on all available epochs at the positions of sources detected on a deep i-band co-add (chapter 2). We developed a classification method (chapter 3) based on photometric information alone (colors, variability metrics, and morphology), which we validated with a spectroscopically complete 55 $\deg^2$ sub-region augmented with 102 new spectroscopic observations of quasars $z > 3.4$ (chapter 4). We defined a statistical sample of 570 quasars with $i \leq 22.0$ between $3.75 < z < 4.5$ where our estimated completeness is $\sim 90\%$. We estimated the QLF accounting for all sources of incompleteness, presenting the first variability-selected QLF at high redshift. We concluded the following:

- With Stripe 82, we estimate the luminosity function over a 4-mag wide luminosity range ($-27.5 < M_{1450} < -23.5$) with one homogenous, uniformly-selected sample providing the best constraints on the break luminosity to date: $M^*_{1450}(z = 4) = -26.7$.

- Comparing constraints on M^*_{1450} and Φ^* at $z = 4$ with other redshifts, we see LEDE: the break luminosity brightens and density decreases from $z = 2.5 \rightarrow 5$.

- There is no clear evidence for evolution in the slopes of the QLF.

- The $z \sim 4$ QLF shows good agreement with the faint-end measurements from COSMOS (Ikeda et al. 2011; Masters et al. 2012); but does not show agreement with the higher-

density estimates from DLS+NDWFS (Glikman et al. 2011) and GOODS-S (Giallongo et al. 2015).

- The best-fit QLF at $z \sim 4$ predicts that quasars contribute $\sim 100\%$ of the ionizing photons required to maintain hydrogen ionization, and have sufficient UV flux to complete He II reionization by $z \sim 3$.

- Purely photometric selection methods for statistical samples of quasars are viable for future surveys.

6.1 Future Directions and Outlook

Clustering: The luminosity function specifies the number of quasars per redshift and per luminosity and constrains both the ionizing emissivity and the evolution of the number of quasars. However, alone it does not constrain the fraction of black holes or dark matter halos that host an active quasar (Veale et al. 2014). Constraining the so-called "duty cycle" requires a measurement of the clustering strength to compare to the clustering strength of the underlying dark matter distribution.

There have been a number of quasar-clustering measurements in the optical for $z < 2.2$ across a range of scales (e.g. Croom et al. 2005; Hennawi et al. 2006; Porciani et al. 2004; Porciani & Norberg 2006; Myers et al. 2007, 2008; Ross et al. 2009). At redshifts between $2.2 < z < 4$, relatively weak constraints exist (Shen et al. 2007; Eftekharzadeh et al. 2015; White et al. 2012). At higher redshifts, quasars are too rare for typical correlation-function estimators, and clustering constraints are therefore made using serendipitously discovered quasar pairs.

The $z \sim 4$ quasar sample presented in chapter 5 is sufficiently well-defined and large to make a measurement of the clustering strength at $z \sim 4$ via the two-point correlation function. With these measurements, we can answer outstanding questions: whether the brightest quasars reside in massive dark matter halos, and how the fraction of SMBHs that are active at a given epoch evolves.

Simulating Quasar Variability: Previous high-redshift QLF surveys have either been wide (SDSS) *or* deep (COSMOS). To make significant progress in constraining the faint high-redshift QLF, the survey must be both. The upcoming LSST will cover 20,000 $\deg^2$ to a single-epoch point source depth of $r \sim 24.5$, and to a co-added depth of $r \sim 27.5$. With an étendue an order of magnitude larger than any other survey, it will image half the sky every week, with over 150 repeated observations in each passband at each position. Because of this large sky coverage, depth, and temporal information, the LSST will produce an AGN sample that will dwarf even the largest current AGN samples. The survey expects sample sizes of over 10^6 quasars at $z \sim 4.5$, 10^5 quasars at $z \sim 5.5$, and 10^4 quasars at $z \sim 6.5$ (in bins of $\Delta z = 1.0$)(Collaboration et al. 2009).

Complete spectroscopic followup will be infeasible at these numbers and faint magnitudes, and we must prepare to select and characterize quasars using photometry alone. We have demonstrated that variability selection is a promising technique for compiling statistical samples sufficiently pure and complete that spectroscopic confirmation is no longer needed.

In this work, we estimated the selection function using the empirical distributions of quasars' variability metrics. We find that estimates of quasar variability depend on the survey cadence, and estimating the selection function required a training set of known $z \sim 4$ quasars from the same survey (with identical cadence) as the new candidates. In the future, we will need to understand quasar variability well enough to develop models of time-varying quasar spectra that capture the breadth of correlations between colors and lightcurves. These models will improve the selection function of variability-selected samples, and will allow us to connect time domain surveys with different cadences.

BIBLIOGRAPHY

Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2003, The Astronomical Journal, 126, 2081

—. 2004, The Astronomical Journal, 128, 502

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, The Astrophysical Journal Supplement, 182, 543

Akaike, H. 1974, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, AC-19, 716

Alexander, D. M., & Hickox, R. C. 2011, arXiv.org, 93

Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, The Astrophysical Journal, 794, 120

Avni, Y., & Bahcall, J. N. 1980, Astrophysical Journal, 235, 694

Axelrod, T., Kantor, J., Lupton, R. H., & Pierfederici, F. 2010, in Proceedings of the SPIE, ed. N. M. Radziwill & A. Bridger, Steward Observatory, United States (SPIE), 774015–774015–10

Baldwin, J. A. 1977, 214, 679

Barris, B. J., Tonry, J. L., Novicki, M. C., & Wood-Vasey, W. M. 2005, The Astronomical Journal, 130, 2272

Battiti, R. 1994, IEEE Transactions on neural networks

Becker, G. D., Bolton, J. S., Haehnelt, M. G., & Sargent, W. L. W. 2011, Monthly Notices of the Royal Astronomical Society, 410, 1096

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

Bosch, J. 2010, The Astronomical Journal, 140, 870

Bovy, J., Myers, A. D., Hennawi, J. F., et al. 2012, The Astrophysical Journal, 749, 41

Box, G. E. P., & Cox, D. R. 1964, Journal of the Royal Statistical Society. Series B (Methodological, 211

Boyle, B. J., Shanks, T., & Peterson, B. A. 1988, Monthly Notices of the Royal Astronomical Society (ISSN 0035-8711), 235, 935

Breiman, L. 2001, Machine learning, 45, 5

Butler, N. R., & Bloom, J. S. 2011, The Astronomical Journal, 141, 93

Caplar, N., Lilly, S. J., & Trakhtenbrot, B. 2015, The Astrophysical Journal, 811, 148

Choi, Y., Gibson, R. R., Becker, A. C., et al. 2014, The Astrophysical Journal, 782, 37

Collaboration, L. S., Abell, P. A., Allison, J., et al. 2009, arXiv.org, 201

Conroy, C., & White, M. 2013, The Astrophysical Journal, 762, 70

Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, 356, 415

Croom, S. M., Richards, G. T., Shanks, T., et al. 2008, arXiv.org, 19

—. 2009, Monthly Notices of the Royal Astronomical Society, 399, 1755

D'Aloisio, A., Sanderbeck, P. R. U., McQuinn, M., Trac, H., & Shapiro, P. R. 2016, arXiv.org, arXiv:1607.06467

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, A&A, 145, 10

Duchon, C. E. 1979, Journal of Applied Meteorology, 18, 1016

Eftekharzadeh, S., Myers, A. D., White, M., et al. 2015, A&A, 453, 2779

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, A&A, 142, 72

Fan, X. 1999, The Astronomical Journal, 117, 2528

Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001, The Astronomical Journal, 121, 54

Ferrarese, L., & Merritt, D. 2000, The Astrophysical Journal, 539, L9

Freund, Y., & Schapire, R. E. 1997, Journal of computer and system sciences, 55, 119

Frieman, J. A., Bassett, B., Becker, A., et al. 2008, The Astronomical Journal, 135, 338

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, Astronomical Journal v.111, 111, 1748

Gebhardt, K., Bender, R., Bower, G., et al. 2000, The Astrophysical Journal, 539, L13

Giallongo, E., Grazian, A., Fiore, F., et al. 2015, Astronomy & Astrophysics, 578, A83

Gibson, R. R., Brandt, W. N., Gallagher, S. C., Hewett, P. C., & Schneider, D. P. 2010, The Astrophysical Journal, 713, 220

Gibson, R. R., Brandt, W. N., Schneider, D. P., & Gallagher, S. C. 2008, The Astrophysical Journal, 675, 985

Gillies, S. 2007

Glikman, E., Djorgovski, S. G., Stern, D., et al. 2011, The Astrophysical Journal, 728, L26

Gonzalez, R. C., & Woods, R. E. 2011, Digital Image Processing (Pearson Higher Ed)

Graham, M. J., Djorgovski, S. G., Drake, A. J., et al. 2014, Monthly Notices of the Royal Astronomical Society, 439, 703

Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, A&A, 698, 198

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, The Astronomical Journal, 116, 3040

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, The Astronomical Journal, 131, 2332

Hennawi, J. F., Strauss, M. A., Oguri, M., et al. 2006, 131, 1

Hickox, R. C., Mullaney, J. R., Alexander, D. M., et al. 2014, The Astrophysical Journal, 782, 9

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, The Astrophysical Journal Supplement, 208, 19

Huff, E. M., & Graves, G. J. 2014, The Astrophysical Journal Letters, 780, L16

Huff, E. M., Hirata, C. M., Mandelbaum, R., et al. 2011, arXiv.org, 1111.6958v1

Ikeda, H., Nagao, T., Matsuoka, K., et al. 2011, The Astrophysical Journal Letters, 728, L25

Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton University Press)

Ivezić, Ž., Lupton, R. H., Schlegel, D., et al. 2004, Astronomische Nachrichten, 325, 583

Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, The Astronomical Journal, 134, 973

Jee, M. J., Tyson, J. A., Schneider, M. D., et al. 2013, The Astrophysical Journal, 765, 74

Jiang, L., Fan, X., Cool, R. J., et al. 2006, The Astronomical Journal, 131, 2788

Jiang, L., Fan, X., Annis, J., et al. 2008, The Astronomical Journal, 135, 1057

Jiang, L., Fan, X., Bian, F., et al. 2009, The Astronomical Journal, 138, 305

—. 2014, arXiv.org, 12

Kaiser, N. 2001, Addition of Images with Varying Seeing, Tech. rep., Honolulu, Hawaii

Kantor, J., Lim, K.-T., Lupton, R. H., et al. 2015, arXiv.org, 1512.07914v1

Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, The Astrophysical Journal, 698, 895

Kirkpatrick, J. A., Schlegel, D. J., Ross, N. P., et al. 2011, The Astrophysical Journal, 743, 125

Koo, D. C., Kron, R. G., & Cudworth, K. M. 1986, Publications of the Astronomical Society of the Pacific, 98, 285

Kormendy, J., & Ho, L. C. 2013, Annual Review of Astronomy and Astrophysics, 51, 511

Kormendy, J., & Richstone, D. 1995, Annual Review of Astronomy and Astrophysics, 33, 581

Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010, The Astrophysical Journal, 708, 927

Kron, R. G., & Chiu, L. T. G. 1981, Astronomical Society of the Pacific, 93, 397

Lang, D., Hogg, D. W., Jester, S., & Rix, H.-W. 2009, The Astronomical Journal, 137, 4400

Lundgren, B. F., Wilhite, B. C., Brunner, R. J., et al. 2007, The Astrophysical Journal, 656, 73

Lupton, R., Gunn, J. E., Ivezić, Ž., Knapp, G. R., & Kent, S. 2001, Astronomical Data Analysis Software and Systems X, 238, 269

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, The Astrophysical Journal, 721, 1014

MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, The Astrophysical Journal, 728, 26

Madau, P., & Haardt, F. 2015, The Astrophysical Journal Letters, 813, L8

Madau, P., Haardt, F., & Rees, M. J. 1999, The Astrophysical Journal, 514, 648

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, The Astronomical Journal, 115, 2285

Masters, D., Capak, P., Salvato, M., et al. 2012, The Astrophysical Journal, 755, 169

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, The Astrophysical Journal, 768, 105

McQuinn, M. 2009, A&A, 704, L89

—. 2015, A&A

McQuinn, M., Oh, S. P., & Faucher-Giguère, C.-A. 2011, The Astrophysical Journal, 743, 82

Meisner, A. M., & Finkbeiner, D. P. 2014, The Astrophysical Journal, 781, 5

Merloni, A., & Heinz, S. 2008, A&A, 388, 1011

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616

Myers, A. D., Brunner, R. J., Nichol, R. C., et al. 2007, 658, 85

Myers, A. D., Richards, G. T., Brunner, R. J., et al. 2008, A&A, 678, 635

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, The Astrophysical Journal Supplement, 208, 5

Oke, J. B., & Gunn, J. E. 1983, Astrophysical Journal, 266, 713

Opitz, D., & Maclin, R. 1999, Journal of Artificial Intelligence Research, 11, 169

Page, M. J., & Carrera, F. J. 2000, Monthly Notices of the Royal Astronomical Society, 311, 433

Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, Astronomy & Astrophysics, 530, A122

Palanque-Delabrouille, N., Magneville, C., Yeche, C., et al. 2012, arXiv.org, 1209.3968v1

Pâris, I., Petitjean, P., Aubourg, E., et al. 2012, Astronomy & Astrophysics, 548, A66

Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, Astronomy & Astrophysics, 563, A54

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Peters, C. M., Richards, G. T., Myers, A. D., et al. 2015, The Astrophysical Journal, 811, 95

Pier, J. R., Munn, J. A., Hindsley, R. B., et al. 2003, The Astronomical Journal, 125, 1559

Porciani, C., Magliocchetti, M., & Norberg, P. 2004, 355, 1010

Porciani, C., & Norberg, P. 2006, 371, 1824

Postman, M., Coe, D., Benítez, N., et al. 2012, The Astrophysical Journal Supplement, 199, 25

Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992, Astrophysical Journal, 385, 404

Richards, G. T., Weinstein, M. A., Schneider, D. P., et al. 2001, The Astronomical Journal, 122, 1151

Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, The Astronomical Journal, 123, 2945

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, The Astronomical Journal, 131, 2766

Richards, G. T., Myers, A. D., Gray, A. G., et al. 2008, The Astrophysical Journal Supplement Series, 180, 67

Ross, B. C. 2014, PLoS ONE, 9, e87357

Ross, N. P., Shen, Y., Strauss, M. A., et al. 2009, A&A, 697, 1634

Ross, N. P., McGreer, I. D., White, M., et al. 2013, The Astrophysical Journal, 773, 14

Sandage, A., & Luyten, W. J. 1967, Astrophysical Journal, 148, 767

Sandage, A., Véron, P., & Wyndham, J. D. 1965, Astrophysical Journal, 142, 1307

Sanderbeck, P. R. U., D'Aloisio, A., & McQuinn, M. J. 2015, arXiv.org, 1885

Schapire, R. E. 2003, in Nonlinear Estimation and Classification (New York, NY: Springer New York), 149–171

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, The Astrophysical Journal, 500, 525

Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010, The Astrophysical Journal, 714, 1194

Schmidt, K. B., Rix, H.-W., Shields, J. C., et al. 2012, The Astrophysical Journal, 744, 147

Schmidt, M. 1968, Astrophysical Journal, 151, 393

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, The Astronomical Journal, 139, 2360

Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, The Astronomical Journal, 134, 2236

Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. 2009, The Astrophysical Journal, 690, 20

Shen, Y., & Kelly, B. C. 2012, The Astrophysical Journal, 746, 169

Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, The Astronomical Journal, 133, 2222

Shull, J. M., France, K., Danforth, C. W., Smith, B., & Tumlinson, J. 2010, 722, 1312

Smith, J. A., Tucker, D. L., Kent, S., et al. 2002, The Astronomical Journal, 123, 2121

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, The Astronomical Journal, 123, 485

Syphers, D., Anderson, S. F., Zheng, W., et al. 2011, A&A, 726, 111

Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, The Astrophysical Journal, 565, 773

Tucker, D. L., Kent, S., Richmond, M. W., et al. 2006, Astronomische Nachrichten, 327, 821

Ulrich, M. H., Maraschi, L., & Urry, C. M. 1997, 35, 445

van Dokkum, P. G. 2001, arXiv.org, 1420

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, The Astronomical Journal, 122, 549

Veale, M., White, M., & Conroy, C. 2014, Monthly Notices of the Royal Astronomical Society, 445, 1144

Vestergaard, M., & Wilkes, B. J. 2001, The Astrophysical Journal Supplement Series, 134, 1

White, M., Myers, A. D., Ross, N. P., et al. 2012, A&A, 424, 933

Willott, C. J., Delorme, P., Reylé, C., et al. 2010, The Astronomical Journal, 139, 906

Wittman, D., Tyson, J. A., Dell'Antonio, I. P., et al. 2002, arXiv.org, 73

Wolf, C., Wisotzki, L., Borch, A., et al. 2003, arXiv.org, 499

Worseck, G., Xavier Prochaska, J., McQuinn, M., et al. 2011, The Astrophysical Journal, 733, L24

Zackay, B., & Ofek, E. O. 2015, A&A, astro-ph.IM

Zu, Y., Kochanek, C. S., Kozłowski, S., & Udalski, A. 2013, The Astrophysical Journal, 765, 106