Clustering of High-Redshift Quasars

A Thesis

Submitted to the Faculty

of

Drexel University

by

John D. Timlin III

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

March 2018

Dedications

To my Family and Friends,

for love and support during my

time as a graduate student.

SDG

Acknowledgments

First, and foremost, I would like to thank my research advisor, Prof. Gordon Richards, for teaching me how to be an effective researcher and giving me the opportunity to advise students as well. Thank you for expanding my knowledge of the subject matter and providing me with research opportunities that many graduate students do not get. I would also like to thank Dr. Nic Ross who guided me through this work, and provided answers to my many questions despite living in a different time zone. Also, a big thank you to the rest of my committee – Prof. Stephen McMillan, Prof. Dave Goldberg, and Prof. Jian-Min Yuan – for your input and advise regarding both this work and for preparing me for postdoctoral jobs moving forward. I would also like to thank Prof. Adam Myers for his correspondence on the clustering work and to the students that I have advised, Jared Haughton and Andrew Pellegrino, for all of their hard work.

I am also very grateful for the support of the graduate department here at Drexel led by Prof. Michael Vogeley. The culture in this department kept me excited to come to work every day and interact with the other graduate students. I am especially grateful for the students in my class, namely Christopher Brown, Cindy Lin, and Rob Stone who helped me along this journey through both their intellectual and emotional support, as well as Jackie Moreno and Josh Wall who kept me encouraged and excited about my research. I'm also very grateful for my group mates, both past and present, who I learned a lot from through many conversations and coffee breaks. Finally, I want to thank Laura D'Angelo, Jacqueline Sampson, Janice Wilhelm for making all of this possible.

Without the love, support, and prayers of my family, I would not be in this position. Dad, Mom, Sarah thank you and I love you all very much. I am also very grateful for the love and support of my extended family, particularly my cousins Steve, Jeff, Julia, and Kevin who were on-call to hang out when I needed them. To all of you, thank you.

Table of Contents

List of Tables

List of Figures

Abstract

Clustering of High-Redshift Quasars
John D. Timlin III
Dr. Gordon Richards

In this work, we investigate the clustering of faint quasars in the early Universe and use the clustering strength to gain a better understanding of quasar feedback mechanisms and the growth of central super-massive black holes at early times in the history of the Universe. It has long been understood (e.g., Hopkins et al. 2007a) that the clustering of distant quasars can be used as a probe of different feedback models; however, until now, there was no sample of faint, high-redshift quasars with sufficient density to accurately measure the clustering strength. Therefore we conducted a new survey to increase the number density of these objects. Here, we describe the *Spitzer*-IRAC Equatorial Survey (SpIES) which is a moderately deep, large-area *Spitzer* survey which was designed to discover faint, high-redshift ($2.9 \leq z \leq 5.1$) quasars. SpIES spans $\sim$115 deg^2 in the equatorial "Stripe 82" region of the Sloan Digital Sky Survey (SDSS) and probes to 5σ depths of 6.13 μJy (21.93 AB magnitude) and 5.75 μJy (22.0 AB magnitude) at 3.6 and 4.5 microns. At these depths, SpIES is able to observe faint quasars, and we show that SpIES recovers $\sim 94\%$ of the high-redshift ($z \geq 3.5$), spectroscopically-confirmed quasars that lie within its footprint. SpIES is also ideally located on Stripe 82 for two reasons: It surrounds existing infrared data from the *Spitzer*-HETDEX Exploratory Large-area (SHELA) survey which increases the area of infrared coverage, and there is a wide range of multi-wavelength, multi-epoch ancillary data on Stripe 82 which we can use together to select high-redshift quasar candidates. To photometrically identify quasar candidates, we combined the optical data from the Sloan Digital Sky Survey and the infrared data from SpIES and SHELA and employed three machine learning algorithms. These algorithms were trained on the optical/infrared colors of known, high-redshift quasars. Using this method, we generate a sample of 1378 objects that are both faint ($i \geq 20.2$) and high-redshift ($2.9 \leq z \leq 5.1$) which we use to compute the angular two-point correlation function. We fit a single power-law model with an index

of $\delta = 1.39 \pm 0.618$ and amplitude of $\theta_0 = 0.71 \pm 0.546$ arcmin to the correlation function, as well as a dark matter model with a bias of $b = 6.78 \pm 1.79$. The bias in our investigation suggests a model of quasar feedback that considers quasar activity as an intermittent phase in galaxy evolution. If this model is correct, quasar feedback is strong enough to periodically halt the accretion of gas onto the central super-massive black hole of the quasar, which shuts down quasar activity and causes the black hole to stop growing, however it is not strong enough to completely shut down the quasar in the early Universe.

Chapter 1: Introduction

The cosmological principle assumes that the density distribution in the universe at large scales is both homogeneous and isotropic. Homogeneity is the idea that the average density of the universe is the same at any location, and isotropy means that there is no preferred direction to observe. In other words, the large scale distribution of matter is the same at any location in the universe, and will be the same no matter which direction one looks. While this is true on very large scales (hundreds of mega-parsecs), observations have shown that the density of matter on smaller scales is, in fact, quite clustered. For example, Figure 1.1[1] shows that the observed galaxy distribution in the Sloan Digital Sky Survey (SDSS; York et al. 2000), as represented as green and red points, tend to lie in large groups and along filaments that connect the groups. So, while at large scales the universe may be uniform, at smaller scales it certainly is not.

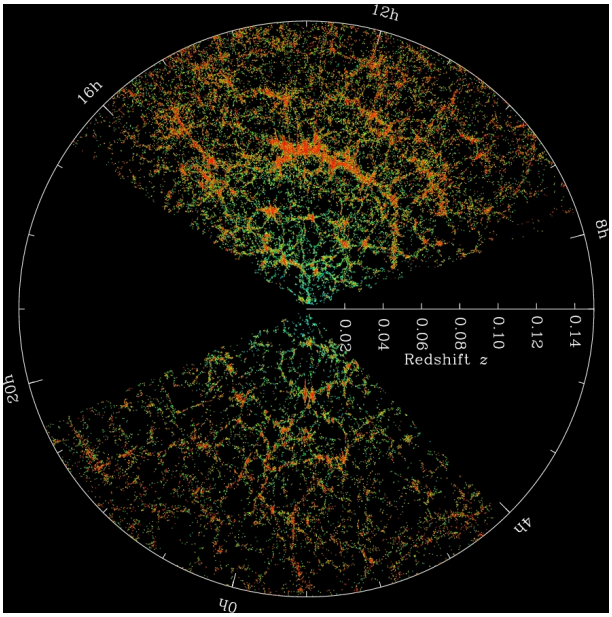

Figure 1.1 Observed distribution of galaxies in the local universe from SDSS. Each point depicts the location of a spectroscopically-confirmed galaxy. This image shows that the distribution of galaxies is not homoegenous on small scales, and matter tends to cluster together. Images from surveys like this fueled new theories about the large scale evolution of the universe. Image Credit: M. Blanton and SDSS

Observations of large scale structure, like the one from SDSS, provide a benchmark for theories

[1] http://www.sdss.org/science/orangepie/

and simulations in the pursuit to understand how the universe evolved from early times to the current epoch. The prevailing theory of universal evolution suggests that, instead of having a uniform density everywhere, the dark matter distribution instead harbored regions of slight over-densities in the early universe. Over time, and under the influence of gravity, these over-dense regions grew hierarchically, meaning as the small over-densities grew, they became more massive which, in turn, attracted more matter. Over time, the small perturbations grew into the over-dense regions observed today. The Millennium Simulation (Springel et al., 2005), which is a large simulation designed to track the evolution of dark matter and galaxies over cosmic time, has shown that this is a plausible solution. Figure 1.2[2] depicts the simulated dark-matter distribution at various epochs. As the age of the universe increases (looking from left to right), the relatively smooth universe becomes clumpy. Moreover, the results of this simulation at the current epoch show that the distribution of visible matter traces the distribution of the underlying dark matter, as depicted in Figure 1.3. Visually, the two distributions look identical, but to properly compare the observed galaxy distribution and the underlying dark matter distribution over cosmic time, we must first quantify the 'clumpiness' of these two distributions.

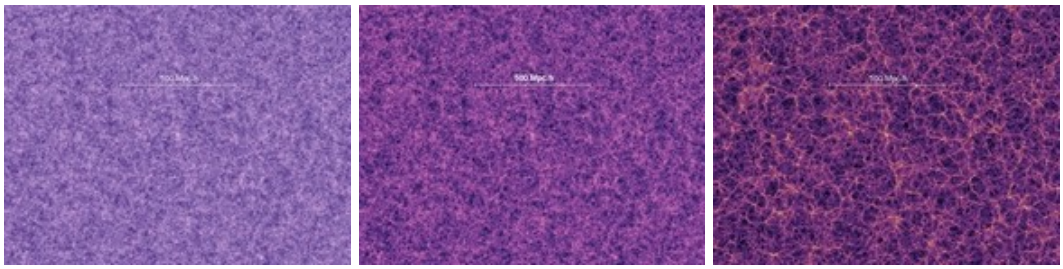

Figure 1.2 Evolution of the dark matter density field over time from the Millennium Simulation, where the age of the universe increases from the left panel to the right panel. Over this period of time, the universe goes from relatively smooth (left) to more noticably more clustered (right). Image Credit: The Millennium Simulation Team (`https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/`)

One of the most fundamental measurements used to measure the degree of 'clumpiness' of both the dark matter and visible matter distribution is the two-point correlation function. This statistic measures the excess probability, above random, of finding a pair of objects separated by a particular distance (Totsuji & Kihara 1969; Peebles 1980). Mathematically, the correlation function is defined

[2]https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/

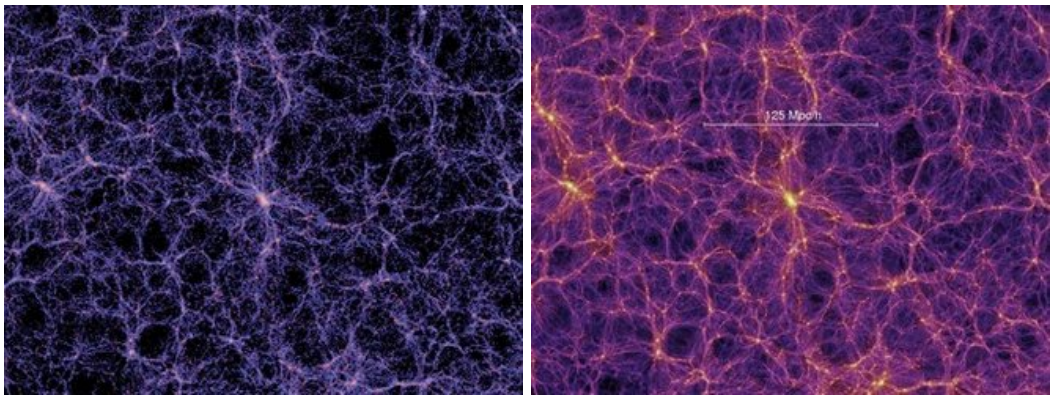

Figure 1.3 Distribution of visible matter (left) and dark matter (right) in the Millennium Simulation. This simulation shows that the distribtion of the visible matter traces the density of the underlying dark matter in the current universe. Measurements of the distribtuion of visible matter must be made over cosmic time to track the evolution of large scale structure. Image Credit: The Millennium Simulation Team (`https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/`)

as the inner product of the density perturbations in an underlying density field. For a homogeneous field (i.e. there is no perturbation in the density), the correlation function is equal to zero. With the results from simulations such as the aforementioned Millenium Simulation, which assumes a flat universe with cold dark matter and a cosmological constant, the correlation function of the dark matter can be computed at any epoch using the mathematical definition directly. Computing the correlation function of matter, namely galaxies, however is slightly different. The process by which galaxies form and evolve is still uncertain, but we know that gravity is not the only force acting on the system. The baryonic matter in the galaxy interacts electromagnetically with its surroundings which, in turn, effects the dynamics of the galaxy throughout its evolution. Due to these other forces, there is no particular reason why the galaxy distribution should trace the underlying dark matter (Liddle & Lyth, 2000). To rectify this problem, Kaiser (1984) introduced the idea that massive galaxies are rare objects that lie only at the peaks of the dark matter distribution, and thus can be considered biased tracers of the underlying dark matter. Therefore, despite the complex physics that happens during galaxy formation and evolution, the clustering of these galaxies differs from the clustering of dark matter by only a scale factor called the clustering bias. This bias parameter can be estimated, then, by comparing the theoretical prediction of the dark matter correlation function with an accurate measurement of the galaxy correlation function at the same epoch.

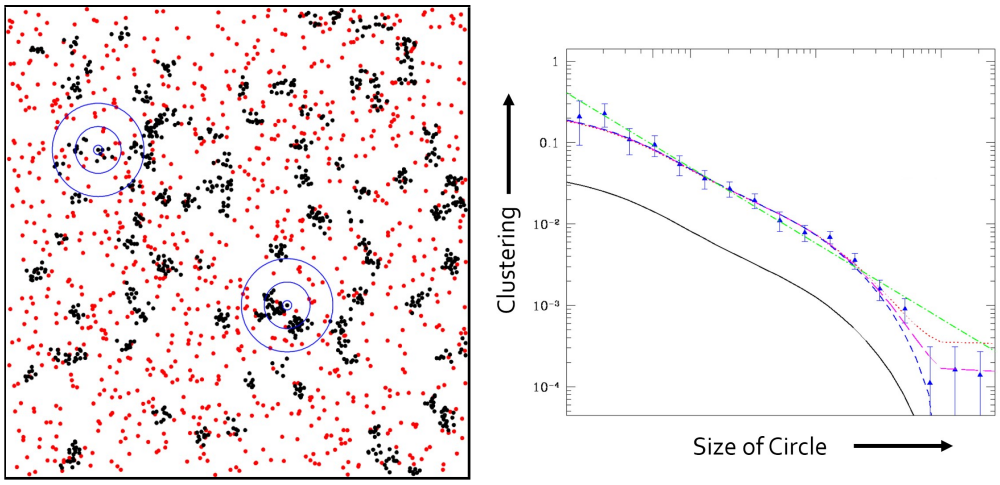

Figure 1.4 Left: Mock data representing a random sample of objects in the sky (red) and a highly clustered set of data (black). To estimate the correlation function, pairs of points are counted inside of circles of increasing radii (as represented by the blue circles). Right: Example clustering signal from Myers et al. (2007) which we use to represent the clustering of the black points in the left hand panel. We show that as the size of the circle increases, the strength of the clustering signal decreases. This implies that, at larger scales, the distribution becomes more homoegeneous. Image credit: Adam Myers

Estimating the correlation function of a population of objects is different than measuring that of dark matter because we cannot directly measure density perturbations from the visible matter. We can, however, measure the position of galaxies in space and use different estimators to compute the correlation function. These estimators require that we compare the positions of the galaxy with a distribution of random points. A toy example of this scenario is depicted in the left panel of Figure 1.4, where we show a distribution of random data (red points), which exhibit no clustering signal, and a distribution of clustered data (black points). To estimate the correlation function of the clustered data, one counts the number of pairs of data and random points inside annuli of increasing radius, as represented by the blue circles. The correlation function is, then, a linear combination of these different pair counts. An example correlation function from Myers et al. (2007) is presented in the right hand panel of Figure 1.4. From this Figure we see that as the radius of the circle increases, the clustering amplitude decreases. This measurement implies that, as the scale grows, the observed distribution of visible matter becomes more homogeneous. Since the estimators only require the positional information, the correlation function can be computed for a variety of astronomical objects such as stars, galaxies, quasars, and even the cosmic microwave background.

Despite being a simple statistic that reports whether or not a distribution is random, the correlation function can can have many different physical implications depending on which population is being analyzed.

In this work, we are particularly interested in studying the clustering of quasars, which are actively accreting super-massive black holes at the center of massive galaxies (Salpeter 1964; Lynden-Bell 1969; Rees 1984). The widely accepted toy model of quasars, as set forth by Antonucci (1993), depicts the average quasar having the following properties: a central super-massive black hole surrounded by an accretion disk which shines from X-ray to optical wavelengths, both broad and narrow line regions responsible for spectral emission lines that can be used to estimate systemic redshift (or distance from the observer), and a dusty region which absorbs light from the disk and reprocesses it into the infrared wavelengths. Radiation from the accretion disk is the result of frictional heating of the gas, spiraling at speeds fractional to the speed of light around the black hole. This friction heats the disk to temperatures hot enough to produce X-ray radiation near the center of the disk, where the gas travels the fastest, as well as ultraviolet and optical light further out. The radiation from the disk also interacts with the surrounding dust, which absorbs and reprocesses the light into infrared wavelengths. Quasars are unique in they they brightly shine at nearly every wavelength in the electromagnetic spectrum. As a result, quasars are some of the most luminous objects in the universe, and thus can be seen out to great distances. Since light travels at a finite speed, observing distant objects is equivalent to looking at the universe earlier in its history. Also, since quasars live at the centers of massive galaxies, they reside in the peaks in the dark matter distribution, and are thus biased tracers of the underlying dark matter (Sheth & Tormen, 1999). Therefore, quasars are excellent probes of the structure in the early universe.

In the early universe, the majority of luminous quasars (such as cataloged by SDSS) are born due to the major merger of two massive, gas-rich galaxies. Figure 1.5 depicts a series of snapshots of a simulation designed to track this type of merger (image courtesy Phil Hopkins). The simulation begins in the top left panel, which depicts two galaxies of approximately the same mass that are gravitationally bound. As the galaxies pass through each other, the gas begins to heat (signified

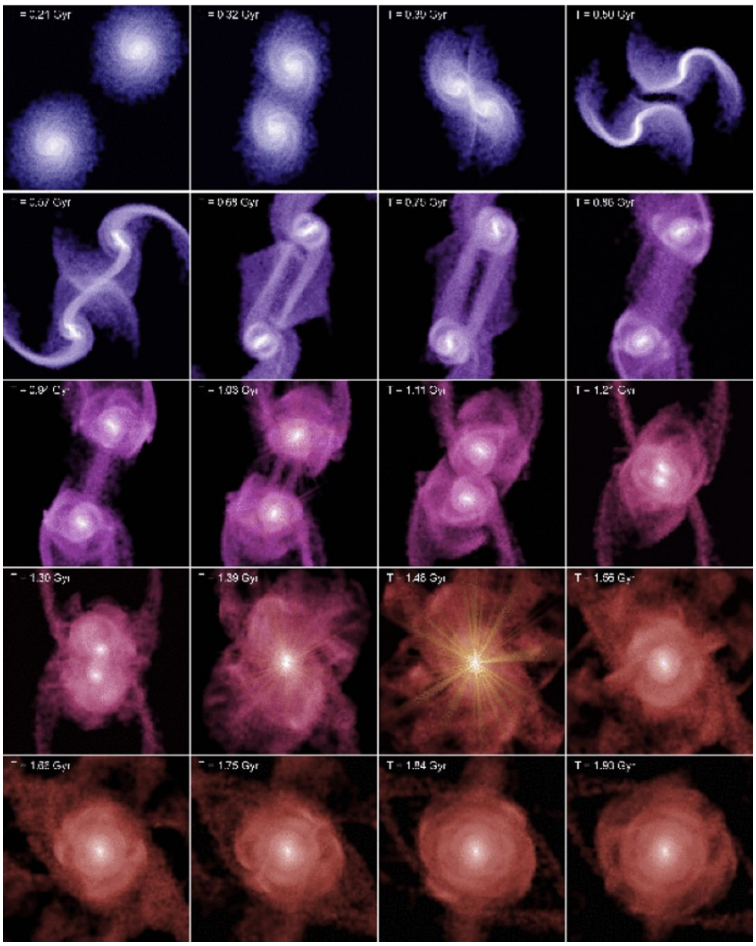

Figure 1.5 Simulation of a pair of galaxies undergoing a major merger. The simulation begins in the top left panel with two gas-rich galaxies falling towards each other due to gravity. As time evolves, the galaxies undergo a major merger, which heats the gas (indicated by the red color) and causes the gas to become perturbed in the system. As the central black holes of the galaxies merge, the gas begins to fall into the central region. This gas is then accreted by the new black hole at the center of the merged galaxy which triggers a luminous quasar (shown in the panel in the fourth row, third column). The radiation from the central quasar interacts with the surrounding gas, pushing it out of the gravitational potential well, which stops accretion and ends the quasar activity. Image credit: Phil Hopkins.

by a change in color from blue to red) and becomes perturbed. Over time, the supermassive black holes of these two galaxies merge forming a single black hole, and the perturbed gas begins to fall in toward the center of the newly formed galaxy. The center of the galaxy begins to shine brightly, due to the gas being accreted onto the supermassive black hole, forming a quasar (shown in the fourth row of Figure 1.5). It is in this quasar phase where the supermassive black hole grows significantly because of the gas that is falling into the center. If the quasar shines bright enough, the radiation from the central source will interact with the surrounding gas in a process called quasar feedback.

When the radiation pressure on the gas is large enough to push the gas out of the potential well of the black hole, accretion ceases and the quasar shuts down (shown in the bottom row of Figure 1.5). Quasar feedback, therefore, is a way for the quasar to self-regulate: the brighter it shines, the larger the pressure will be on the surrounding gas blowing it out of the central region. If there is no more gas in the central region for the black hole to accrete, the black hole no longer grows: therefore quasar feedback can also regulate black hole growth over time.

While quasar feedback is a well accepted idea, there are still questions about how big of a role this plays in the evolution of the quasar and of the host galaxy, particularly in the early universe. To address this question, Hopkins et al. (2007a) sought to test different feedback models in large simulations, and used the clustering strength (bias) as the metric to compare to observations. Their models, shown in Figure 1.6, were designed to align with the known quasar clustering results from observational studies of low-redshift quasars, and predict the clustering strength of different models at high-redshift. The left panel of Figure 1.6 depicts the strongest feedback model where the radiation from the quasar is powerful enough to completely blow the gas from the central region, implying that the quasar is a singular phase in the evolution of the galaxy. The middle panel presents the standard feedback model, which suggests that feedback is powerful enough to push gas away from the black hole, but not completely, and thus the gas will eventually fall back in to the center. This model predicts that the quasar phase in a galaxy is intermittent, turning off only when the quasar shines brightly. Finally, the right panel of Figure 1.6 shows the model where feedback lags, or feedback is inefficient at blowing gas away from the central source until the observed epoch of downsizing begins. While the solid lines in Figure 1.6 predict three divergent clustering results, the dotted and dot-dashed lines predict degenerate results for observations that do not sample the faint end of the quasar distribution. If, for example, the observed quasars are brighter than magnitude $i = 20.2$, all of the clustering results at high-redshift will follow the dotted line. However, if one creates a sample of quasars that breaks this degeneracy between $3 \leq z \leq 4$, then a measurement of the quasar correlation function can constrain these three divergent models.

In this work we present the first measurement of the two-point autocorrelation function of faint

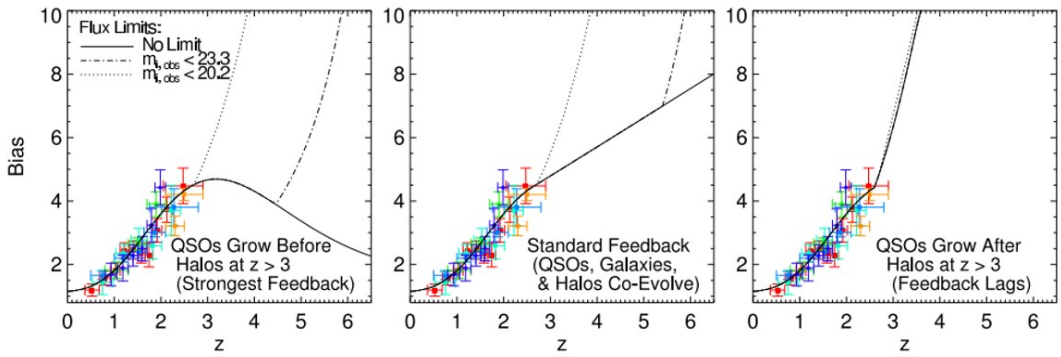

Figure 1.6 Prediction of the clustering strength as a function of redshift for three different feedback models from Hopkins et al. (2007a). The solid line in the left panel shows the predicted bias for quasars that exhibit 'strong feedback' which is capable of blowing the gas from the central region and shutting off the quasar. The same line in the center panel depicts the bias for quasars with 'standard feedback', where the radiation is strong enough to push the gas from the center, but not strong enough to completely remove the gas from the potential well of the black hole. In this model the quasar intermittently turns on and off at high-redshift. The solid line in the right panel shows a model in which 'feedback lags', which means that the radiation is not strong enough to blow the gas away from the central region. These models were designed to fit observations at low-redshift (depicted by the points), however diverge at higher-redshift. Additionally, the dotted and dot-dashed lines depict the predicted bias from the feedback models if the quasar sample used to compute the bias only contains bright quasars. To break this degeneracy, the clustering of faint, high-redshift quasars must be computed.

quasars ($i \leq 20.2$) at high-redshift ($2.9 \leq z \leq 5.1$) to test these quasar feedback models from Hopkins et al. (2007a). Until now, there has not been a comprehensive quasar sample with a sufficient amount of objects that satisfied all of the required magnitude and redshift constraints to break the degeneracy. In Chapter 2 we describe the technical details of a survey we conducted using the *Spitzer* Space Telescope, that was specially designed to help discover faint quasars over a wide area on the sky. Included in this section is an overview of the Stripe 82 field, our image stacking method, and we compare to our study to other infrared surveys in this field. We also describe our source extraction method, give estimates of the depth of the survey, and discuss our reported astrometry and photometry, including a discussion of the reported photometric errors. Finally, we present our full photometric catalog and test how robust the catalog is at detecting high-redshift quasars. The infrared data and optical data are combined in Chapter 3 where we discuss photometric selection algorithms which we use to classify quasars in these catalogs. We present the machine learning algorithms used in this study to select faint, high-redshift quasars, as well as the catalog of quasars we used to compute the correlation function. In Chapter 3 we also provide a more rigorous definition of the correlation function and how we compute the clustering strengths of

our quasar set. Finally, we compare our clustering result to other recent results, we describe the feedback models from Hopkins et al. (2007a), and present our final clustering result. We conclude in Chapter 4 with a summary of results and a discussion of future work that can be done with this data set.

Chapter 2: The *Spitzer* IRAC Equatorial Survey

Abstract

We describe the first data release from the *Spitzer*-IRAC Equatorial Survey (SpIES); a large-area survey of ∼115 deg^2 in the Equatorial SDSS Stripe 82 field using *Spitzer* during its 'warm' mission phase. SpIES was designed to probe sufficient volume to perform measurements of quasar clustering and the luminosity function at $z \geq 3$ to test various models for "feedback" from active galactic nuclei (AGN). Additionally, the wide range of available multi-wavelength, multi-epoch ancillary data enables SpIES to identify both high-redshift ($z \geq 5$) quasars as well as obscured quasars missed by optical surveys. SpIES achieves 5σ depths of 6.13 μJy (21.93 AB magnitude) and 5.75 μJy (22.0 AB magnitude) at 3.6 and 4.5 microns, respectively—depths significantly fainter than *WISE*. We show that the SpIES survey recovers a much larger fraction of spectroscopically-confirmed quasars (∼98%) in Stripe 82 than are recovered by *WISE* (∼55%). This depth is especially powerful at high-redshift ($z \geq 3.5$), where SpIES recovers 94% of confirmed quasars, whereas *WISE* only recovers 25%. Here we define the SpIES survey parameters and describe the image processing, source extraction, and catalog production methods used to analyze the SpIES data. In addition to this survey paper, we release 234 images created by the SpIES team and three detection catalogs: a 3.6 μm-only detection catalog containing ∼6.1 million sources, a 4.5 μm-only detection catalog containing ∼6.5 million sources, and a dual-band detection catalog containing ∼5.4 million sources.

2.1 Introduction

The *Spitzer* Space Telescope (Werner et al., 2004) has been paramount in understanding the Universe at mid-infrared wavelengths. During its primary mission, *Spitzer* observed at 3.6, 4.5, 5.8, and 8.0 μm using the Infrared Array Camera (IRAC; Fazio et al. 2004), at 24, 70, and 160 μm using the Multiband Imaging Photometer for *Spitzer* (MIPS; Rieke et al. 2004) camera, and had a dedicated infrared spectrograph (IRS; Houck et al. 2004) covering wavelengths from 5.3 to 38 μm. Since the

exhaustion of its cryogen in 2009, *Spitzer* has run its 'warm' mission phase, taking images with the two shortest IRAC passbands (3.6 and 4.5 μm).

Spitzer IRAC has been a valuable tool in the creation of deep, relatively small area surveys through campaigns like the $\sim$2 deg^2 Spitzer-COSMOS survey (S-COSMOS; Sanders et al. 2007) and the $\sim$10 deg^2 *Spitzer* Deep, Wide-field Survey (SDWFS; Ashby et al. 2009) utilizing all four of the IRAC bands. *Spitzer* continues to delve deeper in its 'warm' phase with the IRAC ultradeep filed (IUDF; Labbe et al. 2015), the $\sim$1.2 deg^2 *Spitzer* Large Area Survey with Hyper-Suprime-Cam (SPLASH; Steinhardt et al. 2014), and the $\sim$18 deg^2 *Spitzer* Extragalactic Representative Volume Survey (SERVS; Mauduit et al. 2012).

Despite having a relatively small 5.′2×5.′2 field of view (FOV), IRAC has also effectively and efficiently run larger-area programs throughout its lifetime such as the $\sim$65 deg^2 SIRTF Wide-Area Infrared Extragalactic Survey (SWIRE; Lonsdale et al. 2003). Recently, *Spitzer* has made an effort to run larger-area surveys in the 'warm' phase with the $\sim$26 deg^2 *Spitzer*-HETDEX Exploratory Large Area (SHELA; Papovich et al. 2016) and the $\sim$94 deg^2 *Spitzer* South Pole Telescope Deep Field (SSDF; Ashby et al. 2013) mission which, until now, had the largest area of any *Spitzer* survey.

These large-area campaigns are made possible by the IRAC mapping mode strategy, which aligns the arrays on a positional grid, allowing observations to overlap through successive motions in the grid. This approach differs from other observing strategies, many of which forced the telescope to slew to a single position multiple times to observe the same location on the sky in a different channel (see Section 3.2 of the IRAC Instrument Handbook[1]). Mapping mode decreases slew time, allowing for larger area surveys to be performed while still reaching interesting flux limits.

Spitzer is not the only telescope performing large area, mid-infrared observations of the Universe. The Wide-field Infrared Survey Explorer (*WISE*; Wright et al. 2010) telescope has been mapping the entire sky in four channels, two of which have nearly the same wavelength as 'warm' *Spitzer* (3.4 and 4.6 μm). While *WISE* covers essentially the entire sky, it lacks both the depth and the spatial resolution that *Spitzer* IRAC surveys can achieve.

[1] http://irsa.ipac.caltech.edu/data/SPITZER /docs/irac/iracinstrumenthandbook/

Table 2.1. The *Spitzer* IRAC Equatorial Survey (SpIES) key parameters

Parameter	Value
Imaging	IRAC Ch1 and Ch2
Wavelength	3.6 and 4.5 μm
Area[a]	$\sim$115 deg^2
No. of IRAC pointings	$\sim$70,000
Exposure Time at each pointing	60s
Total Observation Time	820hr
Typical Zodiacal Background	$0.09 - 0.23$ MJy sr^{-1}
IRAC PSF FWHM[b]	$1\farcs95$, $2\farcs02$
Total number of objects[c]	$\sim$5,400,000
Limiting AB Magnitude[d] (5σ)	21.93, 22.0
Data URL: `http://www.physics.drexel.edu/~gtr/spies/`	

Note. — [a] Total survey area covered by both detectors. The area covered by a single detector decreases due to their separation on IRAC (details in Section 2.3). [b]5σ dual-band detection catalog (see Section 2.5). [c]Total number of objects in the dual-band catalog. [d]Values are for the 3.6 μm, 4.5 μm detectors.

In this paper, we describe the *Spitzer* IRAC Equatorial Survey (SpIES) parameters and catalogs. SpIES mapped a large portion of the Sloan Digital Sky Survey (SDSS; York et al. 2000) equatorial S82 field (Stoughton et al. 2002; Annis et al. 2014; Jiang et al. 2014), utilizing the *Spitzer* 3.6 and 4.5 μm bands (often referred to as Ch1 and Ch2 respectively). Collecting $\sim$115 deg^2 over $\sim$820 hours, SpIES is the largest area *Spitzer* survey, probing to depths comparable to SWIRE. Table 2.1 contains the key parameters of SpIES such as the wavelengths and point spread function of IRAC, along with the observation times, area, and depth of the SpIES survey. With this release, we present three SpIES source catalogs consisting of $\sim$6.1 million objects detected only at 3.6 μm, $\sim$6.6 million objects detected only at 4.5 μm, and a dual-band detection catalog which contains $\sim$5.4 million detections in both bands. We also release the images generated by the SpIES team used to build the catalogs described herein.

The combined depth and area of the SpIES, along with the wealth of multi-wavelength, multi-epoch ancillary imaging and spectroscopic data on Stripe 82 (S82; Stoughton et al. 2002; Annis

et al. 2014; Jiang et al. 2014), make it a powerful tool for addressing a wide range of topics in contemporary astrophysics. In particular, we seek to use the data to: probe the population of obscured quasars at high redshift (e.g., Alexandroff et al. 2013; Glikman et al. 2013; Assef et al. 2015); use high-redshift unobscured quasars to investigate how quasar feedback contributes to galaxy evolution (e.g., Hopkins et al. 2007b; White et al. 2012); improve the removal of foreground objects from maps of the cosmic microwave background (Wang et al., 2006); better constrain the stellar masses of Lyman Break Galaxies (e.g., Daddi et al. 2007); improve stellar population modeling for hosts of supernovae (e.g., Sullivan et al. 2010; Fox et al. 2015); and enable discovery of cool stars (e.g., Lucas et al. 2010).

We begin our discussion by describing the existing data covering the S82 footprint in Section 2.2, followed by the *Spitzer* observation strategy used for SpIES in Section 2.3. We discuss the data products from *Spitzer* and our image stacking process in Section 2.4. The SpIES catalogs are described Section 2.5, which includes source extraction techniques, photometric errors, and astrometric reliability. This section also discusses the completeness, number counts, and depth of the SpIES detection catalog. Finally, in Section 2.6, we match SpIES objects to various quasar catalogs to test the SpIES recovery fraction of high-redshift quasars. We also provide a summary of the SpIES survey and links to the data products in Appendix 2.7.

We calculate magnitudes on the AB scale, which has a flux density zeropoint of 3631Jy (Oke & Gunn, 1983a). These are denoted as [3.6] and [4.5], respectively. Conversion to Vega magnitudes is given by [3.6]-2.779 and [4.5]-3.264, respectively (calculated using the Vega zeropoint flux density values of 280.9 Jy at 3.6 μm and 179.7 Jy at 4.5 μm from Table 4.1 in the IRAC Handbook[1]).

2.2 The Stripe 82 Region

The observational goal of the SpIES project was to map S82 in order to provide a suitably large "laboratory" in which to conduct the types of experiments that involve rare objects, as noted above. S82 is located on the Celestial Equator spanning a range of $-60^\circ \leq \alpha \leq 60^\circ$ and $-1.25^\circ \leq \delta \leq 1.25^\circ$. The SpIES observations cover approximately one third of this region centered on $\delta = 0^\circ$ and spanning the range from $-30^\circ \leq \alpha \leq 35^\circ$, with a break in coverage between $13.9^\circ \leq \alpha \leq 27.2^\circ$ where deeper

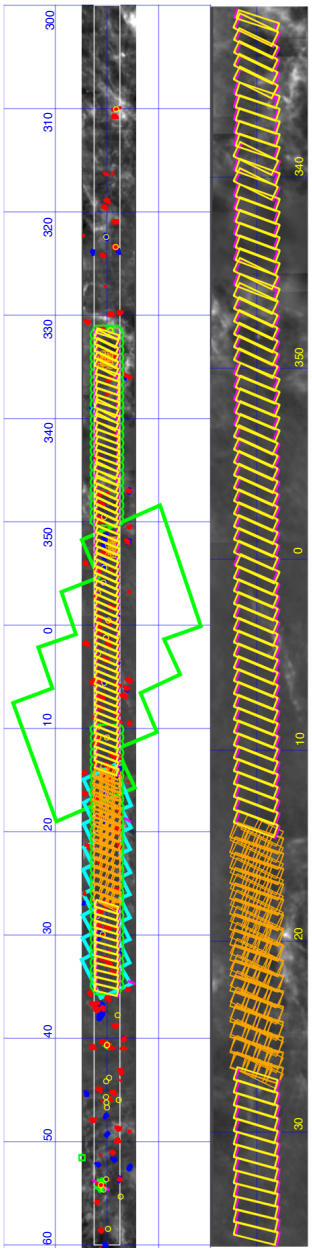

Figure 2.1 Top: We show the SpIES coverage area (yellow and purple rectangles) atop the 100μm IRAS dust map (Schlegel et al., 1998) of the full SDSS Stripe 82 region (white box). Many different surveys have covered this region of the sky and overlap with SpIES. Displayed are the HeLMS (green box) and HeRS (light blue) survey footprints (Oliver et al., 2012; Viero et al., 2014), the regions observed by *XMM-Newton* (yellow and orange circles) and *Chandra* (red circles were observed with the ACIS-S arrays and blue circles with the ACIS-I arrays; LaMassa et al. 2013a,b), the VLA (green scallop) from Hodge et al. (2011), and the SHELA observations (orange boxes) by Papovich et al. (2016), as a few examples of many surveys that cover the S82 region. More details about other surveys on S82 can be found in Table 2.2. Bottom: Detailed SpIES 3.6 μm (yellow) and 4.5 μm (purple) coverage of Stripe 82 along with SHELA coverage (orange). Both panels are centered on δ=0 and α values are given in J2000 degrees.

IRAC data exists from the SHELA (Papovich et al., 2016) survey. Within those RA limits, SpIES completely covers S82 from $-0.85° \leq \delta \leq 0.85°$ with irregular coverage outside of that declination range due to the orientation of observations (see Figure 2.1). The SpIES footprint was chosen to take advantage of the SHELA footprint and for its relatively low background at mid-infrared wavelengths. As described in more detail in Section 2.5.5, background noise can drastically decrease the depth of the survey, which makes observing the faintest sources prohibitively difficult.

SDSS observed S82 in five optical filters (*ugriz*; Fukugita et al. 1996) to find variable objects and to obtain deeper imaging than the wider-area SDSS observations in the Northern Galactic Cap (York et al. 2000; Frieman et al. 2008; Annis et al. 2014). SDSS-I/II observed the full S82 field ∼80 times over 8 years resulting in photometry which reaches nearly two magnitudes fainter than the other fields in the survey (Annis et al. 2014, Jiang et al. 2014). S82 has also been observed multiple times with the SDSS spectrographs (Smee et al., 2013) as part of the SDSS-I/II (York et al., 2000) and SDSS-III/BOSS (Eisenstein et al., 2011) campaigns, along with spectra from other facilities such as 2dF, 6dF, and AUS (Croom et al. 2004, 2009), WiggleZ (Drinkwater et al., 2010), the Virmos-VLT Deep Survey (VVDS; Le Fèvre et al. 2005), the VIMOS Public Extragalactic Redshift Survey (VIPERS1; de la Torre et al. 2013), DEEP2 (Davis et al., 2007), and the Prism Multi-Object Survey (PRIMUS; Coil et al. 2011). In total these facilities have collected ∼125,000 high quality spectra across its entire area.

In addition to the collection of deep SDSS optical imaging (reaching a 5σ AB magnitude of 24.6 in the r-band) and spectra, S82 contains a vast amount of multi-wavelength imaging taken over many epochs. The two panels of Figure 2.1 show several multi-wavelength surveys that overlap with the SpIES region. At radio wavelengths, in addition to full coverage by the Faint Images of the Radio Sky at Twenty-centimeters (FIRST; Becker et al. 1995, Helfand et al. 2015) survey, Hodge et al. (2011) provided $1\farcs8$ resolution data down to 52μJy at 1.4GHz (L-band) over ∼90 deg^2 of Stripe 82 (twice the resolution and three times the depth of FIRST). Additional radio data will be forthcoming at lower resolution (e.g., Jarvis et al. 2014) and at higher frequency (Mooley et al., 2014).

Table 2.2. Deep imaging data available on Stripe 82

Waveband $\lambda_{\rm eff}$ (μm)	Origin	Depth	Coverage (deg^2)	Reference
2-10 keV	XMM-*Newton*	$4.7\times^{-15}$ erg s^{-1} cm^{-2}	31.3[a]	LaMassa et al. (2015)
0.5-2 keV	XMM-*Newton*	$8.7\times^{-16}$ erg s^{-1} cm^{-2}	31.3[a]	LaMassa et al. (2015)
FUV, 1350–1750 Å	GALEX	$m_{\rm AB} \simeq 23$	∼200	Martin et al. (2005)
NUV, 1750–2750 Å	GALEX	$m_{\rm AB} \simeq 23$	∼200	Martin et al. (2005)
0.355 (u)	SDSS	$m_{\rm AB} = 23.90$	∼300	Jiang et al. (2014)
0.5 (g)	SDSS	$m_{\rm AB} = 25.10$	∼300	Jiang et al. (2014)
	HSC[b]	$m_{\rm AB} = 26.50$	∼300	Miyazaki et al.
	DES	$m_{\rm AB} = 26.50$	∼300	Diehl et al. (2014)
0.6 (r)	SDSS	$m_{\rm AB} = 24,60$	∼300	Jiang et al. (2014)
	HSC[b]	$m_{\rm AB} = 26.10$	∼300	Miyazaki et al.
	DES	$m_{\rm AB} = 26.00$	∼300	Diehl et al. (2014)
0.7 (i)	SDSS	$m_{\rm AB} = 24.10$	∼300	Jiang et al. (2014)
	HSC[b]	$m_{\rm AB} = 25.90$	∼300	Miyazaki et al.
	CS82	$m_{\rm AB} = 24.00$	∼170	Kneib et al. in prep.
	DES	$m_{\rm AB} = 25.30$	∼300	Diehl et al. (2014)
0.9 (z)	SDSS	$m_{\rm AB} = 22.80$	∼300	Jiang et al. (2014)
	HSC[b]	$m_{\rm AB} = 25.10$	∼300	Miyazaki et al.
	DES	$m_{\rm AB} = 24.70$	∼300	Diehl et al. (2014)
1.00 (Y)	ULAS[c]	$m_{\rm AB} = 20.93$	277.5	Lawrence et al. (2007)
	HSC[b]	$m_{\rm AB} = 24.40$	∼300	Miyazaki et al.
	DES	$m_{\rm AB} = 23.00$	∼300	Diehl et al. (2014)
	VHS	$m_{\rm AB} = 21.20$	∼300	McMahon et al. (2013)

Table 2.2 (cont'd)

Waveband $\lambda_{\rm eff}$ (μm)	Origin	Depth	Coverage (deg^2)	Reference
1.35 (J)	ULAS[c]	$m_{\rm AB} = 20.44$, 24 μJy	277.5	Lawrence et al. (2007)
	VICS82,	$m_{\rm AB} = 22.70$	150	Geach et al. in prep.
	VHS	$m_{\rm AB} = 22.20$	$\sim$300	McMahon et al. (2013)
1.65 (H)	ULAS[c]	$m_{\rm AB} = 19.98$, 37 μJy	277.5	Lawrence et al. (2007)
	VHS	$m_{\rm AB} = 20.60$	$\sim$300	McMahon et al. (2013)
2.20 (K_s)	ULAS[c]	$m_{\rm AB} = 20.10$, 33 μJy	277.5	Lawrence et al. (2007)
	VICS82	$m_{\rm AB} = 21.60$	150	Geach et al. in prep.
	VHS	$m_{\rm AB} = 21.50$	$\sim$300	McMahon et al. (2013)
3.6 (Ch1)	**SpIES**	$m_{\rm AB} = 21.90$	$\sim$115	**this paper**
	SHELA	$m_{\rm AB} = 22.05$	$\sim$26	Papovich et al. (2016)
4.5 (Ch2)	**SpIES**	$m_{\rm AB} = 22.00$	$\sim$115	**this paper**
	SHELA	$m_{\rm AB} = 22.05$	$\sim$26	Papovich et al. (2016)
250	*Hershel*/SPIRE	64.0, 64.0 mJy	270, 79	Oliver et al. (2012); Viero et al. (2014)
350	*Hershel*/SPIRE	64.5, 64.5 mJy	270, 79	Oliver et al. (2012); Viero et al. (2014)
500	*Hershel*/SPIRE	74.0, 74.0 mJy	270, 79	Oliver et al. (2012); Viero et al. (2014)
1100 (277 GHz)	ACT[d]	$\sim$6.4 mJy	300	analysis under way
1400 (218 GHz)	ACT[d]	$\sim$3.3 mJy	300	Gralla et al. (2014); Das et al. (2014)
2000 (148 GHz)	ACT[d]	$\sim$2.2 mJy	300	Gralla et al. (2014); Das et al. (2014)
21,000 (L-band)	VLA[e]	260 μJy	92	Hodge et al. (2011)
30,000 (S-band)	VLA[e]	400 μJy	$\sim$300	Mooley et al. (2014)

Note. — [a]Includes 7.4 deg^2 of archival *Chandra* data, [b]Hyper Suprime-Cam (see `http://www.naoj.org/Projects/HSC/surveyplan.html` for more details), [c]UKIDSS Large Area Survey, [d]Atacama Cosmology Telescope, [e]Very Large Array

In the far-infrared, the *Herschel Space Observatory* performed the HerMES Large Mode Survey (HeLMS; Oliver et al. 2012) and the Herschel Stripe 82 Survey (HerS; Viero et al. 2014) to study galaxy formation and correlations between galaxies and dark matter haloes. Existing mid-infrared observations of S82 include SHELA (Papovich et al., 2016), which contains deep imaging data for dark energy measurements, and the AllWISE observations from *WISE* (Wright et al., 2010). Near-infrared measurements of S82 have been performed by the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007), the VISTA Hemisphere Survey (VHS; McMahon et al. 2013)—which is matched to the SDSS coadd photometry in the catalog presented in Bundy et al. (2015)—and the deeper J- and K-band coverage from the VISTA-CFHT Stripe 82 Survey over 130 deg^2 of S82 (VICS82; Geach et al. in prep.). In addition to SDSS, Stripe 82 has high-resolution imaging (median

seeing of $0\farcs6$) from the CFHT Stripe 82 Survey (CS82; Kneib et al. in prep.) and is part of the Dark Energy Survey[2] (DES) footprint.

S82 was also mapped in the ultraviolet as part of the *GALEX* All-sky Imaging Survey and Medium Imaging Survey, and a few locations were imaged with the Deep Imaging Survey as outlined in Martin et al. (2005). *Chandra* and *XMM-Newton* have been used to observe partly contiguous regions over a wide area at X-ray wavelengths, searching for high luminosity quasars (LaMassa et al. 2013a,b), with the most recent large-area X-ray catalog release covering $\sim$31deg^2 with *XMM-Newton* (LaMassa et al., 2015). More observations are cited in Table 2.2 which lists some properties of the deepest imaging data of S82 at various wavelengths. The combination of all of the multi-epoch, multi-wavelength spectroscopic and photometric data on S82 provides a powerful tool to aid in our understanding of the Universe by painting a multi-wavelength and multi-epoch picture of matched objects between these surveys.

2.3 Data Acquisition

SpIES data were obtained as part of Cycle 9 (2012-2014) of the *Spitzer* 'warm' post-cryogenic mission utilizing the first two channels of IRAC. IRAC is a wide-field camera with four channels, each 256×256 pixels with a $5\farcm2\times5\farcm2$ field of view (Fazio et al., 2004). The first two arrays (3.6 and 4.5 microns) are designed to observe the sky simultaneously, which decreases observation time and ensures that the epochs of measurement are roughly the same for both channels. *Spitzer* has been operating in 'warm' mode long enough to measure and report the differences in IRAC performance between the cryogenic and 'warm' observations[3]. The changes in performance, including changes in PSF, sensitivity levels, and constant values such as gain and flux conversion, are minor and the overall performance of IRAC has not degraded substantially with time (see Mauduit et al. 2012).

The SpIES observation strategy was motivated by the strategies of previous *Spitzer* campaigns such as SDWFS (Ashby et al., 2009), SWIRE (Lonsdale et al., 2003), SERVS (Mauduit et al., 2012), and SSDF (Ashby et al., 2013). Similar to these surveys, SpIES observations were separated into

[2]http://www.darkenergysurvey.org/
[3]http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/warmimgcharacteristics/

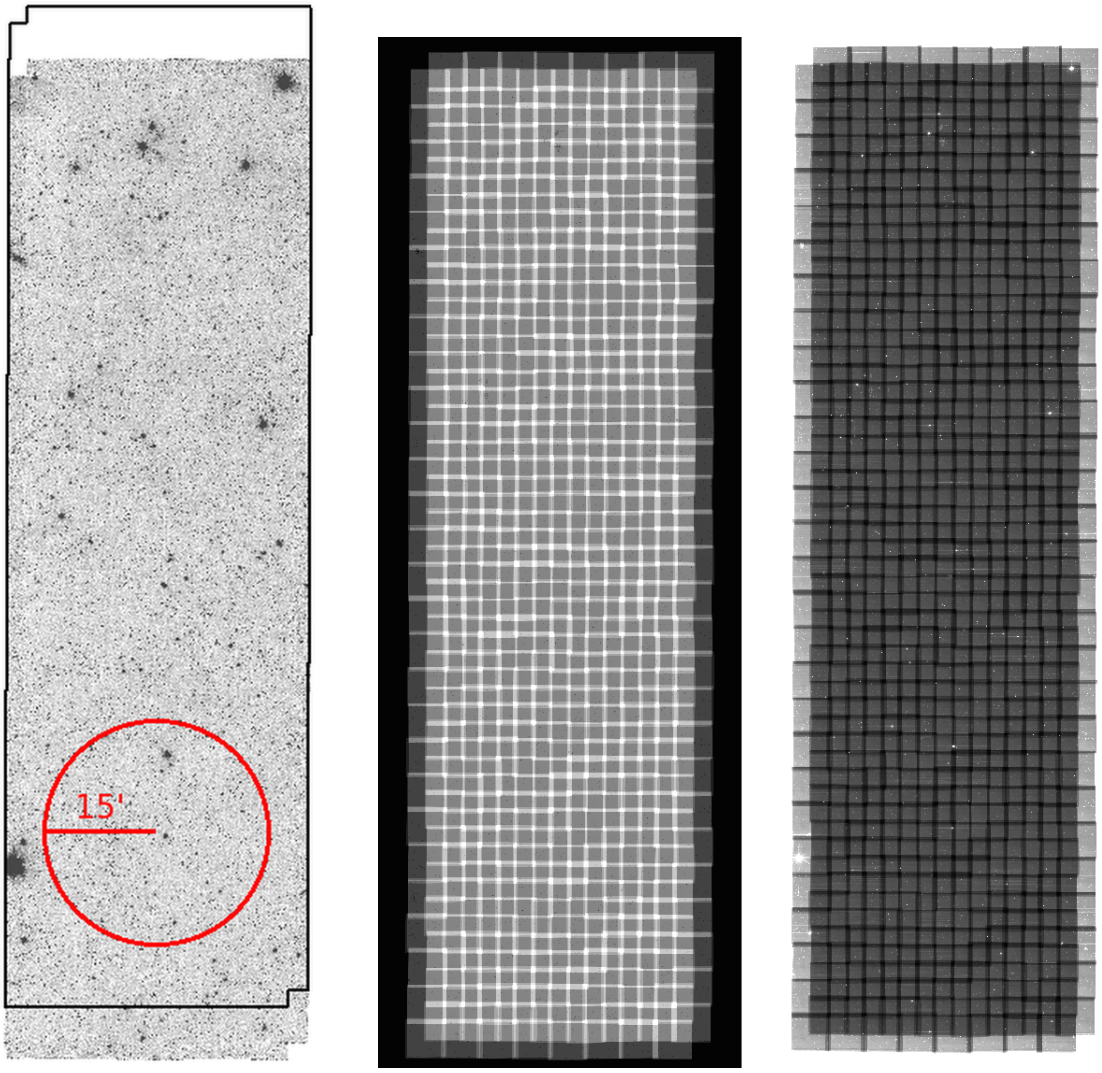

Figure 2.2 Left: One SpIES 3.6 μm, double-epoch, stacked AOR from which we extract sources. This is one of 77 stacked AORs (154 single epoch AORs divided by two epochs) that are strung together (see Figure 2.1) to cover the entire SpIES field. The red circular region illustrates the angular size of the Moon, and the black region shows the coverage of the same AOR at 4.5 μm. Center: An example of the coverage map of the AOR, showing where the individual pointings of IRAC overlap when they are combined to form the AOR. These maps are unique to each AOR and are used as weighted images during source extraction. Pixels with lighter colors have more coverages. The AOR footprint has been padded with a band corresponding to zero coverage. Right: The flux density uncertainty map of each AOR, where the values only take into account details in pipeline processing error propagation, not source extraction. In this map, darker colors correspond to lower uncertainties in flux density. The lower uncertainties align with the higher coverage values shown in the central panel.

Table 2.3. Astronomical Observation Request (AOR) Time Table

Operation	Time (s)
Exposure time at each pointing	30
×2 dithering	60
× ∼224 pointings	13440
+ Slew Time	∼2400
+ Settle Time	∼2400
+ Overhead(Slew and Download)	∼600
×2 epochs	∼37700
×77 AORs	$\sim 2.9\times10^6$
Total Observation Time	∼820hr

Note. — Approximate exposure time breakdown for SpIES for each detector (the larger AORs required more time than estimated). The two dithers and the two epochs combined with 30s exposures each lead to a total AOR exposure time of $2\times2\times30 = 120$s for both channels. SpIES spent ∼70% of the time in observation and ∼30% in motion to other fields.

individual Astronomical Observation Requests (AORs), which are self-contained exposure sequences executed independently of each other. AORs are comprised of sequential pointings of IRAC which are stacked to form a single image. AORs overlap slightly, to form the entire field (see the SpIES regions in Figure 2.1). Most of the SpIES AORs consist of a map of 8×28 IRAC FOVs, corresponding to a total area of ∼1.63 deg^2 per AOR (see Figure 2.2). There were, however, a few AORs which needed to be adjusted in width due to changes in position angle between AOR observations (observations separated by ∼6 months have a field rotation of ∼180°), to connect with their neighboring AORs and form a continuous strip. Four of our AORs were increased to 9×28 pointings, two were increased to 10×28 pointings, and one was decreased to 5×28 pointings. The size differences can be identified by an increase or decrease of the given AOR integration time in Appendix 2.8. In total, SpIES is comprised of 154 AORs observed over two epochs (77 AORs per epoch) which corresponds to ∼70,000 IRAC FOVs spanning the full survey area.

Each AOR was built by successively pointing and dithering IRAC until the 8×28 map was

complete, using a small-cycle dither pattern. This pattern offsets the observations by up to 11 pixels ($\sim$13″) to obtain overlapping coverage while eliminating some instrumental problems such as bad pixel detections and bright star saturation (Mauduit et al., 2012). Built into the cycle dither pattern is a sub-pixel dither pattern of half a pixel, which improves the 1.″2 per pixel sampling to 0.″6 per pixel after the images are stacked. This oversampling reduces effects that bad pixels and bright star saturation have on the image. This issue must be accounted for when calculating source flux error in Section 2.5.2.

Images are taken simultaneously at 3.6 μm and 4.5 μm with a $\sim$6.′7 offset between the two channels due to the physical placement of the arrays. This offset leads to a section around the perimeter where objects are detected in one band and not the other (as shown in Figure 2.2). The catalogs described in Section 2.5.3 indicate which objects lack a counterpart in the other band due to these regions without overlapping dual-band coverage. Additionally, the survey area changes slightly due to this offset. The quoted area of $\sim$115 deg^2 is the coverage where SpIES detects sources at either 3.6 μm or 4.5 μm. The coverage of each individual detector is $\sim$107 deg^2 where the coverage of the overlap of the two detectors (detections at both 3.6 μm and 4.5 μm) is $\sim$100 deg^2. This is important when computing number densities in Section 2.5.5.

Observations were performed over two distinct epochs separated by no less than five hours in time (see Appendix 2.8) and shifted by half a FOV in both right ascension and declination. Multiple epoch observations allow for detection of transient objects, and the spatial offset ensures that detected objects are observed on different regions of the array, allowing for more accurate photometry. In most cases, the second epoch of observation was taken directly after the first, where the observation time for the first epoch of a full AOR ($\sim$5 hours including slew and settle time) was sufficient to significantly separate the two epochs. For a typical asteroid, which moves at $\sim$25″ hr^{-1} (Ashby et al., 2009), a five-hour temporal separation leads to $\sim$2′ spatial separation, which is easily detected in separate epochs. The SpIES field is covered with at least four exposures at each pixel, providing both deep and reliable photometry across the large area of observation—with an exception around the perimeter where the second epoch has been shifted by half a FOV.

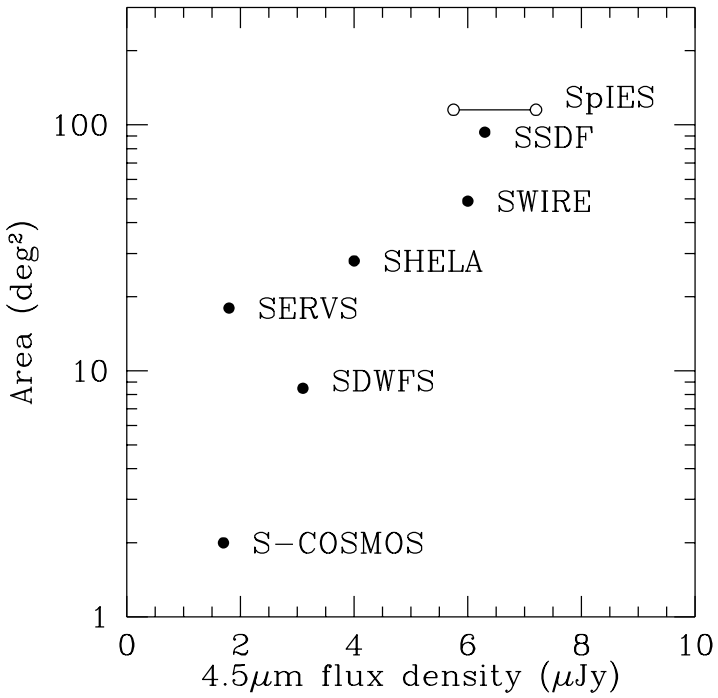

Figure 2.3 Comparison of the calculated 4.5 μm 5σ depth to area of the major *Spitzer* surveys. Depths are calculated using the *Spitzer* Sensitivity Performance Estimation Tool (SENS-PET) assuming a low background. At $\sim$115 deg^2 in area SpIES is the largest *Spitzer* survey and probes SWIRE depths (Lonsdale et al., 2003). Open circles show the measured depth (left; see Table 2.9) and calculated depth from SENS-PET with a medium background (right) for SpIES.

The SpIES AORs were constructed to maximize area while maintaining a depth comparable to that of SWIRE (Lonsdale et al., 2003). To achieve this goal, each AOR was observed for a total of 60 seconds, split evenly among the two dithered pointings of 30 seconds each. The limiting flux does not reach the IRAC confusion limit, and therefore confusion noise, which does not decrease as the square root of exposure time (Surace et al., 2005), is small (see Section 2.5.7 for more detail). The total observation time for the SpIES survey was $\sim$820 hours (Table 2.3) split among the 154 AORs. Figure 2.3 demonstrates that the SpIES survey is both the largest *Spitzer* survey to date and reaches approximately to SWIRE depths, fulfilling two of the projects primary goals.

2.4 Image Reprocessing

Observations from *Spitzer* are downlinked to the *Spitzer* Science Center (SSC) where the raw images are sent through the "Level 1" processing pipeline. This pipeline corrects for known instrumental signatures in the images (dark subtraction, ghosting, and flatfielding) and flags possible cosmic ray

hits. Additionally, the observed counts units (ADU) are converted into flux density units (MJy sr^{-1}), creating the Basic Calibrated Data (BCD) images (see Section 5 of the IRAC Handbook[1]). These BCD images are processed one 5.′2×5.′2 field at a time through a secondary pipeline to correct for other artifacts seen in IRAC images such as stray light (masking of scattered light from stars outside the array location) and column pulldown (a bright pixel causing a low background in the CCD array column; Figure 2.4). The resulting Corrected-BCD (cBCD) images (Section 6 of the IRAC Handbook) were used to create stacked AORs in SpIES (see Figure 2.2). A single cBCD image only covers one IRAC FOV; however, after accounting for the dithers and the two epochs, we have a total of four cBCD images which cover roughly the same region of the sky. The cBCD images are stacked to create the larger AOR mosaics using the SSC Mosaicing and Point-source Extraction (MOPEX[4]) software.

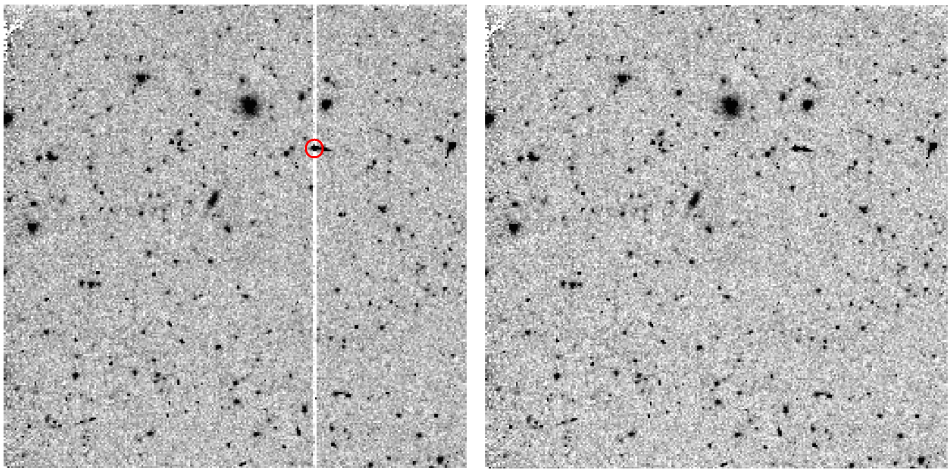

Figure 2.4 Left: Typical SpIES Level 1 BCD image from the SSC before corrections. The bright pixel (red circle) causes its whole column to drop to a low background value (causing the white line across the full array). Right: A cBCD image, which is the BCD image after it has been corrected for known signatures, such as the column pulldown in the left panel. The cBCD images are the size of an IRAC FOV (5.′2×5.′2) and are mosaicked together to form the larger AORs seen in Figure 2.2. Both images are centered at (α, δ)=(32.611, -0.887) degrees.

The MOPEX software was developed by the SSC specifically to process *Spitzer* BCD and cBCD images. This package contains several pipelines which can be used to process, stack, and extract sources from *Spitzer* images; however, we only relied on the mosaic pipeline to combine cBCD images onto a common frame. There are five stages of combination in the mosaic pipeline which transform

[4]http://irsa.ipac.caltech.edu/data/SPITZER/docs/dataanalysistools/tools/mopex/mopexusersguide/

Table 2.4. Parameter values for Mopex and SExtractor

Program	Parameter	Value
MOPEX	Fatal_Bitpattern	27392[a]
SExtractor	DETECT_THRESH	1.25
SExtractor	DETECT_MINAREA	4
SExtractor	DEBLEND_NTHRESH	64
SExtractor	DEBLEND_MINCONT	0.005
SExtractor	PHOT_APERTURES[b]	4.8, 6.4, 9.63, 13.6, 19.2, 40
SExtractor	PIXEL_SCALE	0.6
SExtractor	BACK_SIZE	64
SExtractor	BACK_FILTERSIZE	5
SExtractor	GAIN	4429.37, 3788.29[c]
SExtractor	WEIGHT_TYPE	MAP_WEIGHT
SExtractor	WEIGHT_IMAGE	mosaic_cov.fits
SExtractor	WEIGHT_GAIN	Y
SExtractor	FILTER	Y
SExtractor	FILTER_NAME	default.conv

Note. — Parameters that were changed from the default MOPEX or SExtractor configuration files. These parameters were used in the stacking and source extraction of the SpIES images.
[a]DCE_Status_Mask_Fatal_BitPattern with bits 8,9,11,13,14 are turned on.
[b] The diameter of the aperture in pixels.
[c]Gain values for the 3.6 μm, 4.5 μm detector. See Section 2.5.2 for more details

a list of cBCD images to a full mosaic. First, an interpolation technique is run on the input images, determining the location of each pixel and forming a fiducial frame for the output image. Next, an outlier rejection script is run which flags or masks bad pixels from the final image. These flags are applied to the fiducial frame with a re-interpolation technique. Co-addition of pixel values is performed on tiles of pixels that make up the full image using a method defined by the user (for SpIES, pixels were co-added using a straight average). Finally, a script combines the tiles from the co-addition stage together to form a single image. Along with a combined image, MOPEX provides an option to output other datasets such as a coverage map and uncertainty map similar to those shown in Figure 2.2. The SSC also provides these images as "Level 2" post-BCD (pBCD) images which have been processed by MOPEX and thus can be used for source extraction and photometry; however, they are only single epoch images, thus do not achieve the full depth of our survey.

To achieve our full depth, we created images by submitting the cBCD images of the two overlapping epochs as well as their corresponding bit mask (bimsk) images and the uncertainty (cbunc) into MOPEX. The pipeline was run using the default parameters with the exception of the DCE_Status_Mask_Fatal_BitPattern (see Table 2.4) which tells MOPEX which pixels to mask in the final mosaic based on the bit value of those pixels in the input bit mask. For example, the 3.6 μm 'warm' IRAC images suffer from latent images[5] (typically after exposure to bright stars) which remain at the same pixel location on the detector for the next set of observations (see Figure 2.5). If left unchecked, these objects appear in a different sky location in the final image, and will be detected as individual sources. To prevent contamination in the final AOR, the SSC pipeline locates latent objects in each BCD, and flags the corresponding pixels in the bit mask[6] for that BCD. We then set the DCE_Status_Mask_Fatal_BitPattern (which reads the bit masks) to mask any objects that have that particular flag set in the final combined image (see Figure 2.6). Since latent objects do not appear in our final stacked images they are not present in our final catalogs.

The SSC-produced BCD, cBCD, and pBCD images, as well as all ancillary data images (uncer-

[5]http://irsa.ipac.caltech.edu/data/SPITZER/
docs/irac/iracinstrumenthandbook/63/
[6]http://irsa.ipac.caltech.edu/data/SPITZER/
docs/irac/iracinstrumenthandbook/44/#_Toc410728355

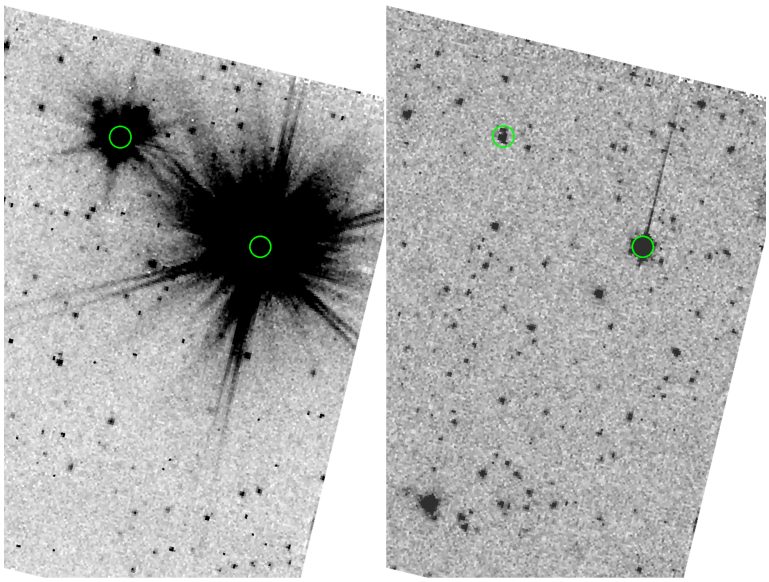

Figure 2.5 Shown on the left is an example of two bright stars in a $\sim 3' \times 4\farcm5$ cutout of a 3.6 μm cBCD (centered at (α, δ)=(34.464, -0.169) degrees). The image in the right panel is the next observation (centered at (α, δ)=(34.482, -0.247) degrees) showing the latent images from the bright stars in the previous observation (left panel). The green circles highlight the pixel location of the latent objects in IRAC from subsequent observations at different sky locations.

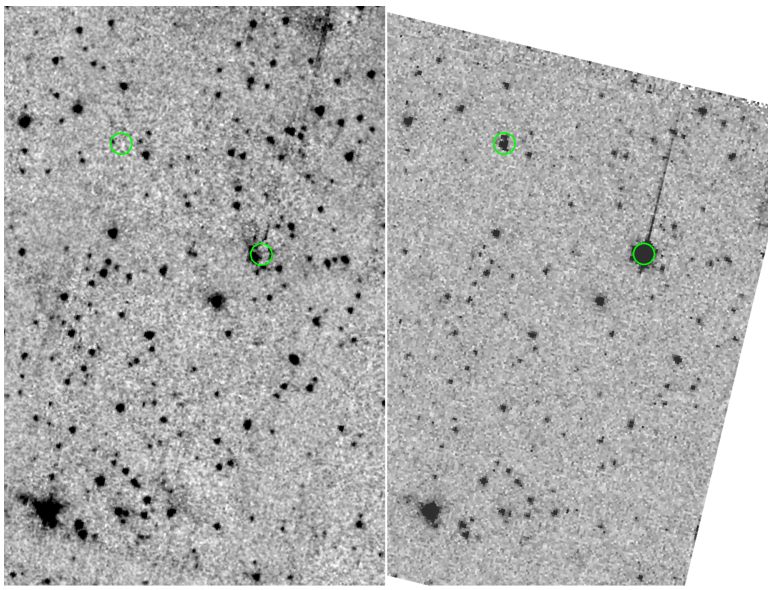

Figure 2.6 Here, the left panel shows a portion of the final stacked AOR image after sky matching to the right panel in Figure 2.5 (also the right panel of this figure) with the latent object locations outlined in green. The latent objects in the cBCD (right panel) are masked in the final stacked image (left panel) because the latent image bits were turned off in the MOPEX processing pipeline (see Table 2.4), therefore, they do not appear in the final catalogs.

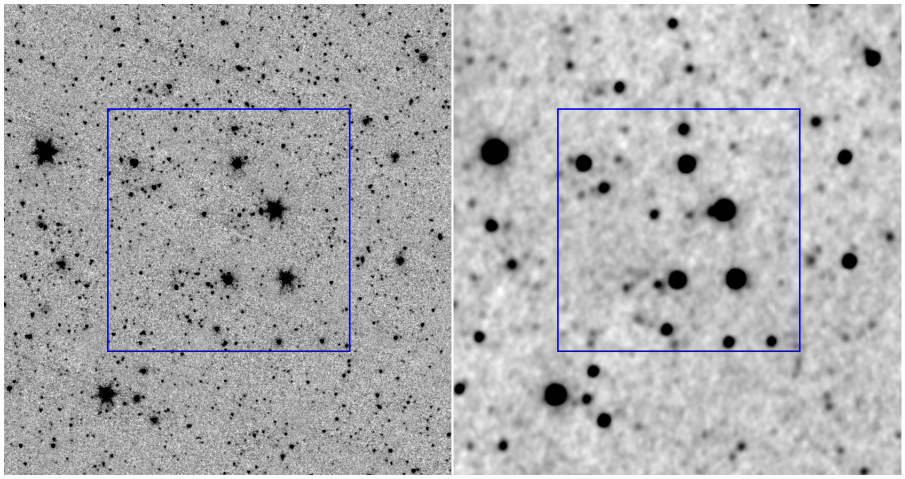

Figure 2.7 Comparison of a ∼100 arcmin^2 box of a SpIES 4.5 μm image and a 4.6 μm image which cover approximately the same central wavelength. 'Warm' IRAC 4.5 μm has a PSF of 2.″02 compared to 6.″4 for *WISE* 4.6 μm, allowing SpIES to resolve objects that are blended in *WISE*. Additionally, the superior depth of SpIES (AB magnitude of ∼22 in [4.5] compared to ∼18.8 in W2) yields more sources above the background (∼1400 in the dual-band catalog) in the field shown compared to *WISE* (∼350 in AllWISE). The blue boxes represent a single FOV of IRAC (5.′2 × 5.′2).

tainty maps, coverage maps, etc.), are publicly available on the *Spitzer* Heritage Archive[7] (SHA) website. The images created by the SpIES team are publicly available (see Appendix 2.7). There are a total of 231 images created by the SpIES team consisting of 154 individual epoch AOR mosaics and 77 combined epoch mosaics (stacking the two overlapping individual epoch images). Source extraction and photometry were performed on each of these 231 images. The final catalogs were constructed by running our source extraction techniques on the 77 combined epoch AORs to take advantage of the full depth of SpIES. To illustrate the depth of SpIES, Figure 2.7 compares a region from a full-depth 4.5 micron AOR and the same region from *WISE* 4.6 micron ($W2$).

2.5 Catalog Production

2.5.1 Source Extraction

The SpIES catalogs were constructed by running Source Extractor (SExtractor; Bertin & Arnouts 1996) on each combined-epoch AOR mosaic, creating 77 AOR source catalogs for the 3.6 μm detections and 77 for the 4.5 μm detections. SExtractor uses a six-step source extraction routine which

[7]http://sha.ipac.caltech.edu/applications/Spitzer/SHA/

efficiently generates catalogs from large images. First, a robust 3σ clipped background estimation is performed on the entire image, which has been inspected through an output background map. This step is followed by a thresholding algorithm which extracts objects at a certain, user-specified standard deviation above the background. SExtractor then runs a deblending routine to separate potentially blended sources, filters the image using an input filtering routine, and performs photometry on detected sources within user specified apertures. Finally, SExtractor attempts to classify objects as point-like (stars) or extended (galaxies) based on the input pixel scale and stellar FWHM of the survey.

Each step is controlled through an input configuration file and an output parameter file. There are a variety of parameters that can be changed in the configuration file, some of which can significantly change the source extraction results. The final configuration file was a mix of parameters extensively tested on the SpIES images and parameters adopted from previous programs such as the SERVS (Mauduit et al., 2012) and SWIRE (Lonsdale et al., 2003) surveys. Table 2.4 lists the configuration parameters used in our processing.

Previous *Spitzer* surveys also used the coverage map created in MOPEX as a weighted image during source extraction. These images hold information about the number of times a particular pixel in the AOR was observed, which is related to the effective exposure time at each pixel. Since the signal-to-noise ratio of an object increases with the square root of exposure time in these data, the coverage maps assign pixels with more coverages (i.e., longer exposures) a higher weight. Following this convention, the coverage maps were input as weight maps, converted into a variance map by SExtractor through the inverse relationship between weight and variance, and scaled to an absolute variance map created internally by SExtactor. This processing is also controlled through the input configuration file during source extraction.

SExtractor can be run in either single-detection mode, which performs source detection, aperture definition, and photometry on the same image, or dual-detection mode, which finds sources and defines apertures on a first input image (for example, a $3.6\,\mu$m AOR) and performs photometry on a second input image (the same AOR observed using the $4.5\,\mu$m detector). All of the SpIES AOR

Table 2.5. Aperture correction for SpIES

Band	1.″4	1.″9	2.″9	4.″1	5.″8
3.6 μm	0.584	0.732	0.864	0.911	0.950
4.5 μm	0.570	0.713	0.865	0.906	0.946

Note. — Measured aperture corrections for SpIES objects with good flags matched to the 2MASS point source catalog. These corrections are nearly identical to those used in SERVS (Mauduit et al., 2012) for identical aperture radii.

mosaics were run in single-detection mode, creating 77 double-epoch catalogs for each channel. Full-area, single-channel catalogs were made by concatenating the 77 individual AOR catalogs using the Starlink Tables Infrastructure Library Tool Set (STILTS)[8]. These single-channel catalogs are designed to contain a single row for each object in the SpIES survey, so when two objects match within 1″ between two AORs (which is possible since the AORs overlap) we report the average position, the weighted average of the flux density values (using the errors as weights), and the errors added in quadrature in a single row in the catalog (the overlapping regions between AORs account for ∼10% of the total survey area). Though we report objects that are detected 5σ above the calculated background, many objects have a signal-to-noise (S/N) less than 5 due to Poisson noise.

Photometry on SpIES sources was performed in six circular apertures of radii 1.″4, 1.″9, 2.″9, 4.″1, 5.″8, and 12″, reported as diameter in pixels in the SExtractor configuration file in Table 2.4. The first five apertures (which are the same size as the SERVS apertures) contain only a fraction of the light from each source, while the sixth contains "all" the light from the source (see Section 4.11 of the IRAC Handbook[1]). The aperture correction factors in Table 2.5 are measured for the SpIES survey for objects with good flags (discussed in more detail in Section 2.5.3) matched to the 2MASS Point Source Catalog (PSC) to ensure that measurements were performed on point sources only. We then took the ratio of the light in the smaller apertures to the light in the largest aperture,

[8]http://www.star.bris.ac.uk/∼mbt/stilts/

made a histogram of the resulting factors for each aperture, and fit a Gaussian to that histogram to measure the peak and spread of the distribution. The location of the peak of the Gaussian was used as the correction factor. The corrections measured for SpIES differ by less than 1% of those used in SERVS (Mauduit et al., 2012) for the exact same aperture radii. Aperture corrections are useful for finding faint objects with a radius much less than the large $12''$ radius aperture, because in these cases the background noise in the aperture would dominate the object. We primarily use the $1\farcs9$ radius aperture for analysis in the following sections as it corresponds to a $\sim$70% curve of growth correction (the curve showing how the flux density ratio changes with aperture size) in both channels.

After objects are extracted from the images, the surface brightness values are converted from the *Spitzer* image unit of $\mathrm{MJy\,sr^{-1}}$ to flux densities (μJy) per pixel using the following conversion:

$$\frac{\mathrm{MJy}}{\mathrm{sr}}\left(10^{12}\,\frac{\mu\mathrm{Jy}}{\mathrm{MJy}}\right)\left(\frac{\pi\mathrm{rad}}{180^{\circ}}\right)^{2}\left(\frac{1^{\circ}}{3600''}\right)^{2}\left(\frac{0\farcs6}{\mathrm{pixel}}\right)^{2}$$

such that,

$$1\mathrm{MJy\,steradian^{-1}} = 8.46\mu\mathrm{Jy\,pixel^{-2}} \tag{2.1}$$

where we multiply by the SpIES pixel size of $0\farcs6$, which is half of the IRAC pixel size due to the image dithering.

This correction factor in Equation 2.1 was applied to each pixel in the image which, when summed in an aperture, yields the total flux density of the source. This value was divided by the appropriate aperture correction from Table 2.5 to produce the final flux density value for the objects in the catalogs.

2.5.2 Photometric Errors

Photometric errors were computed using SExtractor and are reported in the catalog (see Table 2.4). According to Section 10.4 of the SExtractor manual, the 1σ photometric errors are computed via

$$\sigma_{\mathrm{source}} = \sqrt{A\sigma_{rms}^{2} + \frac{F}{g}}, \tag{2.2}$$

where A is the measurement area in pixels, σ_{rms} is the background root-mean-square (rms) value of each pixel, F is the background-subtracted source count value in the measurement aperture, and g is the detector gain. This expression is simply the rms background added in quadrature with the Poisson noise. SExtractor assumes that the signal in the input images is in units of counts, typically a Digital Number (DN) which is the number of photons counted scaled by the detector gain value. *Spitzer* images, however, are converted to physical units during "Level 1" processing. Many previous surveys which have used SExtractor to compute photometric errors exclude the Poisson noise and only report the rms background error, which is also the SExtractor default if no gain is supplied. For bright objects, Poisson noise dominates, and thus using the background error alone dramatically underestimates the true error in the reported flux density. Here we compute and report the full photometric errors from SExtractor for the SpIES survey, correcting for the *Spitzer* image flux units such that both background and Poisson noise are included in the error estimate. Indeed the majority of the sources in our "5σ catalog" will have true soure S/N < 5 (and more typically $\sim$2-3).

To properly incorporate *Spitzer* data into Equation 2.2, we first examine its fundamental components: the noise due to the background and Poisson counting noise. In order to compute the background noise, SExtractor first creates a background map and a background rms map. The background rms map is constructed by calculating the squared rms deviation of each pixel in the background map from the local mean background (whose size is defined by the BACK_SIZE parameter in Table 2.4). The background noise is simply the sum of the background rms pixels inside a given aperture (where $A\sigma^2_{rms}$ in Equation 2.2 is synonymous with the sum over the background rms).

Poisson noise is the discrete counting error which occurs when performing photometry on a source. SExtractor performs photometry on an object inside of an aperture by counting the total pixel value and subtracting the background as follows:

$$F = C - B \tag{2.3}$$

where F is the background-corrected count value of an object, B is the sum of the local background

value in the aperture, and C is the total number of counts in an aperture. Assuming the pixel values in the measurement aperture are uncorrelated (which presents a separate problem that is discussed later in this section), then the error in F can be calculated using the propagation of error equation:

$$\sigma_F^2 = \left(\frac{\delta F}{\delta C}\right)^2 \sigma_C^2 + \left(\frac{\delta F}{\delta B}\right)^2 \sigma_B^2 \tag{2.4}$$

where σ_C and σ_B are the Poisson errors of the total number of counts and background respectively. Taking the derivatives of Equation 2.3 and inserting them into Equation 2.4, we obtain:

$$\sigma_F^2 = \sigma_C^2 + \sigma_B^2. \tag{2.5}$$

The number of electrons measured, the number of counts reported, and the gain are related by:

$$\#e^- = g \times F \tag{2.6}$$

which has an uncertainty,

$$\sigma_{\#e^-}^2 = g^2 \times \sigma_F^2. \tag{2.7}$$

Poisson statistics dictate that the variance of a discrete value (in this case electron number, $\sigma_{\#e^-}^2$) is equal to that value (the number of electrons counted). We therefore relate the number of electrons to the digital count in Equation 2.6 and obtain that the Poisson error for a digital count is:

$$\sigma_F^2 = \frac{\#e^-}{g^2} = \frac{g \times F}{g^2} = \frac{F}{g}. \tag{2.8}$$

This Poisson error (which must have the digital count unit) is the second term in Equation 2.2, and is added in quadrature with the rms background error to generate the total source error found in Equation 2.2.

Spitzer images and SExtractor use two different definitions of the gain. SExtractor is programmed to interpret this parameter as purely the detector gain (which has units of electrons per

digital count) whereas *Spitzer* images have a definition of gain that includes the conversion factor between counts units and physical units. Even though SExtractor expects an image in counts units, we can input *Spitzer* images by incorporating this conversion factor in the gain parameter according to the equation:

$$G = \frac{N \times g \times T}{K} \tag{2.9}$$

where N is average number of coverages estimated from each AOR coverage map, g is the detector gain of 3.7 $e^{-}(DN)^{-1}$ for the 3.6 μm detector and 3.71 $e^{-}(DN)^{-1}$ for the 4.5 μm detector, T is exposure time for one coverage, and K is the conversion factor from digital to physical units found in either the cBCD header or the Warm IRAC Characteristics webpage[9]. For the SpIES images, we calculated the weighted gain, G, to be 4429.37 $e^{-}(\mathrm{MJy\,sr^{-1}})^{-1}$ at 3.6 μm and 3788.29 $e^{-}(\mathrm{MJy\,sr^{-1}})^{-1}$ at 4.5 μm; these values were used in the SExtractor configuration file for source extraction and error estimation. In short, replacing the detector gain, g, with the weighted gain, G, in Equation 2.2 allows a proper determination of both the background and Poisson noise when applying SExtractor to images that have been converted to physical units.

After the gain parameter is replaced, applying simple unit analysis to Equation 2.2 shows that the errors have the same unit as the input image (in this case MJy sr^{-1}). We therefore need to convert the errors from image units of MJy sr^{-1} to μJy using Equation 2.1 in the same way as we did for the flux density values. The error analysis was also done inside apertures of varying radii and therefore also must be aperture corrected by dividing by the values in Table 2.5.

Finally, Equation 2.2 is based on the assumption that the pixels in the images are uncorrelated, which simplifies the SExtractor error calculation. In reality, the SpIES images will have cross correlation terms due to processes such as dithering, reprojection, and stacking, which correlate the count value in overlapping pixels. Since SExtractor does not take correlated noise into account, we corrected the values by multiplying the errors by a factor of two (the ratio of the pre-processed image pixel scale of $1\farcs2$ to the post-processed pixel scale of $0\farcs6$), which accounts for the pixels being sampled twice due to the two dithers in the survey. Although the errors are slightly adjusted to

[9]http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/warmimgcharacteristics/

Table 2.6. SpIES catalog columns

Column Name	Description
RA_ch1	J2000 RA position at 3.6 μm
DEC_ch1	J2000 DEC position at 3.6 μm
FLUX_APER_1_ch1	3.6 μm flux density, 1.″44 radius
FLUX_APER_2_ch1	3.6 μm flux density, 1.″92 radius
FLUX_APER_3_ch1	3.6 μm flux density, 2.″89 radius
FLUX_APER_4_ch1	3.6 μm flux density, 4.″08 radius
FLUX_APER_5_ch1	3.6 μm flux density, 5.″76 radius
FLUX_APER_6_ch1	3.6 μm flux density, 12″ radius
FLUXERR_APER_1_ch1	3.6 μm flux density error, 1.″44 radius
FLUXERR_APER_2_ch1	3.6 μm flux density error, 1.″92 radius
FLUXERR_APER_3_ch1	3.6 μm flux density error, 2.″89 radius
FLUXERR_APER_4_ch1	3.6 μm flux density error, 4.″08 radius
FLUXERR_APER_5_ch1	3.6 μm flux density error, 5.″76 radius
FLUXERR_APER_6_ch1	3.6 μm flux density error, 12″ radius
FLUX_AUTO_ch1	Total 3.6 μm flux density
FLUXERR_AUTO_ch1	Total 3.6 μm flux density error
FLAGS_ch1	3.6 μm SExtractor Flags
CLASS_STAR_ch1	3.6 μm morphology classification
FLAG_2MASS_ch1	3.6 μm object near a bright star
COV_ch1	Number of cBCD coverages
HIGH_REL_ch1	Most reliable objects with good flags

account for oversampling, they should still be considered as lower limits on the true error in each aperture since there are other contributions to the correlated noise in each pixel for which we do not correct (i.e., noise pixels). These photometric error estimates will be used in Section 2.5.6 as one of the ways we measure the depth of the survey.

Table 2.6 (cont'd)

Column Name	Description
RA_ch2	J2000 RA position at 4.5 μm
DEC_ch2	J2000 DEC position at 4.5 μm
FLUX_APER_1_ch2	4.5 μm flux density, 1.″44 radius
FLUX_APER_2_ch2	4.5 μm flux density, 1.″92 radius
FLUX_APER_3_ch2	4.5 μm flux density, 2.″89 radius
FLUX_APER_4_ch2	4.5 μm flux density, 4.″08 radius
FLUX_APER_5_ch2	4.5 μm flux density, 5.″76 radius
FLUX_APER_6_ch2	4.5 μm flux density, 12″ radius
FLUXERR_APER_1_ch2	4.5 μm flux density error, 1.″44 radius
FLUXERR_APER_2_ch2	4.5 μm flux density error, 1.″92 radius
FLUXERR_APER_3_ch2	4.5 μm flux density error, 2.″89 radius
FLUXERR_APER_4_ch2	4.5 μm flux density error, 4.″08 radius
FLUXERR_APER_5_ch2	4.5 μm flux density error, 5.″76 radius
FLUXERR_APER_6_ch2	4.5 μm flux density error, 12″ radius
FLUX_AUTO_ch2	Total 4.5 μm flux density
FLUXERR_AUTO_ch2	Total 4.5 μm flux density error
FLAGS_ch2	4.5 μm SExtractor Flags
CLASS_STAR_ch2	4.5 μm morphology classification
FLAG_2MASS_ch2	4.5 μm object near a bright star
COV_ch2	Number of cBCD coverages at 3.6 μm
HIGH_REL_ch2	Most reliable objects with good flags

Note. — Column descriptions for the three SpIES catalogs. The 3.6 μm-only and 4.5 μm-only catalogs are built in exactly the same manner without the columns from the other channel. All flux density and flux density error columns in this catalog have been converted from MJy sr^{-1} to μJy pixel^{-1} using Equation 2.1, and the first five apertures in each channel have been aperture corrected using the values in Table 2.5.

2.5.3 SpIES Source Catalogs

Using the parameters in Table 2.4 and employing the techniques discussed in previous sections, we generated the SpIES 5σ detection catalogs. Here 5σ refers not to objects with a ratio of flux density to flux density error of greater than five, but rather to objects whose flux density is greater than five times the background. This limit is found by taking the product of the DETECT_MINAREA (minimum number of adjacent pixels to make a source) and DETECT_THRESH (number of standard deviations above the background per pixel) parameters (see Table 2.4 for reference). In fact, the majority of these objects have a S/N of $\sim$2-3, due in large part to the addition of the Poisson noise

as shown in Section 2.5.2.

With this release, we provide three separate detection catalogs: a 3.6 μm-only detection catalog which contains ∼6.1 million objects that are only detected at 3.6 μm, a 4.5 μm-only detection catalog containing ∼6.6 million objects only detected at 4.5 μm, and a dual-detection catalog containing ∼5.4 million sources, comprised of the sources detected at the same positions in both bands. These catalogs were constructed by extracting sources from the 3.6 μm and 4.5 μm AORs separately to generate full object catalogs for each channel. We then matched these two single-band catalogs using a matching radius of 1.″3 (as determined by the Rayleigh criterion), which maximized the number of true matches and minimized the false detections (∼6.5% for the high reliability objects described below) between the two channels to create our combined dual-band catalog. The objects that did not match remained in the single band catalogs. Due to the offset between the detectors in IRAC, there were ∼600,000 objects in 3.6 μm without coverage in 4.5 μm and ∼600,000 objects in 4.5 μm without coverage in 3.6 μm. These objects, however, are retained in their respective single band catalogs. As the majority of the objects in the single-band catalogs have S/N∼2-3, it is perhaps not surprising that they are detected in only one band. However, included among these will be transient objects and mid-infrared/optical dropouts, which are clearly of interest, in addition to spurious sources, which are not. Thus, we recommend using the high reliability flags for the most reliable objects in each catalog (described below).

These catalogs were constructed from the combined epoch AORs, and thus reach the full depth achievable by the SpIES survey. As also noted in the previous section, each row in the catalogs contains a unique source. The columns hold information about the astrometric and photometric values for each source, the flags that were generated during source extraction, and several binary columns which have various meanings (see Table 2.6). The three catalogs are structured in exactly the same way, the only difference being whether or not the object in the catalog is matched between the two channels. A user desiring *all* the 3.6 μm detections can concatenate the 3.6 μm-only and the dual-band catalogs without any changes to the files.

Each row in the catalog contains information about a unique source at a particular J2000 RA

Table 2.7. Sextractor flags

Bit Value	Description
1	The object has neighbors, that significantly bias the photometry, or bad pixels.
2	The object was originally blended.
4	At least one pixel is (nearly) saturated.
8	The object is truncated (close to image boundary).
16	Aperture data are incomplete or corrupted.
32	Isophotal data are incomplete or corrupted.
64	A memory overflow occurred during deblending.
128	A memory overflow occurred during extraction.

Note. — All of the extraction flags from SExtractor. The first five flags are the most common for SpIES as these pertain to issues in source extraction. The last three do not appear in the SpIES data since there are no isophotal aperture measurements and a sufficient amount of memory was allocated for extraction.

and DEC position, which was determined by SExtractor, as reported in the first two columns (both channel positions are reported for matched objects). These positions have been corrected for a slight offset when compared to SDSS point sources (see Section 2.5.4 for more details). The subsequent twelve columns report the flux density values from the six different measurement apertures used in source extraction along with their respective errors. Aperture-corrected flux density values are reported in these columns (except for aperture 6 which is not corrected) and surface brightness units (MJy sr^{-1}) are converted to flux densities (μJy) using Equation 2.1. Additionally, the errors have been adjusted in the manner described in the previous section. The next two columns (FLUX_AUTO and FLUXERR_AUTO) report the flux density and flux density error in apertures whose size and shape are determined by SExtractor to contain the total flux density from a source. These last two values have been converted to flux densities using Equation 2.1; however, they are not aperture corrected.

The extraction flags are reported in the next column as a 2-dimensional array (see Table 2.7 for more information). Since source extraction was performed on an individual AOR basis, the sources

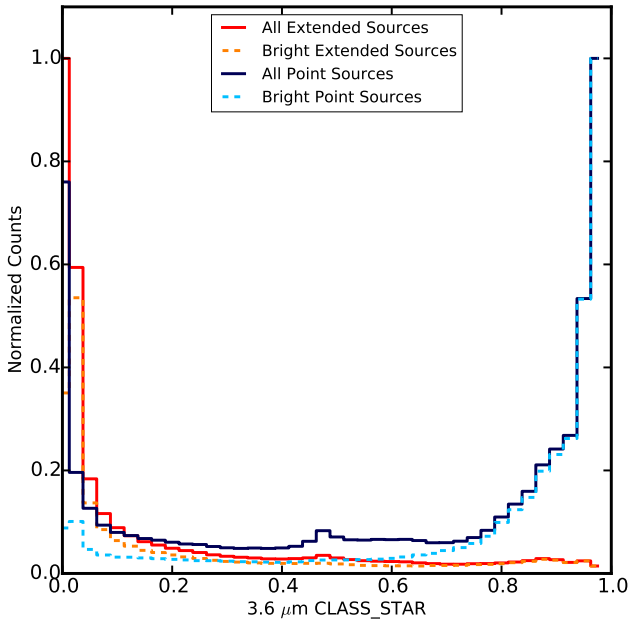

Figure 2.8 Comparisons of the CLASS_STAR parameter at 3.6 μm for objects matched to SDSS sources. We show the distribution for all optically extended sources (red) and all optical point sources (dark blue). Optically extended sources peak at CLASS_STAR$\sim$0, while optical point sources peak at $\sim$1; however there is a small peak at 0.5 implying that SExtractor could not differentiate between point or extended. For bright objects ([3.6] $\leq$ 20.5), however, the extended (orange dashed) and point (light blue dashed) sources still peak at 0 and 1, respectively, but there are far fewer confused classifications. A similar trend occurs for the objects detected at 4.5 μm.

on the edges of AORs have the potential to be detected twice, due to the overlap between AORs, and thus both flags were retained (however there is only one row entry in the catalog for overlapping objects). Sources that do not overlap have a flag value in the first array element and were given a value of -999.0 in the second element in this column to make it clear that this source was detected in only one AOR.

The SExtractor stellar class is reported in the CLASS_STAR column which is a probability that ranges from 0 to 1 and indicates whether an object is resolved (values closer to 1) or extended (values near 0). If the object was detected twice due to the overlap of the AORs, the average value is given in the catalog. We find this measurement to be most reliable for objects with magnitudes brighter than 20.5 ($\sim$1.7 million at 3.6 μm and $\sim$1.5 million at 4.5 μm in the dual-band catalog), with $\sim$40% classified as resolved (CLASS_STAR $\geq$ 0.5) and $\sim$60% as extended (CLASS_STAR $\leq$ 0.5) in both

bands (see Figure 2.8).

Following the SExtractor output columns are a series of flags created after source extraction. The FLAG_2MASS column indicates whether a source is detected within a particular radius (defined by Table 2.8) around a bright star in the 2MASS point source catalog (PSC). Inside this radius there is an excess of artificial sources due to artifacts from the bright star (e.g., diffraction spikes). Flags are assigned to objects near 2MASS stars with K_s-magnitude brighter than 12 (Vega magnitude), where the radii range from $40''$ at the faint end to $180''$ at the bright end. For comparison, the radii used for the SWIRE survey range from $10''$ at the faint end to $120''$ at the bright end using similar (but not the same) K_s-magnitude cuts (see Surace et al. 2005).

The SpIES bright-star flagging radii were empirically determined by cutting the 2MASS PSC into a series of K_s-band magnitude ranges and matching their positions to all SpIES objects within $300''$. We then overlay the positions of all of the stars in a K_s-magnitude bin along with their SpIES matches onto a common coordinate frame and determine the radius which encapsulates the over-dense region around the star. Figure 2.9 shows the result of stacking $6 \leq K_s \leq 7$ Vega magnitude stars and their matches on a coordinate frame. The radial profile plot is presented in Figure 2.10 which clearly shows an excess of detections near bright stars. Objects that fall within the radii in Table 2.8 are given a value of 1 in the catalog to indicate that the source is potentially spurious, and the central star itself is given a value of 2. Using the radii in Table 2.8, we compute the area lost when rejecting such sources is $\sim$5 deg^2 for both bands (which is $\sim$5% of the dual-band catalog area).

We report the number of cBCD coverages (from the coverage maps shown in Figure 2.2) at the centroid position of each source in the COV column. Since AORs overlap, we give an array of two values where, if the object does not overlap, we report -999.0 in the second element (similar to the extraction flags). For the most reliable detection, we recommend using objects which have greater than two coverages in either entry of the reported array.

Finally, we have created a high reliability column which we recommend for users whose science requires that the objects be robust sources and/or have robust photometry. There are three values

Table 2.8. Bright star flagging radius

2MASS (K_s-Magnitude)	Radius ($''$)
≥ 12	0
$12 - 10$	40
$10 - 9.0$	60
$9.0 - 8.0$	90
$8.0 - 7.0$	120
$7.0 - 6.0$	150
≤ 6.0	180

Note. — Objects that fall within the radii given are flagged as bright star contaminants. These values are empirically determined by making K_s-magnitude cuts on 2MASS stars and studying figures like Figure 2.9 and Figure 2.10. The K_s-magnitudes are computed in Vega magnitudes.

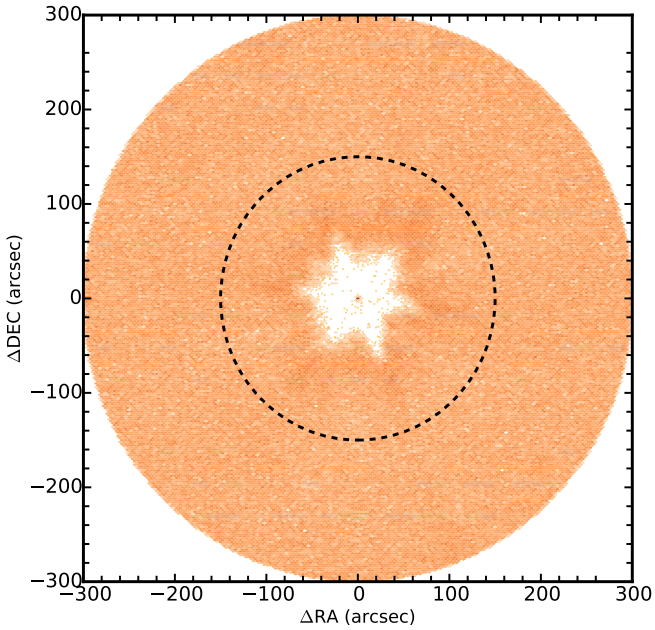

Figure 2.9 The 335 stacked $6 \leq K_s$-magnitude≤ 7 stars matched to SpIES within $300''$. The black dashed circle shows the radius out to which we flag objects as potentially contaminated.

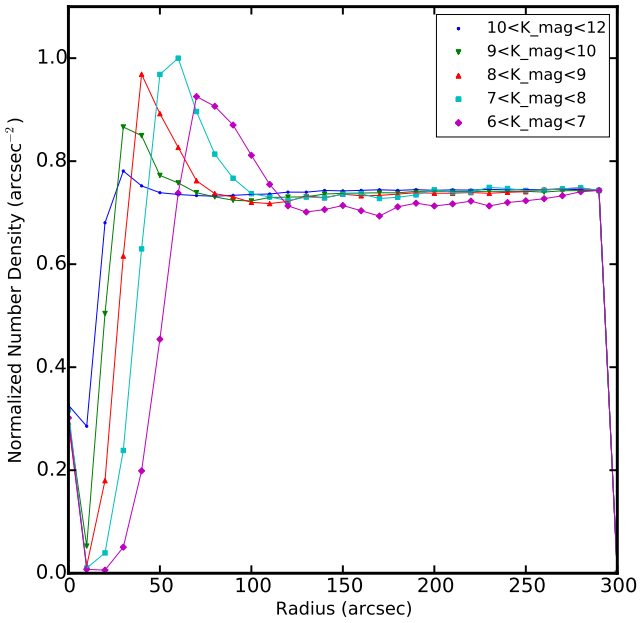

Figure 2.10 Radial profiles of the number density of objects within 300″ of 2MASS stars in magnitude ranges given in Table 2.8, showing how the number density of detected objects around bright stars changes as a function of distance from the center of the star. The peak in these curves is the over-dense region where there are spurious detections due to artifacts such as diffraction spikes. We cut at the radius where the curves approach a constant value of number density for each magnitude.

in this column indicating whether a source is a real object (flagged with a value of 1 or 2), has good photometry (flagged with a value of 2), or does not satisfy the following good flag conditions (flagged with a 0). To be regarded as a real source, the SExtractor flags must be less than or equal to 4, the objects must have flag 0 or 2 in the FLAG_2MASS column, and there has to be greater than or equal to two coverages for each source. For an object to have good photometry, we further require that the SExtractor flags be less than or equal to 0 or equal to 2 (i.e., −999.0, 0, and 2), FLAG_2MASS must be 0, and it must satisfy the same coverage conditions as before. These flags cause holes in the coverage across the survey, thus changing the total coverage area. In total, SpIES has ∼115 deg^2 of coverage in both wavelengths, of which, each band covers ∼107 deg^2 (since there is an offset in the arrays discussed in Section 2.3) and there is ∼100 deg^2 of dual-band coverage. For HIGH_REL>0, the areas are ∼106 deg^2, ∼101 deg^2, and ∼94 deg^2, while for HIGH_REL=2, the areas drop to 105 deg^2, ∼100 deg^2, and ∼89 deg^2. While our catalog only includes sources more than 5σ above the background, full error analysis means that individual objects can have S/N (as computed by FLUX/FLUXERR) less than 5. Some users may want to apply a cut on S/N in addition to using the HIGH_REL flag. For a cut at S/N>3 and HIGH_REL>0, we retain ∼1.4, ∼3.9, and ∼1.4 million objects in the 3.6 μm-only, dual-band, and 4.5μm-only, respectively.

2.5.4 Astrometric Reliability

The astrometric reliability of SpIES was tested by comparing the centroid positions of point sources in SDSS with matched objects in the SpIES dual-band catalog (within 2″). We found the difference in position for objects which have good flags in SDSS (BITMASK=0 and PHOTOMETRIC=1), are bright in the r-band ($r \leq 21$), and have good flags in SpIES (HIGH_REL=2). Fitting a Gaussian to the histograms in Figure 2.11, we find that the mean difference in RA is $-0\farcs112 \pm 0\farcs0008$ and in DEC is $0\farcs0372 \pm 0\farcs0006$ for these objects. These values were then used to correct the astrometry in all three SpIES catalogs. We also matched the SpIES data with the 2MASS PSC and found that the mean astrometric offsets (ΔRA= $-0\farcs086 \pm 0\farcs0006$ and ΔDEC= $0\farcs011 \pm 0\farcs0005$) are slightly smaller than the calculations from SDSS, however confirm the direction of the SpIES positional shifts.

To see if the astrometric offset changes with brightness, we performed the same measurement

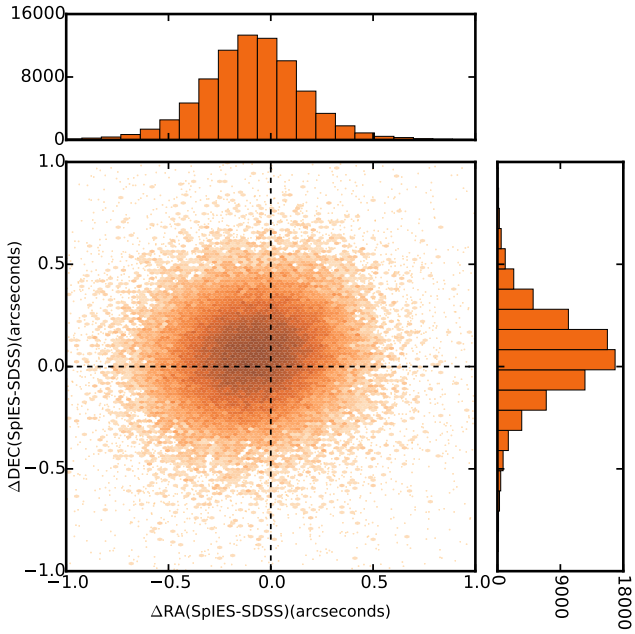

Figure 2.11 Comparison of the SpIES and SDSS astrometry for matched point sources with good flags in both surveys. Darker regions and histograms show the approximate point density. We use the mean offsets of the ΔRA and ΔDEC distributions shown here to correct the SpIES astrometry.

using the SDSS matched point sources for bright and faint sources in [4.5]. We find that the astrometric offsets to be rather consistent both for faint ([4.5] $\geq$ 20 mag) objects with ΔRA= $-0\farcs112 \pm 0\farcs0009$ and ΔDEC= $-0\farcs0370 \pm 0\farcs0007$ and for bright objects ([4.5] $\leq$ 20 mag) with ΔRA= $-0\farcs112 \pm 0\farcs0014$ and ΔDEC= $-0\farcs0376 \pm 0\farcs0012$. Regardless of magnitude, with the $0\farcs6$ pixel scale of the SpIES images, the astrometric offset is approximately one sixth of a pixel, which is similar to the values calculated in Ashby et al. (2009) where the SDWFS astrometry was compared to 2MASS.

2.5.5 Completeness and Number Counts

To estimate the completeness of our detection strategy, we employed a Monte Carlo approach where we simulated 15,000 sources (between 4% and 6% of the total number of sources) with random magnitudes between 14.5 and 28 at random positions on each AOR. The simulated sources were allowed to fall anywhere on the image, including on top of other sources, thus our completeness estimates are robust against confusion noise (see Ashby et al. 2013). Each source was modeled as

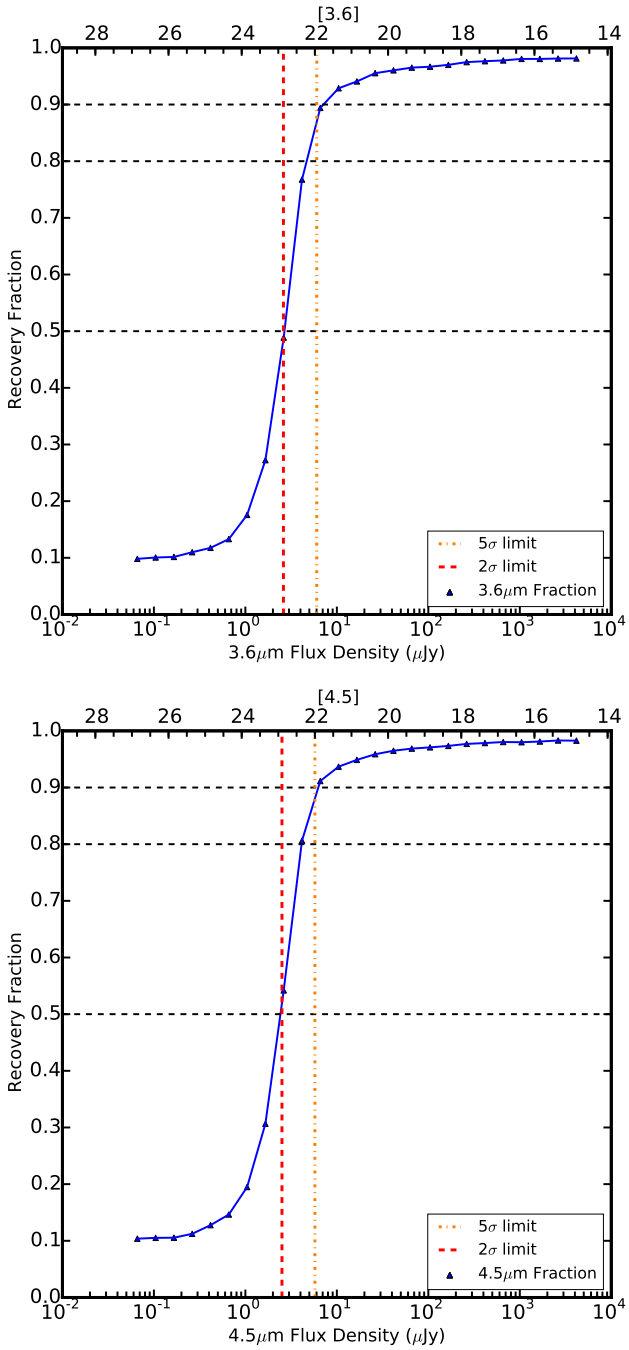

Figure 2.12 Completeness as a function of 3.6 μm flux density (and [3.6]; top) and 4.5 μm flux density (and [4.5]; bottom) of our simulated sources. The orange dot-dashed line marks the faintest detection of (5σ) objects at 6.13 μJy and 5.75 μJy at 3.6 μm and 4.5 μm, respectively; the red dashed line shows (2σ) objects at 2.58μJy and 2.47μJy at 3.6 μm and 4.5 μm, respectively, as measured from the curves in Figure 2.14. The completeness curves are less affected by artifacts at faint magnitudes since the analysis is done with simulated sources, and thus are better estimates of depth than the number counts.

a point source, having a Gaussian profile with the same FWHM as IRAC. We ran SExtractor on these simulations in the exact manner described in Section 2.5.1 and matched to a file containing the position and magnitude for each source. The tables of recovered sources for each AOR were then concatenated as before to cover the full footprint of SpIES. Number counts as a function of magnitude were plotted for both the recovered object catalog and the full simulated source catalog and the ratio of counts in each bin was calculated to estimate the completeness of the survey. Figure 2.12 presents the SpIES completeness curve for each passband, and the 90, 80, and 50 percent completeness values are quoted in Table 2.9. These measurements are performed for the entire survey field, however SpIES spans a wide range in right ascension. We therefore evaluated the completeness at different ranges in right ascension to evaluate how it changes with position. We found that the differences between the completeness curves that were computed for the full survey in Figure 2.12 and the curves computed at different locations in the SpIES survey were not significantly different, and that the differences in the 90, 80, and 50 percent complete values do not exceed $\sim$0.15 magnitudes for both the 3.6 μm and 4.5 μm measurements.

Differential number count histograms provide a visual representation of the distribution of objects of different magnitudes in a survey. They can be used to approximate the number of particular objects (stars, quasars, galaxies, etc.) that should be detected in the survey and can provide a rough estimate of the depth of the survey. The number of objects per square degree per magnitude is plotted as a function of flux density and AB magnitude in Figure 2.13 for SpIES objects detected in each band that satisfy the condition HIGH_REL>0. Shown for comparison are the differential number counts from SSDF (Ashby et al., 2013), which has a similar depth as SpIES, along with counts from the SERVS XMM field (Mauduit et al., 2012) and the S-COSMOS survey (Sanders et al., 2007), both of which are deeper than SpIES. Additionally, we show the contribution of Milky Way stars to these number counts estimated using the DIRBE Faint Source Model (FSM; Arendt et al. 1998; Wainscoat et al. 1992). At the bright end, the four surveys and the FSM all tend to align and follow a similar linear trend, indicating that the bright objects in the SpIES catalog are well represented and are mostly attributed to light in the Milky Way. The "turn over" in these

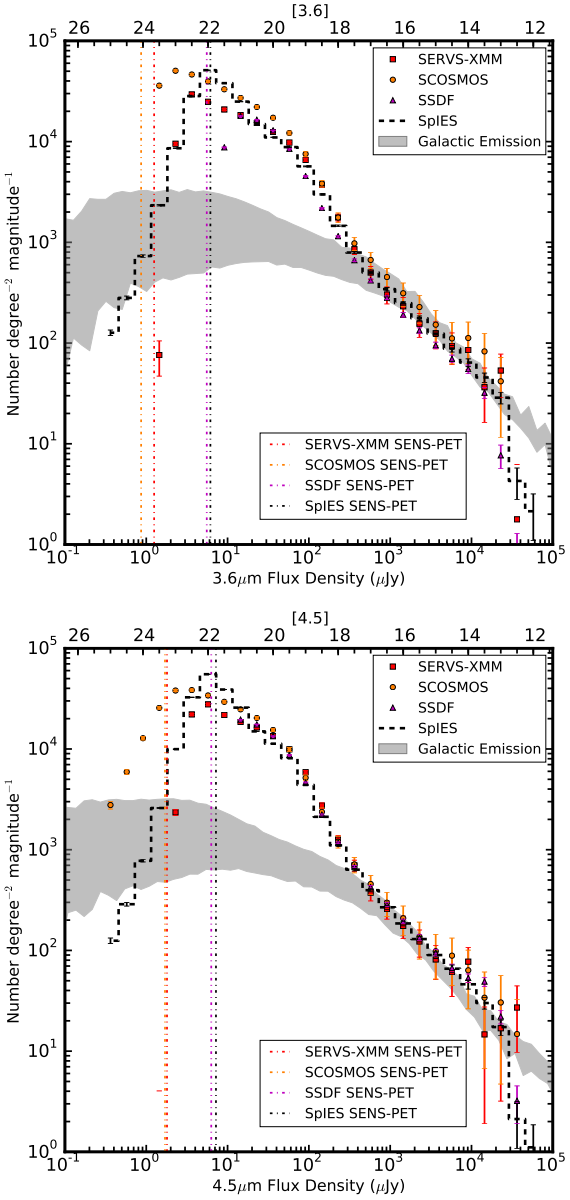

Figure 2.13 Differential number counts per magnitude over the full SpIES field for all objects with a HIGH_REL > 0. In both panels, we divide the counts by an area of 101 deg^2 which is the area covered for this footprint in each detector. Top: SpIES 5σ catalog (black dash) histogram of number of objects per square degree vs flux density (μJy) for all objects detected at 3.6 μm. Also shown are the number counts from the SERVS XMM field (Mauduit et al. 2012; red squares), S-COSMOS (Sanders et al. 2007; orange circles), and SSDF (Ashby et al. 2013; purple triangles) as comparisons. The vertical dot-dashed lines represent the SENS-PET predicted depth for each survey. As we include objects that are more than 5σ above the background, but have S/N < 5, the excess relative to other surveys near the 90% completeness limit is likely an indication of contamination by low probability sources. Bottom: The 4.5 μm number counts similar to the left panel. The grey shaded region shows the contribution of Milky Way stars using the DIRBE Faint Source Model (Arendt et al., 1998; Wainscoat et al., 1992).

histograms indicates the location of the approximate value of the depth of the survey. This is, however, an imperfect measure of the depth since artifacts tend to increase at the faint limits of a survey, resulting in more counts at fainter magnitudes.

The SpIES differential number counts in Figure 2.13 are computed for the full footprint of the survey. The spatial extent of SpIES is large enough, however, that it intersects the Galactic plane at different angles which has a small effect on the number counts, particularly for faint objects ($20 \leq \mathrm{AB} \leq 22$). For this reason the FSM, which is calculated for only a small area on the sky, is represented by a grey shaded region. To test the effect of Galactic latitude on the number counts, we split SpIES into different regions at different Galactic latitudes ($0 \leq b \leq 15$, $15 \leq b \leq 30$, and $b \geq 30$) and recompute the number counts as a function of magnitude. We find fewer faint objects are recovered for low Galactic latitudes, however as we look further off of the Galactic plane the SpIES number counts become consistent with those for surveys of similar depth (i.e., SSDF).

2.5.6 Depth

There are multiple ways of determining the depth of a survey, and the optimal value to use depends on the intended application. We computed the depth in four different ways for our analysis. First, we find the magnitude where the completeness curves turn over (see Figure 2.12). Object detection declines rapidly at this magnitude, making it a useful indicator of survey depth. An estimate of the limiting magnitude using the 90th percentile of completeness for simulated sources is [3.6]=21.75 and [4.5]=21.90. We report the 90, 80, and 50 percent complete values in Table 2.9.

Secondly, we can estimate the 5σ and 2σ depths by plotting the magnitude error as a function of magnitude (see Figure 2.14). From Figure 2.14 we determine the magnitude value where the outer edge of the curve reaches a magnitude error of $\sim$0.2 to obtain the 5σ magnitude limit. For SpIES, this limit occurs at [3.6]=21.93 and [4.5]=22.00, which corresponds to flux density values of 6.13 μJy and 5.75 μJy, respectively.

Another method to estimate depth is to perform empty aperture photometry where we placed random apertures on the images and performed source extraction in each aperture. We then made a histogram of the measurements with negative flux density values in the $1\farcs9$ aperture in an attempt

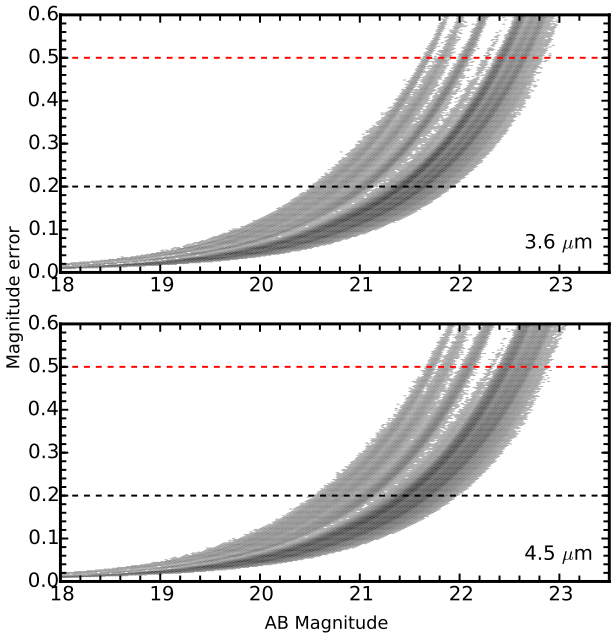

Figure 2.14 Estimation of the SpIES detection limit at 3.6 μm (top) and 4.5 μm (bottom). The grey points indicate the error in magnitude vs. magnitude. The 5σ limit occurs at a magnitude error of 0.2 (black dashed line), and the 2σ limit occurs at a magnitude error of 0.5 (red dashed line). These values are determined by propagating the error in the expression for magnitude, resulting in the ratio of noise to signal as the error in magnitude. The intersection of the right edge of the grey points with the respective magnitude error is the approximate detection threshold. Differences in shading indicates the density of points.

to eliminate contamination from sources to the background measurements. We then fit a Gaussian curve to the data to find the standard deviation in the background, σ_{bg}, across the SpIES field. We find that the $5\sigma_{bg}$ measurements are 8.14 μJy at 3.6 μm and 7.55 μJy at 4.5 μm. While this does not directly measure the depth to which we observe, it is a robust measurement of the noise in the data, including confusion noise since the apertures were randomly placed on our images.

Finally, we use the predicted limits produced by the SENS-PET[10] tool. This estimate calculates the 5σ point source depth given the background level of the survey (depending on the survey location), the exposure time, and number of repeat exposures over a single area. The SpIES depth is estimated at 6.15 μJy at 3.6 μm and 7.2 μJy at 4.5 μm using a medium background, an exposure time of 30 seconds, and four overlaps in the 'Warm IRAC Parameters' section. This tool appears

[10]http://ssc.spitzer.caltech.edu/warmmission/propkit/pet/senspet/

Table 2.9. Completeness levels

Level	3.6 μm		4.5 μm	
90% complete	21.75	7.2μJy	21.90	6.3μJy
80% complete	22.20	4.8μJy	22.37	4.1μJy
50% complete	22.82	2.7μJy	22.91	2.5μJy
5σ	21.93	6.13μJy	22.00	5.75μJy
2σ	22.87	2.58μJy	22.92	2.47μJy
5σ_{bg}	21.62	8.14μJy	21.70	7.55μJy
2σ_{bg}	22.62	3.26μJy	22.70	3.02μJy
SENS-PET 5σ	21.93	6.15μJy	21.76	7.20μJy

Note. — We give the 90, 80 and 50 percent completeness levels in AB Magnitudes and flux density of the SpIES survey from Figure 2.12 as well as the 5σ and 2σ values from Figure 2.14, the empty aperture measurements at 5σ_{bg} and 2σ_{bg}, and the SENS-PET estimates.

to calculate depths that are shallower than the measured depths; however, it is useful for making robust comparisons to other survey fields (for example, see Figure 2.13).

There are multiple reasons for the slight differences between the prediction from SENS-PET and our measurements. First, the noise estimates previously discussed in Section 2.5.2 should be considered a lower limit on the error and therefore the signal-to-noise ratios may be overestimated. Second, an overlap value of 4.0 was inserted into the SENS-PET calculator, whereas in reality the overlap of the SpIES BCD images averages to a value of $\sim$4.5 per pixel. The more coverage, the deeper the observations, so the theoretical value will be slightly brighter than reality. Finally, there could be a disparity between the background model used in SENS-PET and the measured background from the SpIES AORs, which could lead to a difference in the depth.

2.5.7 Confusion

We estimate the threshold for source confusion (the noise attributed to faint or unresolved background sources) by calculating the average number of SpIES beams per source, similar to the technique used in Ashby et al. (2009), and compare with the classical threshold limits determined in

Condon (1974) and Hogg (2001). The SpIES beam size (solid angle) is calculated using $\Omega = \pi\sigma^2$, where σ is the standard deviation of the Gaussian point spread function. Using the relation FWHM $= 2\sqrt{2\ln(2)}\sigma$ and the 'warm' IRAC FWHM values of $1\farcs95$ in the 3.6 μm detector and $2\farcs02$ in the 4.5 μm detector, we obtain a beam size of 2.155 arcsec^2 for the 3.6 μm detector and 2.312 arcsec^2 for the 4.5 μm detector. The total number of beams over the full SpIES area is 6.92×10^8 in the 3.6 μm images and 6.45×10^8 in the 4.5 μm images. Finally, taking the ratio of the number of beams to the number of objects at different detection thresholds yields an estimate for the confusion.

There are a total of $\sim 11.6 \times 10^6$ objects detected at 3.6 μm (combining the 3.6 μm-only catalog and the dual-band catalog) and $\sim 12.1 \times 10^6$ objects detected at 4.5 μm (combining the 4.5 μm-only catalog and the dual-band catalog) before applying flags for known contaminants, thus there are $\sim$60 beams per source and $\sim$53 beams per source for the full 3.6 μm and 4.5 μm detection catalogs, respectively. Taking the inverse of these two results suggest that approximately 1.6% of the detections at 3.6 μm and 1.9% of the detections at 4.5 μm are confused. Condon (1974) and Hogg (2001) found the threshold for confusion to be significant when there are fewer than 30 to 50 beams per source for number counts histograms which have power law slopes of 0.75 to 1.5. The SpIES number counts histograms have slopes of $\sim$0.85 for both bands, therefore, with 60 and 53 beams per source at 3.6 μm and 4.5 μm, respectively, we conclude that SpIES is not significantly affected by source confusion.

2.6 Diagnostics and Summary

2.6.1 Color Distributions

To test the accuracy of our data processing, we examine the distribution of magnitudes and colors of SpIES sources and compare them to known objects and infrared photometry from *WISE*. Mid-infrared color-color diagrams have proven to be effective in classifying objects, for example quasars, as shown in Lacy et al. (2004), Stern et al. (2005), and Donley et al. (2012). Unlike these previous IRAC analyses, which had access to all four channels, SpIES only observes in the first two, thus instead of a color-color diagram, we investigate the color-magnitude space shown in Figure 2.15. All SpIES sources with HIGH_REL=2 (in both bands) from the dual-band catalog are shown, along with stars and spectroscopically-confirmed quasars (drawn from the Richards et al. (2015) "master"

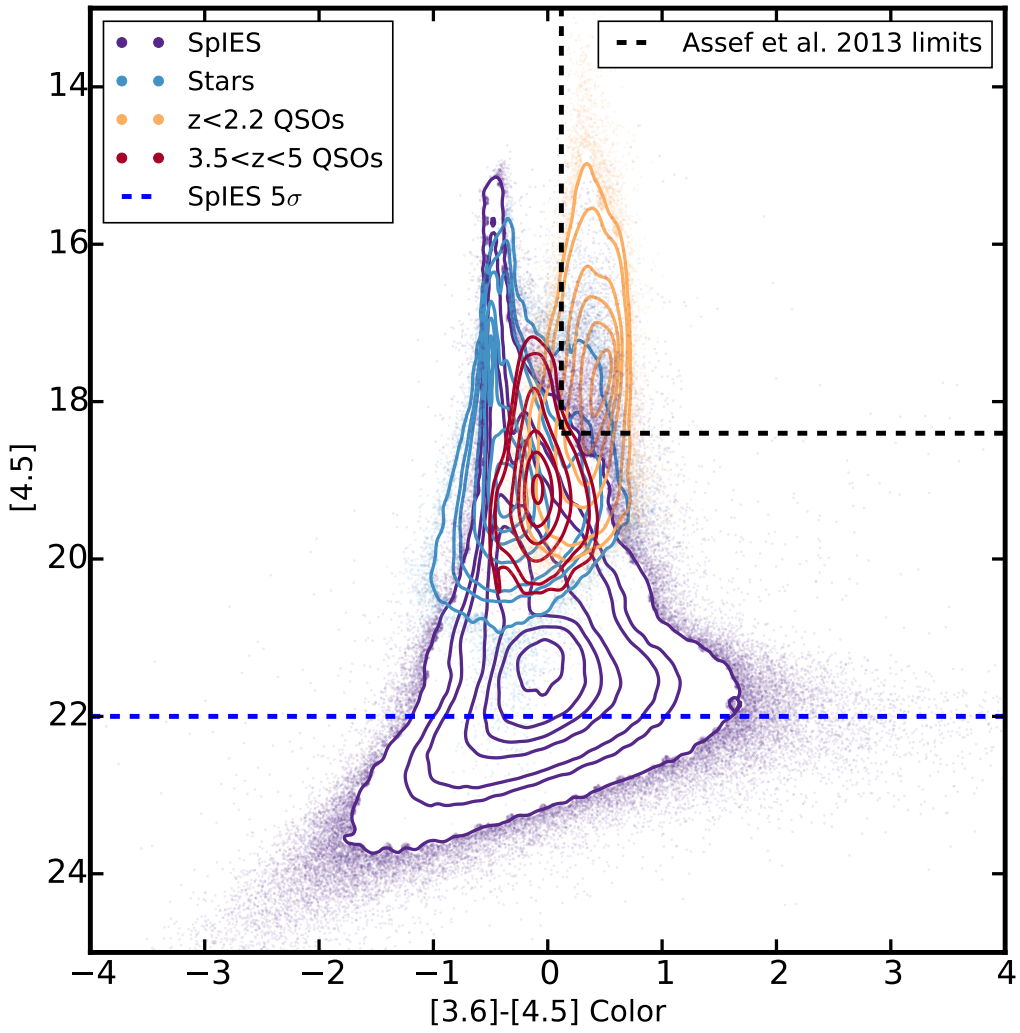

Figure 2.15 Color-magnitude diagram for SpIES objects with good photometry (i.e., HIGH_REL=2; purple). Also indicated are contours of where different objects fall in this color space. The blue contours are stars, light orange contours are known low-redshift quasars ($z \leq 2.2$), and red contours are high-redshift quasars ($3.5 \leq z \leq 5$). These additional contours are not objects matched to SpIES data, rather are SDSS detections which have *Spitzer* color information. We show the superior depth of the SpIES survey (the blue dashed line is the [4.5]=22.00 5σ line) compared to the star and quasar data from the optical. The black dashed lines represent the Assef et al. (2013) criteria for AGN selection in this color space ($W1$-$W2 \geq 0.8$), which, although very complete for low-redshift quasars (obscured and unobscured), misses most high-redshift quasars (e.g., Richards et al. 2015). We draw contours which encapsulate 10 to 90 percent of the data (in 20 percent increments) and 95 percent of the data. We additionally draw 99 percent contours for the SpIES objects (purple) and stars (blue).

quasar catalog) which are detected in both the optical and by *Spitzer*.

The "master" catalog is a combination of spectroscopically-confirmed quasars from SDSS-I/II/III (York et al., 2000; Eisenstein et al., 2011) matched with photometric sources from the AllWISE survey. To the "master" catalog, we have added new z>5 quasars from McGreer et al. (2013) and the SDSS DR12 quasar catalog (Pâris et al. 2016, in preparation). The *WISE* Vega magnitudes in the "master" catalog have been converted to AB magnitudes by adding 2.699 to W_1 and 3.339 to W_2 which is the difference in the respective zero points for the *WISE* detectors. The *WISE* AB magnitudes were then converted to the *Spitzer* AB system using the method in Section 2.3 of Richards et al. (2015) and Table 1 of Wright et al. (2010). The *Spitzer* and *WISE* detectors take images at slightly different wavelengths, and therefore observe emission from an object at slightly different locations in its spectral energy distribution. The conversion factor between the two detectors is, therefore, dependent on the color of the observed object. For our analysis, we adopt the look-up table from Richards et al. (2015) which provides the proper correction for an object with a given color and spectral index (assuming a power-law spectral energy distribution). Figure 2.15 demonstrates that SpIES can be used to distinguish various types of objects in the mid-infrared. Stars, for example, appear bluer ([3.6]-[4.5]<0) than low-redshift ($z \leq 2.2$) quasars, which tend to lie in a redder ([3.6]-[4.5]>0) region of this diagram, despite covering approximately the same magnitude range at 4.5 μm. It is also apparent that SpIES is achieving a depth that exceeds that of the spectroscopic quasar sample shown.

2.6.2 SDSS quasars

Figure 2.16 displays [3.6]−W1 vs [4.5]−W2 for the confirmed quasars in the Richards et al. (2015) "master" quasar catalog. In theory, we might expect the quasar colors to converge at the origin, however there is a deviation of the colors from the origin which can be attributed to a few factors. First, SpIES and the AllWISE surveys were conducted at different times, and thus variable quasars would shift diagonally in this color space. Additionally, there is a well-known flux underestimation bias for fainter objects in the AllWISE data attributed to an overestimation of the background caused by contamination of nearby objects, forcing the *WISE* colors to appear fainter (see the AllWISE

Explanatory Supplement[11] for more detail).

One of the goals of SpIES is to uncover new, faint quasars at high-redshift to use for clustering investigations. From Figure 2.15, it is apparent that cuts in infrared color-magnitude space alone will not cleanly select high-z quasars. However, quasar candidates can be selected using the multidimensional selection algorithm described in Richards et al. (2015) which analyzed the colors of quasars in the optical with SDSS and infrared with AllWISE. They constructed a training set of quasars comprised of objects in the AllWISE catalog that have spectroscopically confirmed quasar counterparts in SDSS (i.e., known quasars), and a test set comprised of AllWISE objects that have SDSS photometry. Using the colors of the known quasars in the training set as a Bayesian prior, probabilities were assigned to the objects in the test set based off of where they lie in the optical-infrared, multidimensional color space. We will follow this technique using, the SpIES data instead of AllWISE since it probes much deeper and has superior resolution, allowing us to better select high-redshift quasar candidates on S82.

Discovery of such objects is beyond the scope of this paper, but we show here that the SpIES data are capable of recovering such objects and have a greater ability to do so than can be achieved with the shallower *WISE* data. Figure 2.17 shows redshift and i-band magnitude histograms of sources using the "master" quasar catalog from Richards et al. (2015) as before. *WISE* only recovers 55% of the quasars in this sample, while SpIES has superior resolution and is sufficiently deep to recover 98%, including objects as faint as 22nd magnitude (i-band) and redshifts as high as 6. As one of the key science goals of the SpIES program is the discovery of faint, high-redshift quasars, we note that SpIES recovers 94% of these quasars with $z \geq 3.5$ as opposed to the 25% recovered by the *WISE* data, and 3.5% recovered after applying the Assef et al. (2013) color cuts.

2.6.3 Summary

The *Spitzer* IRAC Equatorial Survey is supplying large-area, mid-infrared imaging of the Sloan Digital Sky Survey field Stripe 82. Utilizing mapping mode with 'warm' IRAC, SpIES covers a total of $\sim$115 deg^2 of S82 (where there is $\sim$100 deg^2 of coverage in both bands) over two epochs,

[11] http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec6_3c.html#flux_under

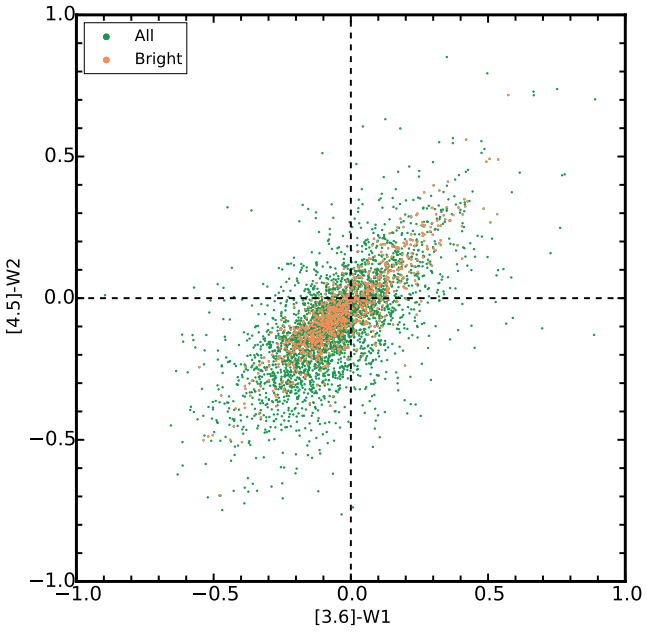

Figure 2.16 Comparison of the SpIES and WISE colors for quasars from the Richards et al. (2015) "master" catalog. WISE Magnitudes have been corrected to the IRAC AB Magnitude system in both channels. The orange points show the color of the brightest quarter of the WISE data (W1≤15.5 & W2≤15.5 WISE Vega magnitudes). In principle, we expect the points to be near the origin, however phenomena such as variability and systematics such as contamination in *WISE* W1 and W2 cause the points to deviate.

and overlaps with a wealth of ancillary data at almost every wavelength. We present the initial source catalogs for SpIES. First, a dual-band catalog containing detections in both 3.6 μm and 4.5 μm. Second, a 3.6 μm-only detected catalog and, third, a 4.5 μm-only detected catalog. In these catalogs, we report positional and photometric information, photometric errors (see Section 2.5.2), and a number of flags which are used to distinguish the high-reliability sources. The structure and analysis of these catalogs are as follows:

- We detect ∼11.6 million sources at 3.6 μm and ∼12.1 million sources at 4.5 μm, ∼5.4 million of which are matched between the two bands and are presented in the dual-band catalog. The remaining ∼6.1 million sources at 3.6 μm and ∼6.6 million sources at 4.5 μm that do not match are retained in the respective single-band only catalogs. ∼1.4, ∼3.9, and ∼1.4 million of these sources (3.6 μm-only, dual-band, 4.5μm-only) are considered reliable (i.e, HIGH_REL>0 and

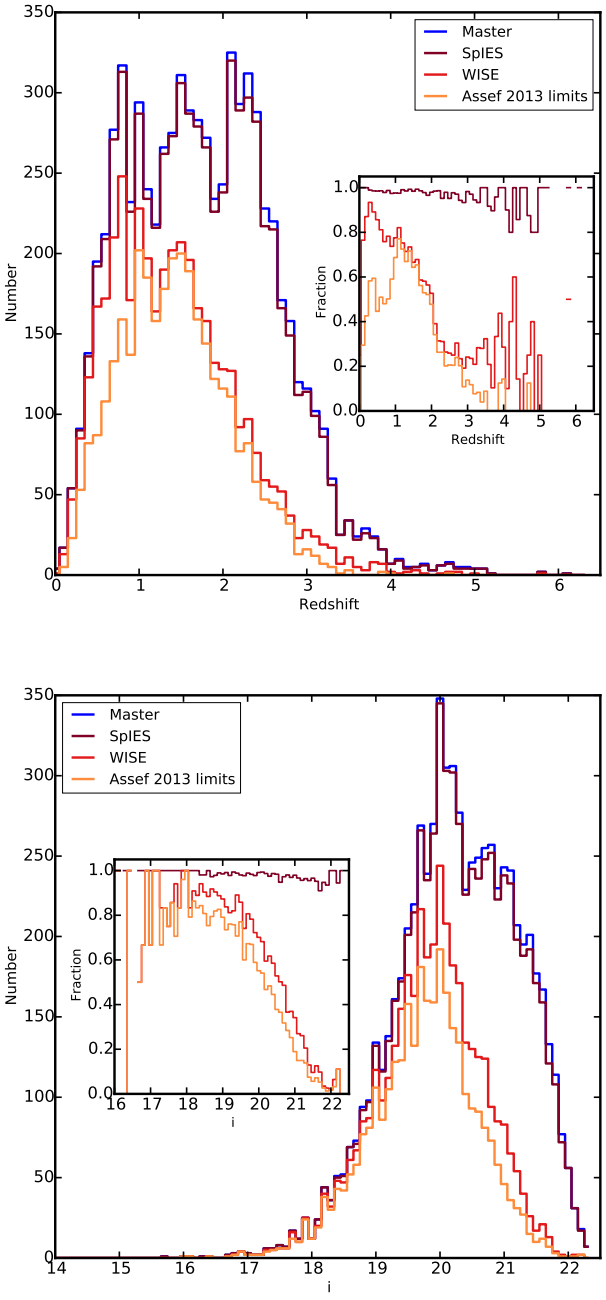

Figure 2.17 Top: Number counts of confirmed quasar redshifts from the optical samples (blue line) in the Richards et al. (2015) "master" catalog, the high-redshift quasars catalog of McGreer et al. (2013), and the SDSS DR12 quasar catalog (Pâris et al. 2016, in preparation). We overplot the redshift distribution of the matched SpIES objects (dark red) and the WISE objects (red) along with the WISE data after applying the Assef et al. (2013) constraints (orange). The number counts have been enhanced by a factor of 5 at $z \geq 3.5$ to emphasize the detections at high redshift. Bottom: The same sample of quasars, using the i-band magnitude as a depth comparison. The inset on both panels is the fraction of objects recovered for SpIES (dark red), WISE (red), and the Assef et al. (2013) objects (orange) with respect to the optical sample.

S/N>3). Much of our data analysis was performed on the dual-band catalog since it contains the most reliable sources in the survey.

- Using the objects in the dual-band catalog, we measured the positional accuracy (Figure 2.11) of the SpIES detections against point sources from SDSS, and have corrected the positions in the three catalogs for the measured offset. The standard deviation of this distribution is $0\farcs0008$ in RA and $0\farcs0006$ in DEC.

- A Monte Carlo estimate of the completeness is given in Figure 2.12, which shows that SpIES is 90% complete at AB magnitudes of 21.75 (7.2 μJy) and 21.90 (6.3 μJy) at 3.6 μm and 4.5 μm, respectively. Additionally, the SpIES number counts are compared with those from previous *Spitzer* surveys (Figure 2.13) which, along with completeness, can be used as a measure of the survey depth.

- An extensive discussion of the depth is given in Section 2.5.6 where we compare some of the different methods typically used to measure depth. We show that SpIES has a calculated 5σ depth of $\sim$6.15 μJy and $\sim$7.2 μJy and an empirical 5σ depth from Figure 2.14 of $\sim$6.13 μJy and $\sim$5.75 μJy at 3.6 μm and 4.5 μm respectively. We report the completeness and depth measurements in Table 2.9.

- One of the mission goals of SpIES was to be deep enough to detect high-redshift quasars. To test how well SpIES detects these objects, we first examined the colors of different objects in the mid-infrared in Figure 2.15, and show that SpIES has the capability to detect these high-redshift quasars from the overlap of their mid-infrared colors. From this plot we also see that SpIES detects objects much fainter than the majority of spectroscopically confirmed high-redshift quasars. Finally, the SpIES data were matched to the known quasars in the Richards et al. (2015) "master" quasar catalog and we show that SpIES detects a high percentage of quasars compared to *WISE*, particularly at $z \geq 3.5$ (Figure 2.17).

The raw imaging data is available on the SHA website, and we now release the mosaics created by the SpIES team and our three detection catalogs for public use (see Appendix 2.7).

2.7 Appendix A

2.7.1 How to Access the Raw Data, Image and Catalogs

Raw Data

The raw data for SpIES can be found on the *Spitzer* Heritage Archive website `http://sha.ipac.caltech.edu/`, where the user can input the SpIES program number (90045) and select the data type (BCD image, pBCD image, AOR).

2.7.2 Catalogs and Images

The three detection catalogs and all of the images created by the SpIES team can be found at `http://www.physics.drexel.edu/~gtr/spies/`. These files have been compressed for convenience.

2.8 Appendix B

2.8.1 The SpIES Astronomical Observation Requests

Table 2.10. Full SpIES AOR list

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
1	22634	46949888	331.760	0.008333	2013-01-15 12:22:18	2013-01-15 15:36:43	11945
2	22647	46945024	331.812	−0.008333	2013-01-14 20:03:50	2013-01-14 23:18:21	11956
3	22915	46973952	332.275	0.008333	2014-09-06 22:29:12	2014-09-07 04:59:07	23788
4	22927	46970880	332.327	−0.008333	2014-09-06 11:07:02	2014-09-06 17:36:43	23772
5	221156	46943744	332.984	0.008333	2014-09-07 10:38:37	2014-09-07 15:50:27	19075
6	22128.	46955520	333.036	−0.008333	2014-09-07 05:25:54	2014-09-07 10:37:40	19069
7	221436	46969856	333.654	0.008333	2014-09-07 21:12:36	2014-09-08 02:24:23	19072
8	221449	46966784	333.706	−0.008333	2014-09-07 16:00:01	2014-09-07 21:11:40	19064
9	221717	46955264	334.324	0.008333	2014-09-08 16:45:17	2014-09-08 21:57:09	19075
10	221730	46950144	334.376	−0.008333	2014-09-08 03:03:36	2014-09-08 08:15:14	19061
11	221958	46980864	335.032	0.008333	2014-09-13 16:18:01	2014-09-13 22:10:02	21503
12	222011	46947584	335.084	−0.008333	2014-09-13 10:24:42	2014-09-13 16:16:39	21500

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
13	222239	46977792	335.664	0.008333	2013-01-15 17:50:48	2013-01-15 23:03:16	19071
14	222251	46972416	335.716	−0.008333	2013-01-15 06:51:52	2013-01-15 12:04:24	19079
15	222520	46962176	336.334	0.008333	2013-01-15 01:38:25	2013-01-15 06:51:06	19091
16	222532	46958080	336.386	−0.008333	2013-01-14 09:22:06	2013-01-14 14:34:48	19101
17	22280.	46948608	337.004	0.008333	2013-01-14 04:08:28	2013-01-14 09:21:20	19111
18	222813	46943488	337.056	−0.008333	2013-01-13 22:45:08	2013-01-14 03:58:33	19115
19	223041	46945536	337.674	0.008333	2013-01-13 17:30:51	2013-01-13 22:44:22	19127
20	223054	46942208	337.726	−0.008333	2013-01-13 12:16:22	2013-01-13 17:29:57	19131
21	223322	46971392	338.344	0.008333	2013-01-13 06:52:39	2013-01-13 12:06:22	19143
22	223335	46968064	338.396	−0.008333	2013-01-13 01:37:58	2013-01-13 06:51:45	19147
23	22363.	46948096	339.014	0.008333	2014-09-14 21:53:31	2014-09-15 04:24:01	23819
24	223615	46953984	339.066	−0.008333	2014-09-14 15:22:11	2014-09-14 21:52:32	23810

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
25	223844	46942464	339.684	0.008333	2013-01-20 02:13:16	2013-01-20 07:25:31	19065
26	223856	46978816	339.736	−0.008333	2013-01-19 15:34:17	2013-01-19 20:46:34	19072
27	224124	46979584	340.354	0.008333	2014-09-17 15:56:43	2014-09-17 21:48:42	21487
28	224137	46976000	340.406	−0.008333	2014-09-17 10:03:25	2014-09-17 15:55:19	21483
29	22445.	46951680	341.024	0.008333	2013-01-18 22:36:06	2013-01-19 03:49:14	19102
30	224418	46962432	341.076	−0.008333	2013-01-18 17:22:04	2013-01-18 22:35:12	19106
31	224646	46952448	341.694	0.008333	2013-01-20 12:50:36	2013-01-20 18:03:20	19089
32	224659	46949376	341.746	−0.008333	2013-01-20 07:26:56	2013-01-20 12:39:42	19093
33	224927	46976512	342.364	0.008333	2013-01-22 12:04:10	2013-01-22 17:16:26	19068
34	224939	46973184	342.416	−0.008333	2013-01-22 06:42:00	2013-01-22 11:54:18	19072
35	22528.	46951168	343.034	0.008333	2013-01-18 06:42:57	2013-01-18 11:56:32	19139
36	225220	46969600	343.086	−0.008333	2013-01-18 12:07:13	2013-01-18 17:20:49	19136

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
37	225448	46960896	343.704	0.008333	2013-01-19 04:56:31	2013-01-19 10:09:46	19135
38	22551.	46952192	343.756	−0.008333	2013-01-19 10:10:41	2013-01-19 15:23:53	19132
39	225729	46946560	344.374	0.008333	2013-01-22 01:28:02	2013-01-22 06:40:47	19104
40	225742	46980352	344.426	−0.008333	2013-01-21 13:15:52	2013-01-21 18:28:43	19111
41	23010	46967040	345.044	0.008333	2014-09-22 01:15:20	2014-09-22 07:07:07	21486
42	23023	46963712	345.096	−0.008333	2014-09-22 07:08:28	2014-09-22 13:00:19	21489
43	23251	46953472	345.714	0.008333	2013-01-20 18:05:04	2013-01-20 23:18:42	19140
44	2333.	46978304	345.766	−0.008333	2013-01-20 23:35:28	2013-01-21 04:49:06	19138
45	23532	46977280	346.345	0.008333	2014-09-19 05:22:35	2014-09-19 11:13:11	21404
46	23544	46974464	346.397	−0.008333	2014-09-19 11:14:33	2014-09-19 17:05:15	21409
47	23812	46963968	347.054	0.008333	2014-09-20 05:39:13	2014-09-20 10:50:41	19061
48	23825	46959360	347.106	−0.008333	2014-09-20 10:51:35	2014-09-20 16:03:08	19064

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
49	231053	46954752	347.724	0.008333	2014-09-21 10:03:17	2014-09-21 15:15:05	19075
50	23116.	46971136	347.776	−0.008333	2014-09-21 15:15:59	2014-09-21 20:27:49	19076
51	231334	46978560	348.394	0.008333	2014-09-24 03:00:32	2014-09-24 08:13:02	19104
52	231347	46975232	348.446	−0.008333	2014-09-23 21:47:10	2014-09-24 02:59:36	19101
53	231615	46964480	349.064	0.008333	2014-09-24 08:23:13	2014-09-24 13:35:39	19097
54	231627	46960640	349.116	−0.008333	2014-09-24 13:36:33	2014-09-24 18:49:04	19101
55	231856	46951936	349.734	0.008333	2014-09-27 06:29:48	2014-09-27 11:42:33	19133
56	23198.	46961664	349.786	−0.008333	2014-09-26 15:23:00	2014-09-26 20:35:32	19124
57	232136	46949120	350.404	0.008333	2013-01-29 18:01:35	2013-01-29 23:13:56	19078
58	232149	46944000	350.456	−0.008333	2013-01-29 12:48:22	2013-01-29 18:00:42	19083
59	232417	46972928	351.074	0.008333	2013-01-27 06:05:06	2013-01-27 11:18:22	19121
60	232430	46968320	351.126	−0.008333	2013-01-27 00:51:00	2013-01-27 06:04:12	19125

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
61	232658	46959104	351.744	0.008333	2013-01-25 07:06:46	2013-01-25 12:20:32	19159
62	232711	46954240	351.796	−0.008333	2013-01-24 10:56:43	2013-01-24 16:10:41	19172
63	232939	46944512	352.414	0.008333	2013-01-27 16:42:59	2013-01-27 21:56:29	19134
64	232951	46979072	352.466	−0.008333	2013-01-27 11:28:32	2013-01-27 16:42:05	19138
65	233220	46979840	353.084	0.008333	2013-01-25 17:44:54	2013-01-25 22:58:59	19171
66	233232	46976256	353.136	−0.008333	2013-01-25 12:21:49	2013-01-25 17:35:53	19176
67	23350.	46966272	353.754	0.008333	2013-09-24 20:07:15	2013-09-25 01:20:29	19129
68	233513	46962944	353.806	−0.008333	2013-09-24 14:53:12	2013-09-24 20:06:19	19126
69	233741	46952960	354.424	0.008333	2013-09-27 05:28:49	2013-09-27 10:42:24	19155
70	233754	46949632	354.476	−0.008333	2013-09-26 23:28:02	2013-09-27 04:41:30	19150
71	234022	46950656	355.094	0.008333	2013-09-26 18:13:41	2013-09-26 23:26:56	19138
72	234035	46945792	355.146	−0.008333	2013-09-26 07:33:02	2013-09-26 12:46:05	19131

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
73	23433.	46965760	355.764	0.008333	2013-09-23 23:03:13	2013-09-24 04:16:06	19092
74	234315	46971648	355.816	−0.008333	2013-09-23 02:21:37	2013-09-23 07:34:09	19076
75	234544	46961152	356.434	0.008333	2013-09-24 09:39:16	2013-09-24 14:51:42	19089
76	234556	46957568	356.486	−0.008333	2013-09-24 04:16:58	2013-09-24 09:29:46	19086
77	234824	46947072	357.104	0.008333	2013-09-27 15:57:49	2013-09-27 21:11:02	19124
78	234837	46943232	357.156	−0.008333	2013-09-27 10:43:48	2013-09-27 15:56:53	19120
79	23515.	46969088	357.774	0.008333	2014-10-01 09:05:27	2014-10-01 14:17:10	19081
80	235118	46980096	357.826	−0.008333	2014-10-01 14:18:04	2014-10-01 19:29:53	19083
81	235346	46970624	358.444	0.008333	2014-10-01 03:44:12	2014-10-01 08:55:37	19062
82	235359	46967808	358.496	−0.008333	2014-09-30 22:32:02	2014-10-01 03:43:17	19052
83	235627	46956288	359.114	0.008333	2014-10-06 05:30:49	2014-10-06 10:43:44	19132
84	235639	46953216	359.166	−0.008333	2014-10-06 00:17:04	2014-10-06 05:29:53	19130

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
85	23598.	46956800	359.784	0.008333	2014-10-05 18:54:42	2014-10-06 00:07:17	19119
86	235920	46974976	359.836	−0.008333	2014-10-05 08:23:09	2014-10-05 13:35:37	19112
87	0148	46974208	0.454	0.008333	2013-02-01 07:42:52	2013-02-01 12:56:56	19178
88	021.	46964736	0.506	−0.008333	2013-02-01 02:27:55	2013-02-01 07:41:58	19184
89	0429	46958848	1.124	0.008333	2013-09-30 00:07:47	2013-09-30 05:20:39	19103
90	0442	46954496	1.176	−0.008333	2013-09-30 05:21:33	2013-09-30 10:34:33	19106
91	0710	46942720	1.794	0.008333	2013-09-30 16:20:56	2013-09-30 21:33:54	19103
92	0723	46978048	1.846	−0.008333	2013-09-30 21:43:31	2013-10-01 02:56:33	19106
93	0951	46969344	2.464	0.008333	2013-10-01 02:57:20	2013-10-01 08:10:19	19101
94	0103.	46965504	2.516	−0.008333	2013-10-01 19:30:29	2013-10-02 00:43:36	19108
95	01232	46964224	3.134	0.008333	2013-10-02 00:53:18	2013-10-02 06:06:01	19104
96	01244	46960384	3.186	−0.008333	2013-10-02 07:03:50	2013-10-02 12:16:38	19107

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
97	01512	46950912	3.804	0.008333	2013-10-03 01:26:08	2013-10-03 06:39:09	19108
98	01525	46946304	3.856	−0.008333	2013-10-03 19:40:08	2013-10-04 00:53:17	19119
99	01753	46975488	4.474	0.008333	2013-10-04 22:50:39	2013-10-05 04:03:46	19127
100	0186.	46953728	4.526	−0.008333	2013-10-05 10:58:43	2013-10-05 16:11:57	19132
101	02034	46972672	5.144	0.008333	2013-10-05 16:12:44	2013-10-05 21:25:58	19127
102	02047	46970112	5.196	−0.008333	2013-10-05 21:35:40	2013-10-06 02:49:00	19129
103	02315	46958336	5.814	0.008333	2014-10-08 18:12:29	2014-10-08 23:24:09	19074
104	02327	46955776	5.866	−0.008333	2014-10-08 12:50:34	2014-10-08 18:02:28	19064
105	02556	46944256	6.484	0.008333	2014-10-09 06:57:36	2014-10-09 12:09:18	19072
106	0268.	46956544	6.536	−0.008333	2014-10-09 01:45:09	2014-10-09 06:56:41	19064
107	02836	46970368	7.154	0.008333	2014-10-12 21:56:43	2014-10-13 03:09:07	19118
108	02849	46967552	7.206	−0.008333	2014-10-13 03:19:55	2014-10-13 08:32:24	19120

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
109	03117	46968576	7.824	0.008333	2014-10-13 08:33:11	2014-10-13 13:45:38	19115
110	03130	46962688	7.876	−0.008333	2014-10-13 13:46:32	2014-10-13 18:59:06	19118
111	03358	46955008	8.494	0.008333	2014-10-15 02:56:10	2014-10-15 08:09:03	19130
112	03411	46948352	8.546	−0.008333	2014-10-14 21:33:39	2014-10-15 02:46:29	19126
113	03639	46979328	9.164	0.008333	2014-10-13 21:13:54	2014-10-14 02:26:13	19102
114	03651	46973440	9.216	−0.008333	2014-10-14 02:27:07	2014-10-14 07:39:31	19103
115	03920	46963200	9.834	0.008333	2014-10-15 08:10:08	2014-10-15 13:22:49	19115
116	03932	46958592	9.886	−0.008333	2014-10-15 19:29:13	2014-10-16 00:42:02	19120
117	0420.	46960128	10.504	0.008333	2014-10-16 06:05:16	2014-10-16 11:17:42	19119
118	04213	46957056	10.556	−0.008333	2014-10-16 00:42:52	2014-10-16 05:55:36	19115
119	04441	46946816	11.174	0.008333	2014-10-16 17:26:02	2014-10-16 22:38:30	19117
120	04454	46942976	11.226	−0.008333	2014-10-16 22:39:24	2014-10-17 03:51:59	19119

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
121	04722	46971904	11.844	0.008333	2014-10-18 16:32:54	2014-10-18 21:45:50	19133
122	04735	46968832	11.896	−0.008333	2014-10-18 05:01:26	2014-10-18 10:14:13	19126
123	0503.	46959872	12.514	0.008333	2014-10-19 03:09:25	2014-10-19 08:22:23	19130
124	05015	46966016	12.566	−0.008333	2014-10-18 21:46:41	2014-10-19 02:59:33	19127
125	05244	46956032	13.184	0.008333	2014-10-17 23:34:29	2014-10-18 04:46:50	19103
126	05256	46952704	13.236	−0.008333	2014-10-17 18:21:21	2014-10-17 23:33:34	19098
127	05524	46980608	13.854	0.008333	2014-10-20 21:34:09	2014-10-21 02:47:05	19136
128	05537	46976768	13.906	−0.008333	2014-10-20 16:20:23	2014-10-20 21:33:14	19133
129	1490	46967296	27.254	0.008333	2013-10-18 17:40:42	2013-10-18 22:51:25	18965
130	14913	46963456	27.306	−0.008333	2013-10-18 22:52:20	2013-10-19 04:03:05	18970
131	15141	46965248	27.924	0.008333	2013-10-19 11:07:27	2013-10-19 16:18:12	18968
132	15154	46961408	27.976	−0.008333	2013-10-19 16:19:07	2013-10-19 21:29:53	18970

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
133	15422	46951424	28.594	0.008333	2013-10-20 13:16:13	2013-10-20 18:27:12	18975
134	15435	46947328	28.646	−0.008333	2013-10-20 18:37:01	2013-10-20 23:47:36	18977
135	1573.	46966528	29.264	0.008333	2013-10-20 23:48:22	2013-10-21 04:58:50	18970
136	15715	46972160	29.316	−0.008333	2013-10-21 10:44:00	2013-10-21 15:54:36	18977
137	15944	46961920	29.934	0.008333	2013-10-21 15:55:22	2013-10-21 21:05:52	18970
138	15956	46957824	29.986	−0.008333	2013-10-21 21:06:45	2013-10-22 02:17:18	18971
139	2224	46977536	30.604	0.008333	2013-10-24 13:46:24	2013-10-24 18:57:49	19021
140	2237	46974720	30.656	−0.008333	2013-10-24 18:58:43	2013-10-25 00:10:09	19022
141	255.	46948864	31.274	0.008333	2013-10-25 00:20:14	2013-10-25 05:31:32	19014
142	2518	46959616	31.326	−0.008333	2013-10-25 08:39:48	2013-10-25 13:51:12	19021
143	2746	46950400	31.944	0.008333	2013-10-27 01:14:33	2013-10-27 06:26:30	19045
144	2759	46946048	31.996	−0.008333	2013-10-27 11:23:55	2013-10-27 16:35:58	19052

Table 2.10 (cont'd)

#	SpIES AOR Label (IRACPC-SPIES-)	AOR Key	RA (J2000) (degrees)	DEC (J2000) (degrees)	Obs. Start	Obs. End	Integration time (s)
145	21027	46944768	32.614	0.008333	2013-10-28 18:25:44	2013-10-28 23:37:49	19059
146	21039	46981120	32.666	−0.008333	2013-10-28 23:38:43	2013-10-29 04:50:49	19061
147	2138.	46945280	33.284	0.008333	2013-10-31 13:24:25	2013-10-31 18:37:17	19090
148	21320	46964992	33.336	−0.008333	2013-10-31 18:47:26	2013-11-01 00:00:17	19093
149	21548	46957312	33.954	0.008333	2013-11-01 18:29:22	2013-11-01 23:42:19	19098
150	2161.	46947840	34.006	−0.008333	2013-11-01 23:43:13	2013-11-02 04:56:10	19100
151	21829	46941952	34.624	0.008333	2014-11-08 22:10:40	2014-11-09 03:23:03	18801
152	21842	46975744	34.676	−0.008333	2014-11-08 16:48:59	2014-11-08 22:01:17	18801
153	22110	46977024	35.294	0.008333	2014-11-08 11:35:49	2014-11-08 16:47:54	18802
154	22123	46973696	35.346	−0.008333	2014-11-08 06:14:41	2014-11-08 11:26:35	18801

Note. — This table shows the single epoch AORs in the SpIES survey. Each set of 2 AORs overlap approximately the same area, for instance, AOR number 1 and 2 overlap with each other to image one region on the sky.

Chapter 3: The Clustering of High-Redshift ($2.9 \leq z \leq 5.1$) Quasars in SDSS Stripe 82

Abstract

We present a measurement of the two-point autocorrelation function of photometrically-selected, high-z quasars over $\sim$ 100 deg^2 on the Sloan Digitial Sky Survey Stripe 82 field. Selection is performed using three machine-learning algorithms in a six-dimensional, optical/mid-infrared color space. Optical data from the Sloan Digitial Sky Survey is combined with overlapping deep mid-infrared data from the *Spitzer* IRAC Equatorial Survey and the *Spitzer*-HETDEX Exploratory Large-area survey. Our selection algorithms are trained on the colors of known high-z quasars. The selected quasar sample consists of 1378 objects and contains both spectroscopically-confirmed quasars and photometrically-selected quasar candidates. These objects span a redshift range of $2.9 \leq z \leq 5.1$ and are generally fainter than $i = 20.2$; a regime which has lacked sufficient number density to perform autocorrelation function measurements of photometrically-classified quasars. We compute the angular correlation function of these data, marginally detecting quasar clustering. We fit a single power-law with an index of $\delta = 1.39 \pm 0.618$ and amplitude of $\theta_0 = 0.71 \pm 0.546$ arcmin. A dark-matter model is fit to the angular correlation function to estimate the linear bias. At the average redshift of our survey ($\langle z \rangle = 3.38$) the bias is $b = 6.78 \pm 1.79$. Using this bias, we calculate a characteristic dark-matter halo mass of 1.70–9.83$\times 10^{12} h^{-1} M_{\odot}$. Our bias estimate suggests that quasar feedback intermittently shuts down the accretion of gas onto the central super-massive black hole at early times. If confirmed, these results hint at a level of luminosity dependence in the clustering of quasars at high-z.

3.1 Introduction

In the present day Universe, super-massive black holes (SMBHs) reside at the center of most, if not all, galaxies with $M_{\star} \gtrsim 10^{10} M_{\odot}$, in which star-formation has almost completely ceased (e.g.,

Bell, 2008; Bower et al., 2017). It is commonly accepted that every massive galaxy has undergone *at least* one quasar phase within its lifetime (Soltan 1982; Richstone et al. 1998). In this quasar phase, baryons in an accretion disk lose angular momentum through mechanisms such as viscous transfer, and eventually are accreted by the SMBH (Salpeter 1964, Lynden-Bell 1969, Rees 1984). The friction in the disk heats the baryons causing the disk to shine in the optical, ultraviolet (UV), and X-rays.

Quasars, defined here as a luminous active galactic nuclei with bolometric luminosity $L_{\rm bol}$ above $\sim 10^{45}$ erg s^{-1}, are among the most luminous objects in the Universe, and therefore, can trace the large scale structure out to high redshift. Galaxies are thought to reside in the peaks in the dark-matter (DM) distribution, and are generally biased tracers of the underlying DM (e.g., Dekel & Lahav, 1999; Sheth & Tormen, 1999; Peacock, 1999). This relationship can be quantified by measuring the linear bias parameter, b. As an initial guide, we define b as:

$$\delta_{\rm Q} = b\ \delta_{\rm DM} \tag{3.1}$$

where $\delta_{\rm Q}$ is the quasar density contrast and $\delta_{\rm DM}$ is the mass density contrast. Defining the two-point auto-correlation function (2PCF) as $\xi(r) = \langle \delta(x)\delta(x+r) \rangle$, where r is the separation between two local over-densities, leads to

$$\xi_{\rm Q}(r) = b_{\rm Q}^2\ \xi_{\rm DM}(r) \tag{3.2}$$

where $\xi_{\rm Q}$ is the quasar two-point correlation function and $\xi_{\rm DM}$ is the DM correlation function. The 2PCF is defined as the joint probability of finding a pair of objects having a particular separation in two volume elements (Totsuji & Kihara 1969; Peebles 1980) and is a statistic commonly employed to measure the spatial distribution of galaxies (e.g., Zehavi et al., 2011), hydrogen gas in absorption (e.g., Bautista et al., 2017) and, in this case, quasars. In practice, the 2PCF is calculated as the excess probability, above a random Poisson distribution, of finding a pair of objects within an annulus between r and $r + \delta r$ (Peebles, 1980; Martinez & Saar, 2002; Feigelson & Babu, 2012).

Table 3.1. Selected quasar clustering measurements.

Survey	Area / deg^2	N_Q	Magnitude range	Selection	z-range	Type[a]	Reference
NDWFS+AGES	7.9	585	$I \leq$21.5, $[3.6] = 6.4\mu$m	X+R+MIR	$0.25 < z < 0.8$	C/s	Hickox et al. (2009)
NDWFS+AGES	9	924	$R \sim$25.0, $[3.6] = 6.4\mu$m	opt+MIR	$0.7 < z < 1.8$	C/b	Hickox et al. (2011)
PRIMUS+DEEP2	$\sim$10	$\sim$1 000	$i_{AB} \sim$23.5	X+R+MIR	$0.2 < z < 1.2$	C/s	Mendez et al. (2016)
SpIES+SHELA	**$\approx$100**	**1 378**	$i \sim$23.5, $[3.6] = 6.1\mu$m	**opt+MIR**	$\mathbf{z > 2.9}$	**A/b**	**This study**
2SLAQ	$\approx$150	6 374	$20.85 < g < 21.85$	cb/UVX	$0.3 < z < 2.9$	A/s	da Ângela et al. (2008)
HSC	172	901	$21.0 < i < 23.5$	opt+NIR	$3.4 < z < 4.6$	C/b	He et al. (2018)
ACTxSDSS	324	$\sim$24 000	$17.75 < i < 22.45$	[b]XDQSO	$z \approx 1.4$	C/p	Sherwin et al. (2012)
2QZ	$\approx$445	13 989	$18.25 < b_J < 20.85$	cb/UVX	$0.8 < z < 2.1$	A/s	Porciani et al. (2004)
2QZ	721	22 655	$18.25 < b_J < 20.85$	cb/UVX	$0.3 < z < 2.2$	A/s	Croom et al. (2005)
eBOSS Y1Q	1168	$\sim$70 000	$g \leq 22.0$ or $r \leq 22.0$	XDQSO	$0.9 < z < 2.1$	A/s	Rodríguez-Torres et al. (2016)
eBOSS	1200	$\sim$69 000	$g \leq 22.0$ or $r \leq 22.0$	XDQSOz	$0.9 < z < 2.2$	A/p	Laurent et al. (2017)
eBOSS BAO	2044	147 000	$g \leq 22.0$ or $r \leq 22.0$	XDQSOz	$0.8 < z < 2.2$	A/s	Ata et al. (2017)

Table 3.1 (cont'd)

Survey	Area / deg^2	N_Q	Magnitude range	Selection	z-range	Type[a]	Reference
SPTxWISE	2500	107 469	$W2 \leq 15$	IR	$\langle z \rangle \sim 1$	C/p	Geach et al. (2013)
BOSSxLyα	3275	61 342	$g \leq 22.0$ or $r \leq 21.85$	XDQSO	$2.0 < z < 3.5$	C/s	Font-Ribera et al. (2013)
WISE	3363	176 467	$W2 < 15.05$	IR	$z \sim 1$	A/p	Donoso et al. (2014)
WISE	3422	175 911	$W2 < 15.05$	IR	$z \sim 1$	A/C/p	[c]DiPompeo et al. (2016)
BOSS DR9	3600	27 129	$g \leq 22.0$ or $r \leq 21.85$	XDQSO	$2.2 < z < 2.8$	A/s	White et al. (2012)
SDSS DR5	∼4000	38 208	$i \leq 19.1$	cb	$0.1 < z < 5.0$	A/C/s	Shen et al. (2009)
SDSS DR5	4013	30 239	$i \leq 19.1$	cb	$0.3 < z < 2.2$	A/s	Ross et al. (2009)
SDSS DR5	4041	4 426	$i \leq 20.2$	cb	$2.9 < z < 5.4$	A/s	Shen et al. (2007)
SDSS DR4	∼6670	∼300 000	$g < 21$	[d]KDE cb	$0.75 < z < 2.28$	A/p	[e]Myers et al. (2007)
BOSS DR12	6950	55 826	$g \leq 22.0$ or $r \leq 21.85$	XDQSO	$2.2 < z < 2.8$	A/s	Eftekharzadeh et al. (2015)

Note. — [a]Measurement of the auto-correlation function(A), the cross-correlation function(C), using photometric(p)/spectroscopic(s) redshifts, or a combination of both (b). The studies in this table take advantage of the properties of quasars in X-ray (X), radio (R), mid-infrared (MIR), near-infrared (NIR), and optical (opt) wavelengths and use color-boxes (cb) and/or with machine-learning techniques for selection.
[b]"Extreme Deconvolution", see Bovy et al. (2011).
[c]DiPompeo et al. (2014, 2015) performed similar analyses on earlier WISE datasets.
[d]Kernel Density Estimator, see Richards et al. (2009).
[e]Myers et al. (2006) performed a similar analysis on SDSS DR1 data.

The 2PCF, and the corresponding bias, have been measured for quasars as a function of different observable properties, including redshift, luminosity and color; Table 3.1 presents a summary of recent results. Studies of quasar clustering as a function of luminosity (da Ângela et al., 2008; Shen et al., 2009; Eftekharzadeh et al., 2015; Chehade et al., 2016) have shown that the bias is very weakly, if at all, dependent on absolute quasar UV/optical luminosity. In fact, both Shen et al. (2013) and Krolewski & Eisenstein (2015) found no luminosity dependence of quasar clustering at low-z by studying the cross-correlation between galaxies and quasars. This result implies that quasars all live in the most massive dark-matter halos, regardless of how bright the quasar shines. Aird et al. (2018), however, suggested that the observed lack of luminosity dependence on quasar clustering may be due to a selection effect depending on the type of galaxy in which the AGN resides (star-forming or quiescent).

Croom et al. (2005), Myers et al. (2007), and Ross et al. (2009) have demonstrated that the bias evolves with redshift, increasing at higher redshift until the peak of quasar activity at $z \sim 2.5$. These studies were performed with large number densities of either spectroscopically-confirmed or photometrically-selected quasars, driving down Poisson noise in the clustering measurement (see Table 3.1). Interestingly, however, due to the evolution of the underlying DM density field, the masses of the halos quasars inhabit remains approximately constant at $M_{\rm halo} \sim 2-3\times10^{12}h^{-1}M_{\odot}$ from redshifts $z \sim 2.5$ to the present day. Shen et al. (2007) performed a similar analysis of the luminous, high-z ($2.9 \leq z \leq 5.4$) confirmed quasars from the Sloan Digital Sky Survey (SDSS; York et al. 2000) Data Release 5. Despite having low number densities (~ 1 quasar deg^{-2}), their study detected a large clustering signal, which implied that the bias increases rapidly beyond $z \sim 2.5$, yielding a large increase in the DM halo mass estimate with redshift.

Clustering has also been studied as function of quasar color, which is a proxy for quasar type. Here the results are not so definitive. Hickox et al. (2011) measured the clustering of both obscured and unobscured quasars, as defined by an optical-to-IR flux ratio (specifically $R_{AB} - [4.5]_{\rm Vega}$=6.0; Hickox et al. 2007), with bluer objects being classed as unobscured quasars. Hickox et al. (2011) reported "marginally stronger clustering" for the obscured quasars compared to the unobscured

population, with the consequence that dust-obscured quasars tend to reside in more massive DM halos than 'dust-free' quasars. Donoso et al. (2014), using a similar selection to Hickox et al. (2011), similarly found that obscured AGNs inhabit denser environments than unobscured AGNs. DiPompeo et al. (2014, 2015, 2016), in finding a less significant difference between the clustering of obscured and unobscured quasars, noted that Donoso et al. (2014) discounted several critical systematics that affect the amplitude of quasar clustering measurements.

Linking the measurements of the 2PCF and of the corresponding bias to quasar and host galaxy physical parameters is paramount in understanding the relationship between the observable Universe and the underlying DM distribution. These observables can then be used to direct theories and models of galaxy and quasar formation and evolution. One model that links the DM distribution, quasar activity and the associated environment was presented in Hopkins et al. (2007a). The simulations in that investigation predicted the clustering of the quasar population through the implementation of three different quasar feedback models. Quasar feedback works against gravity by forcing material away from the SMBH through radiation pressure, thus limiting the material that can accrete onto, and increase the mass of, the SMBH. This process can ultimately shut down the quasar phase and cause the SMBH to cease growing. Measuring the spatial distribution of quasars, particularly in the early Universe, can test the predictions made by Hopkins et al. (2007a).

Testing these models requires surveys to push beyond the redshift peak in the quasar epoch ($2 \leq z \leq 3$; Schmidt et al. 1995; Boyle et al. 2000), *and* delve further down the quasar luminosity function (QLF). Current surveys are underway to address this question, for example the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016), which will be able to select quasars out to $z \sim 3.5$ (Myers et al., 2015); however, the majority of existing quasar surveys are either designed to observe rest-frame UV bright quasars and/or are focused on $z < 2$, thus new data and analysis is needed.

In this paper we present the first measurements of the autocorrelation function of optical+infrared-selected quasars at $z > 2.9$ with the 2PCF. This approach is made possible by the combination of deep optical data from the SDSS Stripe 82 coadded catalog (Annis et al. 2014; Jiang et al. 2014) as

well as new deep, overlapping *Spitzer* coverage from the *Spitzer* IRAC Equatorial Survey (SpIES; Timlin et al. 2016) and the *Spitzer*-HETDEX Exploratory Large-area (SHELA; Papovich et al. 2016) survey. Following the work of Richards et al. (2015), we combine the color information from the optical and mid-infrared (MIR) and employ machine-learning algorithms to classify faint, high-z quasar candidates using their photometric colors.

Traditionally, large numbers of quasars have been detected from *Spitzer* surveys alone. Quasars tend to lie in a specific location in MIR color space so selection can be performed through various color cuts (Lacy et al. 2004; Stern et al. 2005). These constraints, while effective, lead to an increasing amount of contamination, particularly at high-z where quasar colors overlap with the stellar locus, resulting in a higher level of incompleteness in the selection (Assef et al. 2010; Donley et al. 2012). Donley et al. (2012) added a power law selection requirement for classification, which significantly reduced contamination; however, quasar spectra are not necessarily power laws in the MIR (Richards et al., 2015). Similarly, optical-only selections have found a large number of new quasars in SDSS alone; however, these techniques suffer from incompleteness at $z \sim 3.5$ (Richards et al., 2006; Worseck & Prochaska, 2011), where quasars have colors near that of the stellar locus. The combination of optical and infrared colors allows for more robust classifications, particularly at high-z, which is essential for this study (see Section 3.2.2).

In this paper, we measure the clustering strength of photometrically-selected quasar candidates. We compare these measurements to the theoretical predictions for DM clustering to draw inferences on various physical parameters such as DM halo mass and AGN feedback mechanisms in the early Universe. In Section 3.2 we discuss the data used in this study, as well as the techniques to select quasar candidates. Section 3.3 provides further details about the two-point autocorrelation function definition and uses. We present our results in Section 3.4 and discuss the implications of our results, comparing to several quasar feedback models in Section 3.5. We summarize and conclude in Section 3.6. The Appendices give further relevant and supplemental information. Throughout this paper, we assume a spatially flat ΛCDM model, consistent with the latest Cosmic Microwave Background (CMB; Planck Collaboration et al., 2016) and Baryon Acoustic Oscillations (BOSS;

Alam et al., 2016) datasets: $\Omega_{\rm m} = 0.275$, $H_0 = 70$ km s^{-1} Mpc^{-1} and $\sigma_8 = 0.77$, unless otherwise stated. All colors and magnitudes in this data set were corrected for Galactic extinction using the parameters (for $R_V = 3.1$) given in Table 6 of Schlafly & Finkbeiner (2011). We calculate magnitudes on the AB scale, which has a flux density zeropoint of 3631 Jy (Oke & Gunn, 1983b).

3.2 Data and Selection

In this section we describe our datasets including the SpIES and SHELA surveys and the optical data on SDSS Stripe 82. Following that, we describe our 'test' and 'training' sets required to classify our data. Finally, we present the classification algorithms. The final sample of quasar (candidates) we generate is given in Section 3.2.5.

3.2.1 SpIES, SHELA and SDSS Stripe 82

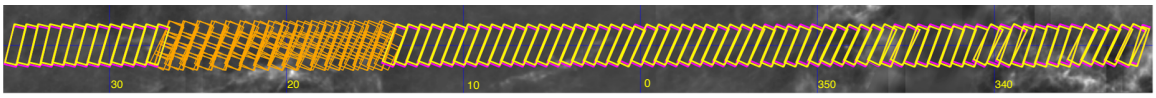

Figure 3.1 Superimposed on the 100μm IRAS dust map (Schlegel et al., 1998), we show the mid-infrared coverage mask on S82 from the SHELA (orange squares) and SpIES (yellow/purple rectangles) survey. These surveys cover $\sim$ 120 square degrees on S82 (approximately on third of the full area) and are deep enough to detect quasars out to $z = 6$. Each SpIES observation (individual yellow/purple rectangle) spans a range of 0.82° in RA (horizontal axis) and 2° in DEC (vertical axis), covering an area of $\sim$ 1.63 deg^2 each.

Covering approximately a third of S82 ($-60° \leq \alpha \leq 60°$; $-1.25° \leq \delta \leq 1.25°$), the SpIES survey was designed to span a large area ($\sim$100 deg^2, centered at $\delta = 0$; see Figure 3.1) and to probe sufficiently deep to select faint, high-z quasars; quasars which were undetected by the Wide-field Infrared Survey Explorer (*WISE*; Wright et al. 2010). The SpIES catalogs reported the photometry and photometric errors for $\sim$5.4 million objects at 3.6 μm and 4.5 μm. Using SpIES, we are able to detect quasars as faint as $i \sim 22$ with high reliability (Timlin et al., 2016). SpIES is also optimally located to surround existing *Spitzer* data from SHELA (Papovich et al., 2016), forming a long stripe of deep MIR coverage on S82 (see Figure 3.1).

The SHELA survey was designed to be used alongside the Hobby-Eberly Dark Energy Experiment (HETDEX; Hill et al. 2008) to perform dark-energy measurements, requiring deep infrared data. With depths greater than that of SpIES, SHELA provided an additional $\sim$24 deg^2 of deep infrared

coverage on S82 (see Figure 3.1). In total, SHELA detected ∼2 million objects down to *Spitzer* magnitude depths of [3.6]=22.0 and [4.5]=22.6 (compared to 21.9 and 22.0, respectively, for SpIES). In tandem, SpIES and SHELA provide ∼120 deg^2 (accounting for overlapping coverage: Figure 3.1) of deep, MIR data on S82; data necessary to, along with optical colors, select faint, high-z quasars.

Optical photometric data come from the full SDSS-I/II (York et al., 2000) data release as well as the SDSS-III/Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011, Dawson et al. 2013). Of particular interest for this study is the S82 coadded catalog (Annis et al. 2014; Jiang et al. 2014). Imaged with the five optical SDSS filters (*ugriz*; Fukugita et al. 1996), S82 was the target for recurring observations to detect variable objects and to obtain deep optical photometry. When the images are stacked, S82 has an optical i-band magnitude limit of $i \sim 24.1$ (Jiang et al., 2014), which is significantly deeper than the rest of the SDSS survey.

Spectroscopically-confirmed quasar data come from the composite quasar catalog of Richards et al. (2015). This catalog is a compilation of spectroscopic quasars from large surveys such as SDSS (York et al. 2000, Eisenstein et al. 2011) and the 2QZ project (Croom et al., 2004) as well as from smaller surveys such as Hectospec (Fabricant et al., 2005). In total, they compiled ∼2 million quasars and quasar candidates (including ∼437,000 spectroscopically confirmed quasars) which span a large range in both redshift and i-magnitude. The catalog encompasses faint, high-z quasars from BOSS (Pâris et al., 2014), which are key to defining the quasar color space used to classify the photometric objects.

Richards et al. (2015) also matched their catalog to infrared catalogs such as AllWISE[1] and various overlapping *Spitzer* surveys to investigate the mid-infrared colors of these known quasars in the full SDSS field. Mid-infrared color-color diagrams have been particularly useful in quasar classification as shown in Lacy et al. (2004), Stern et al. (2005), and Donley et al. (2012), among others. The addition of this mid-infrared data in classification allows for higher number densities of detected quasars, particularly at high-z ($z \geq 2.9$).

The new infrared SpIES and SHELA surveys provide a much larger area where deeper infrared

[1] `http://wise2.ipac.caltech.edu/docs/release/allwise/`

data overlaps the optical, providing the necessary information to classify objects as type-1 quasars; the challenge becomes selecting a clean sample of quasar candidates. However, using the machine-learning techniques demonstrated in Richards et al. (2015), selection of high-z quasar candidates has become much more complete. To generate a final catalog of high-z quasars, we must first assemble a complete sample of all detected objects (i.e., photometric and spectroscopic) to form the *test* set. This test set is then reduced to a subset containing the known (spectroscopically confirmed) high-z quasars used to define the color spaces that train the algorithms along with a fraction of the unknown (photometric) objects (the *training* set). Test objects are then fit using the trained algorithm and are assigned a classification. Presented in Table 3.2 are the demographics of the test and the training sets used in this study, as well as the final selected type-1 quasar candidates.

3.2.2 Test and Training Sets

In this study, the set of objects to be classified (the test set) and the objects used to train the algorithms (the training set) were constructed in much the same way as in Richards et al. (2015). The full test set was built using matched optical+MIR photometric data spanning the full SDSS footprint, where *WISE* photometry (converted to *Spitzer* magnitudes) was used when *Spitzer* data did not exist. Furthermore, to be considered for classification, these objects were required to be SDSS sources with $m_{AB} > 15$ in all optical bands (to remove contamination due to saturation) and to have the 'good' SDSS flags as described in Richards et al. (2015). The full test set contains ~ 50 million objects spanning the full SDSS footprint and includes both spectroscopic and photometric quasars. For this study we further restricted our final test set to objects *only* in S82 since we were particularly interested in candidates where deep *Spitzer* data exists from SpIES and SHELA. After this cut, the final S82 test set is comprised of ~ 2 million objects with optical+MIR color information.

Scranton et al. (2002) demonstrated that SDSS star-galaxy separation is relatively clean to $r \sim$21. The Stripe 82 catalogs are *catalog* co-adds, so the deeper data does not yield improved star–galaxy separation without further work. As our targets are typically $r \sim$22, in Section 3.2.5 and Appendix 3.7.1 we describe the tests intended to exclude low-redshift galaxies acting as interlopers in our

sample.

The quasar training set is constructed by first matching the full test set to the high-z quasars in the Richards et al. (2015) composite catalog. In total, there are 22,737 high-z ($z \geq 2.9$) matches between these two sets which we use to train our algorithms, the majority of which come from SDSS (DR7, DR10, and DR12). We also include 12 high-z ($z \geq 3.7$) spectroscopic quasars from VVDS and McGreer et al. (2013) which were confirmed after the composite catalog was generated. To ensure that the training objects are not confused with other low-z sources, we queried the NASA/IPAC Extragalactic Database[2] (NED) to check the redshifts. We performed a follow up visual inspection of the spectra for the objects which NED reported to be low-z, and removed four objects that had non-quasar spectra. In all, there are 22,745 quasars with $z \geq 2.9$ in the quasar training set to train our machine-learning algorithms.

Additionally, we add to the training set non-quasar sources ('stars'), which do not have spectroscopic information, randomly selected from the full test set. As described in Richards et al. (2015), the 'stars' in the training set can also include previously unclassified quasars, stellar sources, and compact galaxies. The additional 'star' information is important in the classification because it defines the color space boundaries around the high-z quasars in the machine-learning algorithms. The full training set is comprised of ∼700,000 'stars' and confirmed quasars in the SDSS footprint that are as faint as $i \sim 23$ and observed to $z \sim 6$. In this investigation, we split the training set into two redshift ranges; a lower-z ($2.9 \leq z < 3.5$) and a higher-z ($3.5 \leq z \leq 5.2$) range for selection. The colors of the higher-z objects are much more distinct from low-z 'stars' compared to the objects in the lower-z range, thus the selection is much more efficient in the higher-z range. Figure 3.2 depicts the colors of extended and point sources in the training set, and highlights the colors of the known high-z quasars in each color space used to classify the test objects. We also provide the demographics for both the testing and training sets in Table 3.2.

[2] http://ned.ipac.caltech.edu/forms/nnd.html

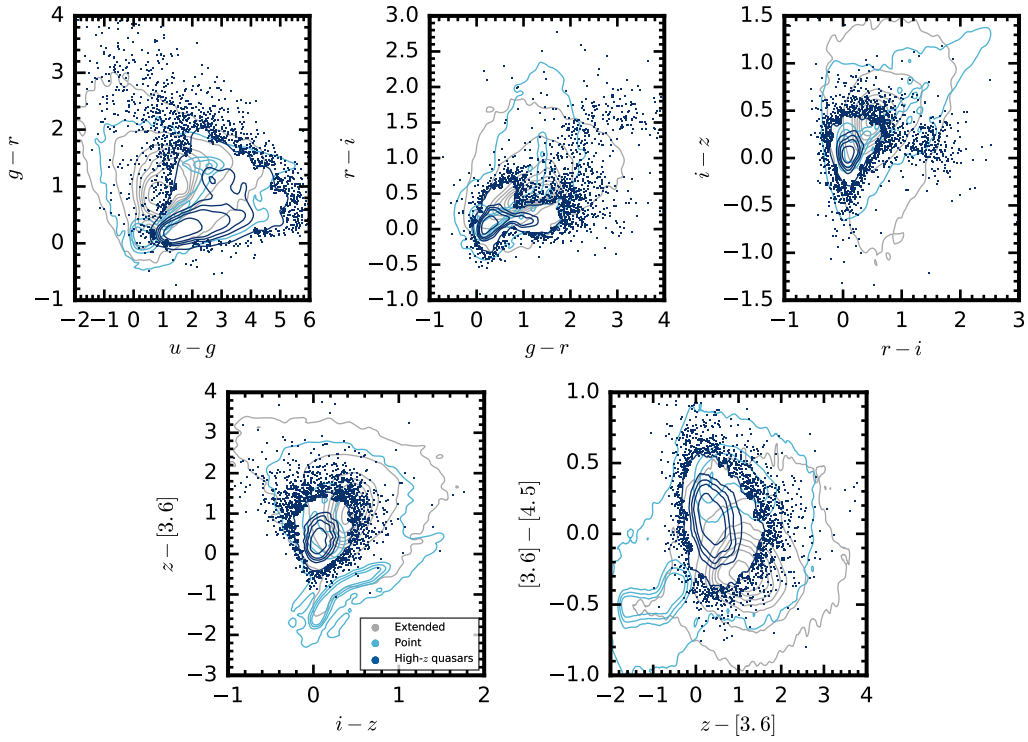

Figure 3.2 Optical and infrared colors (computed from AB magnitudes) of the training set objects used to define the classification color spaces. Extended objects in the training set are outlined by the gray contours and point sources are depicted in light blue. The dark blue contours highlight the region where known, high-z ($2.9 \leq z \leq 5.2$) quasars reside. The overlap of the extended sources and the high-z quasars opens the possibility that we classify, e.g., low-z galaxies as quasars in our algorithms. To remedy this particular situation, we define a metric to identify point sources which eliminates extended object (galaxy) contaminants, and by visual inspection (see Appendix 3.7.1).

3.2.3 Classification Algorithms

The colors of confirmed high-z quasars in the training set (shown in Figure 3.2) are used to teach the machine-learning algorithms where high-z quasars lie in multi-dimensional color space. Colors of the photometric objects in the S82 test set are then input into the trained algorithms to classify them as high-z quasars. For this analysis, we utilize three classification algorithms; *Random-Forest Classification* (RF), *Support-Vector Classification* (SVC), and *Bootstrap Aggregation (Bagging) on K-Nearest Neighbors* (KNN), which we define below. All of these algorithms are openly available in the Scikit-Learn[3] Python package used in this study.

The RF classifier[4] creates a set of N random decision trees, which split the training quasars by

[3]http://scikit-learn.org/stable/
[4]http://scikit-learn.org/stable/modules/ensemble.html#forest

Table 3.2. Training and Test Sets

Data Set	N_{Obj}	z-range	$i$mag
Full Training	$\sim 7.00 \times 10^5$	$0 \leq z \leq 6$	$15 \leq i \leq 24$
High-z quasars	$\sim 2.27 \times 10^4$	$2.9 \leq z \leq 6$	$16 \leq i \leq 23$
Full Test	$\sim 5 \times 10^7$	–	$16 \leq i \leq 23$
S82 Test	$\sim 2 \times 10^6$	–	$16 \leq i \leq 24$
Candidates[a]	1378	$2.9 \leq z^b \leq 5.1$	$18 \leq i \leq 23$

Note. — The training and testing set demographics. The training set is a combination of spectroscopic objects and photometric objects in the full test set. The training quasars is the compilation of spectroscopic quasars used to train the algorithms that are used to classify the S82 test objects. Combining the photometric and spectroscopic quasars, there are 1378 high-z quasars with which to compute the correlation function.
[a] Contains both spectroscopic and photometric quasars
[b]Photometric redshifts

their colors into different branches, with each branch returning a classification (in this case high-z or not). The colors of the test objects are then subject to the splitting that each tree has created, and each of the trees assign a classification based on the conditions that the test objects satisfy. The mode result of all of the trees is used as the final classification for each of the test objects.

We also employ the SVC algorithm[5], which defines an optimal hyperplane that separates two populations of objects by the largest margin. In this case, the training set objects create the six-dimensional color space, and the hyperplane is defined by the plane that maximally separates the known high-z quasars from the 'stars' in the training set. Classification of the test objects is based on the side of the hyperplane they lie in this multi-dimensional color space.

Finally, we use "Bagging" with a KNN algorithm[6], where Bagging is the process of splitting the training set into N different subsets of randomly chosen training objects (with replacement). Each of those subsets is used to train the machine-learning algorithm (KNN in this case), resulting in N trained KNN algorithms. The KNN algorithm assembles the training set color information

[5]http://scikit-learn.org/stable/modules/svm.html#svm
[6]http://scikit-learn.org/stable/modules/neighbors.html#classification

and classifies the test data by analyzing the closest 'k' training objects in color space. Similar to a majority rule, the test object is classified based on the type of the closest 'k' training object (in this case, high-z quasar or not). This analysis is done in all of the Bagging subsets, and the mean result from all of the bags is chosen as the final classification.

To measure the effectiveness of each algorithm, we compute the two key selection parameters: efficiency and completeness. The efficiency of an algorithm relates the number of objects that it classifies correctly to the total number of objects it classifies, and can be used to estimate the contamination of the classified sample by taking the difference from unity. Completeness is a measure of how many quasars are properly classified compared to the total number of known quasars in the data set.

Estimation of the completeness and efficiency of our algorithms requires the full training set to be split into two subsets for cross-validation (CV); a subset with 75% of the data to be used as a CV 'training set', and a subset with 25% of the data to be used as CV 'test' objects. These sets are input into the classification algorithms discussed above. Since the CV test objects contain known quasars, completeness and efficiency can be calculated using the classification results of the known quasars from the CV test set. Ideally, both completeness and efficiency should be maximized to recover all of the high-z quasars, and only the high-z quasars. Practically, however, quasar colors can overlap with stars and low-z galaxies, so contamination and missed classifications are inevitable.

We compare our algorithms to the kernel density estimation (KDE) used in Richards et al. (2015). This study classified photometric objects in the SDSS footprint using optical data along with infrared data from *WISE*. The KDE method used in Richards et al. (2015) first defined a color 'bandwidth' for each class of object (quasar or non-quasar) which acts to smooth the color distributions, and a Bayesian stellar prior which defines the percentage of objects in the test sets thought to be 'stars' (i.e., non-quasars). A probability density function (PDF) is then defined in color space for a class of object, and the likelihood that a test object with certain photometric colors belongs to a class is computed using the bandwidth and a kernel function. The posterior probability that an object is a quasar given its color is computed by applying Bayes' theorem using the defined

Table 3.3. Estimated Completeness and Efficiency

Algorithm	Completeness %	Efficiency %	Contamination %
RF	83/78/80	43/93/86	57/7/14
SVC	82/79/79	40/95/86	60/5/12
Bagging KNN	83/80/80	85/95/88	15/5/12
KDE	−/78/−	−/97/−	−/3/−

Note. — Estimated completeness, efficiency and contamination measured for the three algorithms used in this study compared to the KDE method used in Richards et al. (2015). The first three rows report our algorithms when selecting in a lower redshift range ($2.9 \leq z < 3.5$; left), a higher redshift range ($3.5 \leq z \leq 5.2$; center), and when selecting in a broader redshift range ($2.9 \leq z \leq 5.2$; right). The values in the center are used to compare to Richards et al. (2015). These values are estimates since the actual test set probes slightly fainter than the validation set.

priors and likelihoods for each test object (see Richards et al. 2009 for details). This study classified objects over a wide range of redshifts, however we will compare the performance of our algorithms to their highest redshift classification ($3.5 \leq z \leq 5$).

In our investigation, we split the classification of quasar candidates into a lower-z ($2.9 \leq z < 3.5$) and a higher-z ($z \geq 3.5$) bin. We found that the classification algorithms performed better at higher-z compared to lower-z as reported in Table 3.3. This mainly because quasars begin to drop out of the SDSS u-band filter at $z \sim 3.5$, which significantly alters the $u - g$ color space and helps the machine-learning algorithms efficiently select these objects. The colors of $z \sim 3$ quasars are very similar to those at $z \sim 2.2$, therefore the algorithms tend to confuse low-z quasars with higher redshift quasars. Through cross validation, we found that the "Bagging" algorithm performed the best at lower-z and all three perform equally well at higher-z as reported in Table 3.3. More details are presented in Section 3.2.5 where we describe our final candidate selection.

3.2.4 Photometric Redshifts

Photometric redshifts of our candidates were estimated with Nadaraya-Watson (NW) kernel regression[7]. NW is a natural extension of more familiar regression techniques. Linear regression fits a line to 2-D data. Polynomial regression instead fits a higher order curve. Basis function regression (of which polynomial regression is an example) uses a pre-determined "basis" function to fit the data. NW is just basis function regression using a Gaussian kernel (Ivezić et al., 2014).

The NW algorithm defines the multi-dimensional color space of the training objects with spectroscopic redshifts, then builds a kernel matrix, K, which measures the pairwise distance between the colors of the test objects and the colors of the training objects, where K is the Gaussian kernel:

$$K = \exp\left(\frac{1}{2\sigma^2}\|d_{\rm test} - d_{\rm train}\|^2\right). \tag{3.3}$$

Here, $\|d_{\rm test} - d_{\rm train}\|$ is the Euclidean distance between the colors of the test objects and the color of the training objects, and σ is the bandwidth of the kernel ($\sigma = 0.05$ produced the best self-validation results in this study). From Equation 3.3, if a test object is close to a training object (i.e., if the 6D colors are very similar), the kernel approaches 1; however, the further the colors are from each other, the smaller the Gaussian kernel becomes. Therefore, the kernel matrix is used as weights in the estimate of the photometric redshift, defined by:

$$z_{\rm phot} = \frac{\sum_i K_i \cdot z_{{\rm spec},i}}{\sum_i K_i}, \tag{3.4}$$

where the kernel element in K is multiplied by the spectroscopic redshift corresponding to the training quasar input into Equation 3.3. The final photometric redshift result for a candidate object is then the weighted sum over all the spectroscopic redshifts of the training objects.

To test the effectiveness of this method, we calculate the photometric redshifts of the spectroscopic quasars on S82 over all redshift ranges using the same training set we use for the candidates. Additionally, we split the quasars into a bright and faint subset, where we differentiate between

[7] http://www.astroml.org/modules/generated/astroML.linear_model.NadarayaWatson.html

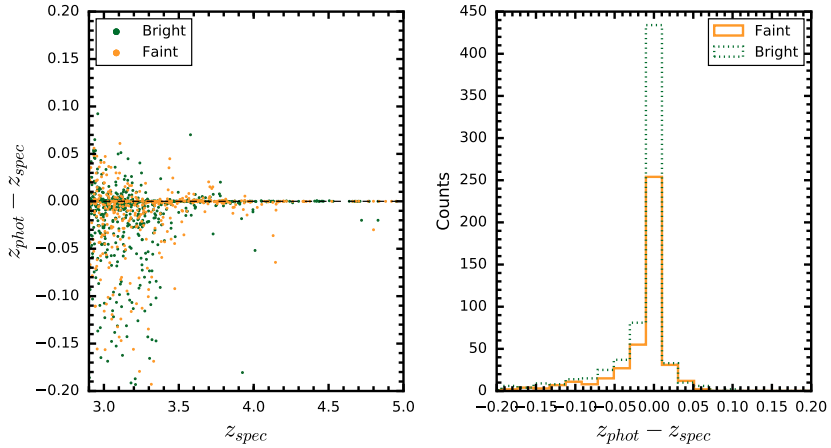

Figure 3.3 Left: Comparison of known high-z quasar spectroscopic redshifts with the estimated photometric redshift using Nadaraya-Watson regression. The known quasars are split into bright ($i \leq 20.2$; green points) and faint ($i > 20.2$; orange points) bins to test the effectiveness of this algorithm for quasars of different brightness. The black dashed line depicts $z_{phot} - z_{spec} = 0$. Right: Difference between the photometric and spectroscopic redshifts for the bright and faint quasars. Approximately $\sim 93\%$ of the high-z quasars are constrained to $|\delta z| \leq 0.1$ in both bins.

bright and faint at i=20.2. The results in Figure 3.3 show that there is a tight correlation between the spectroscopic redshift of the quasar and its estimated photometric redshift for both subsets. In both cases $\sim 93\%$ of the photometric redshifts differ from the spectroscopic redshifts by no more than, $|\delta z| \leq 0.1$. These results are similar to the findings in Richards et al. (2015) for their highest redshift bin, who used an empirical method outlined in Richards et al. (2001) and Weinstein et al. (2004).

Using the NW regression algorithm, each candidate quasar selected with the aforementioned algorithms was assigned a photometric redshift. A comparison of the candidate redshifts to the spectroscopic redshifts is displayed in Figure 3.4. With candidates selected, and their photometric redshifts computed, we now create a final sample of candidates with which to compute the correlation function.

3.2.5 Clustering Sample

Although classification was performed on all of the S82 test objects, and photometric redshifts were computed for all candidates that were selected, we further restricted the data set to create the cleanest sample of faint, high-z quasars with which to compute the 2PCF. First, to retain the faintest

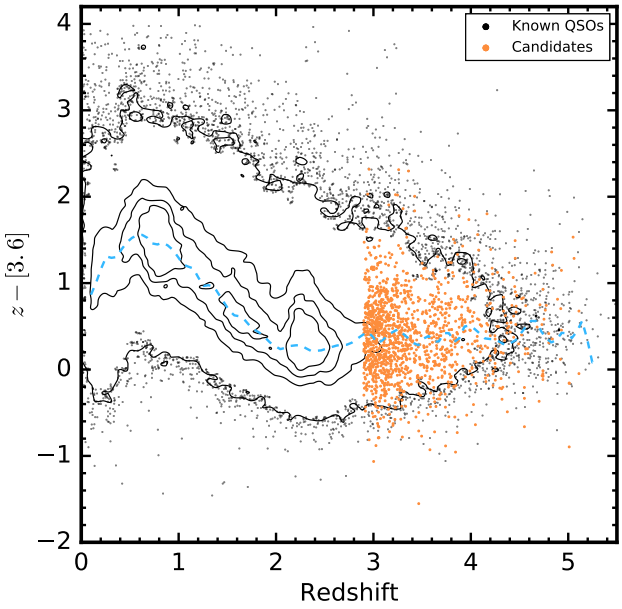

Figure 3.4 Example color-redshift diagram of the spectroscopic training data (black) and the photometric redshifts of the candidates (orange). The photometric redshifts estimated using the NW algorithm share the same color space as the spectroscopic sample on which the algorithm was trained ($2.9 \leq z \leq 5.1$). The blue dashed curve indicates the modal color as a function of redshift of the known quasars.

objects with the deepest photometry, we required that the objects lay within the SpIES/SHELA footprint, where the deep MIR data exists, *and* that they were sufficiently far away from bright stellar sources which contaminate the photometry (see Timlin et al. 2016 for more details). Additionally, candidates were required to have photometric redshifts in the range $2.9 \leq z \leq 5.1$, enabling us to compare our results with Shen et al. (2007): the most recent wide-area spectroscopic study of quasar clustering at redshifts as high as $z \sim 4$.

To ameliorate potential sources of contamination, yet to select as many true high-z quasars as possible, we combined the results of each of the selection algorithms (see Table 3.3). At low-z ($2.9 \leq z < 3.4$), we chose to *only* employ the "Bagging" classifier because of its high efficiency. While including the results from the other two classifiers would have made our sample more complete, it also would have added a large amount of contamination. At high-z ($3.4 \leq z \leq 5.2$), however, we combined the selection results of the three algorithms since they all have low contamination as shown in Table 3.3.

Despite combining the classification results in this manner, the sample still contained contami-

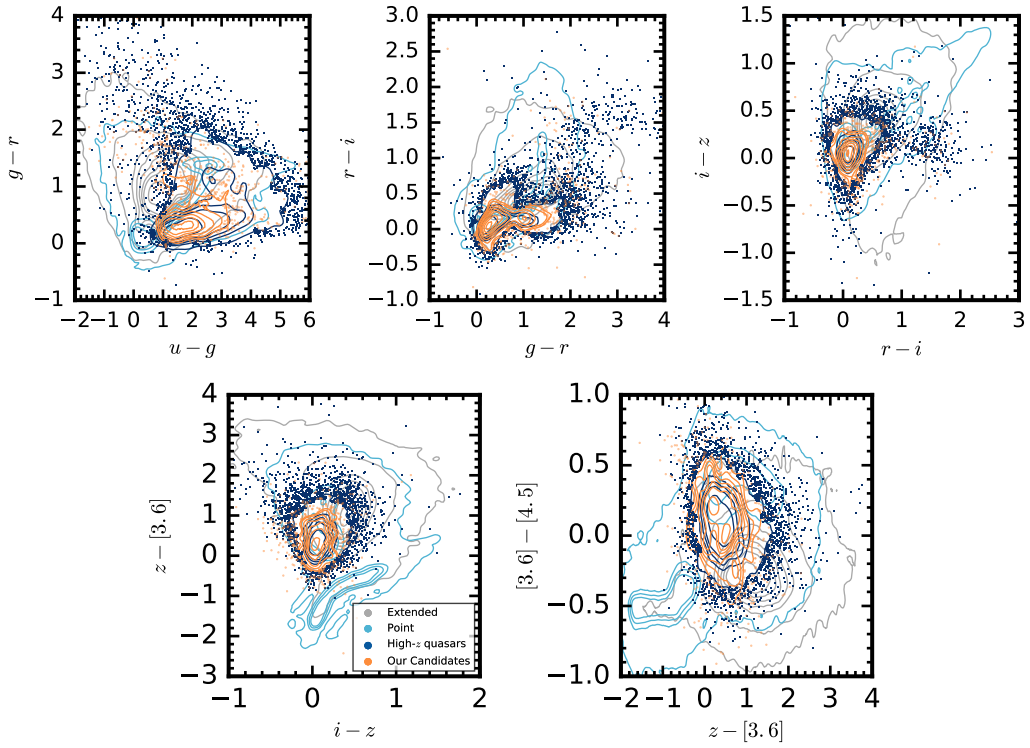

Figure 3.5 Optical and infrared colors of the selected quasars (orange contours). The other contour colors are the same as in Figure 3.2. These panels demonstrate that the location of the candidates in color space overlap with the colors on which they were trained (dark blue contours).

nation from low-z galaxies. To eliminate the obvious galaxies, we restricted our data to point-like sources only. We generated our own metric for high-z quasar point sources by taking the difference between the `PSFMAG`[8] and `cMODELMAG`[8] (δ_{mag}) in the SDSS DR10 i-band. A difference of $\delta_{mag} \leq 0.145$ is used in the SDSS catalogs to label an object as a point source. We found that the known quasars in our lower-z range ($2.9 \leq z < 3.4$) had $\delta_{mag} = 0.2$, whereas the known higher-z ($z \geq 3.4$) had $\delta_{mag} = 0.15$. We apply this morphology cut to the selected objects in the appropriate redshift range after the selection had been performed. This cut eliminated a significant fraction of extended sources, which we consider to be contaminants in our sample (confirmed using visual inspection; see Appendix 3.7.1).

Another source of contamination that we account for is high Galactic extinction objects which can cause low-z objects to be mistaken for high-z quasars (Myers et al., 2006). Removal of these

[8] `http://www.sdss3.org/dr10/algorithms/magnitudes.php`

highly extincted objects is particularly important in this study since the eastern edge of the SpIES field overlaps with the Galactic plane ($330° \leq \alpha_{J2000} \leq 344.4°$). To remove the contamination due to these objects, we elect to cut out this region from our final analysis (see Appendix 3.7.1 for more details). While this process eliminates some area over which we can perform the clustering analysis, it also removes contaminants that are confused for high-z quasars in the machine-learning algorithms (despite the extinction-corrected magnitudes).

Differences in the angular mask of the data and the randoms can also affect our clustering measurement. The edges of the SHELA field are not uniformly covered in the mask, requiring that we cut in declination ($-1.2° \leq \delta_{J2000} \leq 1.2°$) to ensure that the densities of the data and randoms were approximately the same across the field. After cutting out the extinction region and these under-dense regions, our final footprint covers 102 deg^2 on Stripe 82.

Finally, every candidate object (before and after the morphology cut) was visually inspected using the stacked g, r, z images from the Dark Energy Camera Legacy Survey (DECaLS[9]) image cutout tool[10]. DECaLS images to similar depths as the SDSS S82 coadded catalog ($r = 23.4$ compared to $r = 24.6$ on S82), but uses the Dark Energy Camera (DECam[11]), which has a finer resolution than SDSS ($0.26''$ compared to $0.39''$ per pixel). This added resolution enabled us to visually eliminate a small number of obvious low-redshift galaxies which share color spaces with high-z quasars (see examples in Appendix 3.7.1).

After the cuts and visual inspection, 1378 objects remained as high-z quasars (see Table 3.4). Of these, 726 are spectroscopically confirmed from the Richards et al. (2015) comprehensive catalog and we select 652 new high-z quasar candidates with which we can measure the 2PCF. None of the quasars or candidates used in this study were used in the Shen et al. (2007) study. The colors of our selected quasars are presented in Figure 3.5 and compared to the colors of the training objects. The majority of the selected quasars share the same color space as the high-z quasars whose colors were used to train the algorithms, but as our candidates delve fainter than the majority of the training objects, there is some scatter in their colors. While there is stellar contamination in the sample

[9] `legacysurvey.org`
[10] `https://github.com/yymao/decals-image-list-tool`
[11] `http://www.ctio.noao.edu/noao/content/DECam-User-Guide`

(which we will model in Section 3.4), some of the scatter in the colors could be due to contamination from objects such as compact galaxies, which are more difficult to identify from colors alone.

Using the redshifts and the i-band apparent magnitudes, we compute the absolute magnitude of these quasar candidates, and compare them to the spectroscopic sample from Shen et al. (2007) (renormalized to $z = 2$ after K-correcting using the model in Richards et al. 2006) in the top panel of Figure 3.6. The majority of the photometric candidates are fainter than the Shen et al. (2007) quasars and are fainter than $i = 20.2$ (shown in the bottom panel of Figure 3.6), *which is necessary to break the degeneracy in the bias as a function of redshift.* This investigation contains a small number of objects brighter than $i = 20.2$ compared to Shen et al. (2007) because it covers a smaller area (~ 100 deg^2 vs. ~ 4000 deg^2, respectively).

3.3 Clustering

3.3.1 Two-Point Correlation Function

Spatial clustering of a population of objects is quantified using the 2PCF, which is the joint probability of finding an object in two volume elements, dV_1 and dV_2, at some separation $\mathbf{r_{12}}$ (Peebles, 1980). This quantity can be expressed as:

$$dP = n^2[1 + \xi(\mathbf{r_{12}})]\ dV_1\ dV_2 \tag{3.5}$$

where n is the mean number density and $\xi(\mathbf{r_{12}})$ is the correlation function. In this equation, if the 2PCF is zero, the probability shows no excess compared to a Gaussian random distribution. We can derive this statistic for a distribution of objects in a density field, ρ, where the probability of finding an object in that field is $dP = \langle\rho(r)\rangle\ dV$ (Peebles, 1980). The probability of finding a pair of objects in two density fields ρ_1, ρ_2 separated by a distance r is:

$$dP = \langle\rho_1(r)\rangle\langle\rho_2(r)\rangle\ dV_1\ dV_2 \tag{3.6}$$

The density in an expanding Universe is modeled with a linear perturbation $\rho(r) = \bar{\rho}[1 + \delta(r)]$,

Table 3.4. Quasar Candidate Table

α_{J2000} (degrees)	δ_{J2000} (degrees)	Au	ug	gr	ri	iz	$zs1$	$s1s2$	$i$mag (AB; asinh)	$z$spec	$z$photNW	$z$best
31.92786	-0.04275	0.15	2.17	0.933	0.277	0.203	0.811	0.329	21.00	-999.0	3.89	3.89
12.30498	0.69156	0.14	0.23	2.04	0.113	0.030	0.631	-0.543	22.10	-999.0	3.90	3.90
27.56491	0.76547	0.14	1.84	1.50	0.205	0.283	0.600	0.143	20.76	3.90	3.91	3.90

Note. — List of all candidates selected by the three algorithms. Along with positional information, we record the i-band AB magnitude, its u-band extinction parameters, and the optical/infrared color of each object. We also report the spectroscopic redshift if the quasar is a confirmed object (a value -999.0 indicates that there is no spectroscopic redshift). The photometric redshift estimate from the Nadarya-Watson regression algorithm is recorded in the next column followed by the 'best' redshift estimate (records the spectroscopic redshift instead of the photometric estimate, when available). The full version of this catalog can be found at `https://github.com/JDTimlin/QSO_Clustering/tree/master/highz_clustering/clustering/Data_Sets`.

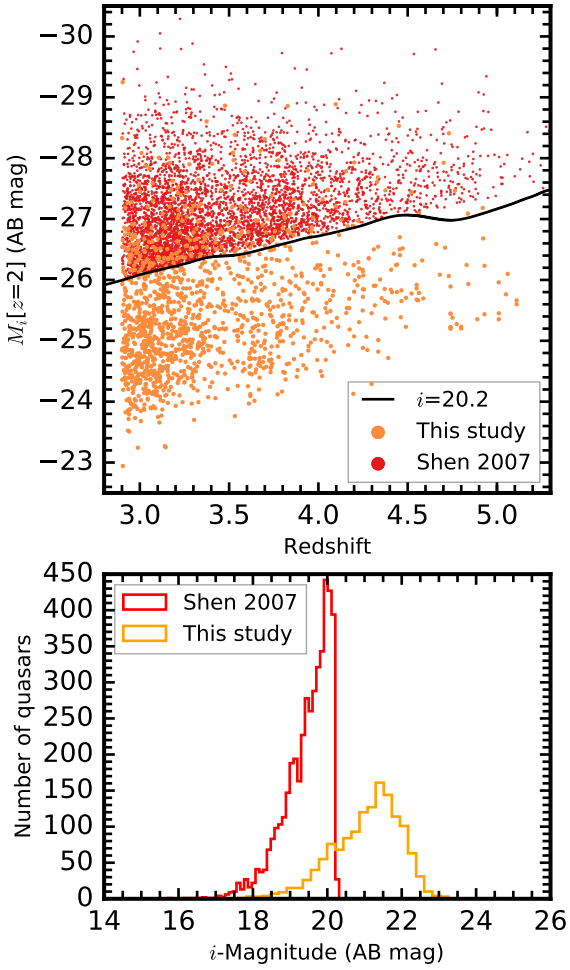

Figure 3.6 Top: Absolute i-magnitude of the quasar candidates (orange points) compared to the spectroscopically-confirmed quasars from Shen et al. (2007) (red points). The solid black curve depicts constant i-magnitude ($i = 20.2$), where the Shen et al. (2007) objects are brighter than this magnitude and the photometric candidates are fainter. The i-magnitudes were corrected for reddening to $z = 2$ using the model from Richards et al. (2006). Bottom: Distribution of i-magnitudes for our candidates (orange) compared to the Shen et al. (2007) candidates (red). We have far fewer bright objects ($i \leq 20.2$) because our survey area is much smaller.

so Equation (3.6) becomes:

$$dP = \langle \bar{\rho}[1 + \delta_1(r')] \rangle \langle \bar{\rho}[1 + \delta_2(r)] \rangle \; dV_1 \; dV_2 \tag{3.7}$$

$$dP = \bar{\rho}^2[1 + \langle \delta_1(r')\delta_2(r) \rangle] \; dV_1 \; dV_2 \tag{3.8}$$

Comparing with Equation (3.5) we see that the correlation function is the ensemble average of the perturbations, $\xi(r_{12}) = \langle \delta_1(r')\delta_2(r) \rangle$. The density field can also be expressed in Fourier space (Bonometto et al., 2002):

$$\delta(r) = \frac{1}{(2\pi)^3} \int \delta(k) e^{-ikr} dk. \tag{3.9}$$

Taking the Fourier transform of the correlation function yields:

$$\langle \delta_1(r')\delta_2(r) \rangle = \frac{1}{(2\pi)^3} \int \langle \delta(k)\delta^*(k) \rangle e^{-ikr} dk \tag{3.10}$$

where the ensemble average of the density modes, $\langle \delta(k)\delta^*(k) \rangle$, is the definition of the power spectrum, $P(k)$. The correlation function is, therefore, the Fourier transform of the power spectrum. We relate our clustering results to the theoretical clustering of DM, which will be obtained through calculation of the DM power spectrum. In this paper, we compute the angular projected correlation function, $\omega(\theta)$, which is a projection from three dimensional (3-D) volume space into two dimensional (2-D) angular space.

3.3.2 Estimating the Correlation Function

To estimate the correlation function, one needs to compare the data set to a set of randomly distributed points. To compute the correlation function we use the estimator from Landy & Szalay (1993):

$$\omega(\theta) = \frac{\langle DD \rangle - 2\langle DR \rangle + \langle RR \rangle}{\langle RR \rangle} \tag{3.11}$$

where $\langle DD \rangle, \langle DR \rangle$, and $\langle RR \rangle$ are the data-data, data-random, random-random pair counts within an angular separation of θ (to measure the three dimensional correlation function, $\xi(s)$ one simply counts pairs within a 3-D comoving separation distances). The pair counts are normalized by the ratio of the number of objects in the data and random sets. To reduce the shot noise in the measurement, we use $\sim$100 times the number of points as the data catalog. The normalization of the pair counts reconciles that there are more random points to match than data.

The random data must lie on an identical angular mask as the data. To generate the random catalog for our candidates, we first construct the angular mask using the MANGLE[12] software (Swanson et al., 2008). This package allows a user to combine polygons from telescope observations to create a continuous mask, with accurate boundaries and holes, on the surface of a sphere. We combine the fields from the SpIES and SHELA surveys (see Figure 3.1), and remove circular regions of varying radii around bright stars from the 2MASS Point Source Catalog as outlined in Timlin et al. (2016). Objects in these regions were excluded from the selection of quasars because they are contaminated by the excess flux from the bright star, so we mask them using MANGLE. Random positions are chosen across the full field, avoiding masked areas, to form the random mask which is used in the LS estimator in Equation 3.11. Figure 3.7 compares the data to the random catalogs within a sample of the field created in MANGLE.

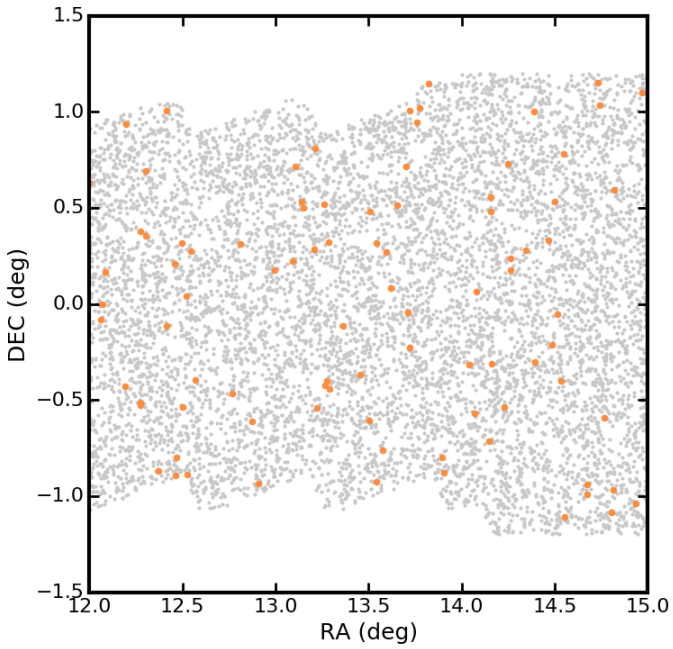

Figure 3.7 High-z quasar data set (orange) and the random mask (grey) used to perform the clustering analysis. The holes in the mask are cutouts of bright stars in the SpIES and SHELA field where the radius of the hole corresponds to the brightness of the star (see Timlin et al. 2016 for more details). The holes and corners in this mask identify locations where candidates cannot be selected; to eliminate bias, we mask these regions using MANGLE. Additionally, we exclude objects in the declination range $-1.2^\circ \leq \delta_{J2000} \leq 1.2^\circ$ due to coverage issues in the edges of the SHELA fields.

[12] `http://space.mit.edu/~molly/mangle/`

3.3.3 Measuring bias

The linear bias in Equation 3.1 is used as a measure of the clustering strength of the population of quasars and has been related to many physical parameters of quasars as well as their DM environments.

Estimating the bias, however, requires that we relate the projected correlation function to the three dimensional power spectrum. To perform this task, we use Limber's approximation which projects the three-dimensional correlation function to two dimensions (Limber, 1953) for objects with small separations ($\theta \ll 1$ rad; Simon 2007). Projecting the correlation function requires that we integrate the three dimensional correlation function along the line of sight of two objects,

$$\omega(\theta) = \iint \xi(r_1, r_2)|r_1|^2|r_2|^2\phi(r_1)\phi(r_2)dr_1dr_2 \tag{3.12}$$

where $|r_1|, |r_2|$ are the magnitudes of the two distance vectors and $\phi(r)$ is a radial selection function. The selection function acts as a probability distribution where the integral of $r^2\phi(r)\ dr$ is normalized to unity (Brewer, 2008). Shifting the coordinate system to one where the unit vectors are along the line of sight, $u = r_1 - r_2$, and across the line of sight, $r = \frac{1}{2}(r_1 + r_2)$, Equation (3.12) can be rewritten as:

$$\omega(\theta) = \int_0^\infty r^4\phi(r)^2 dr \int_0^\infty \xi(\sqrt{u^2 + r^2\theta^2})\ du \tag{3.13}$$

where, for small u, $r_1 \approx r_2$ and for small angles, $\cos(\theta) \approx 1 - \frac{\theta^2}{2}$ (see Peebles 1980, Brewer 2008 for more details). Equation (3.13) is the functional form of Limber's Equation to project the 3-D correlation function into two dimensions.

We transform Limber's equation into familiar cosmological parameters. For instance, the observed number of objects in radial shells can be described in terms of the redshift distribution of a sample of objects by:

$$\phi(r)r^2\ dr = \frac{dN}{dz}\ dz \tag{3.14}$$

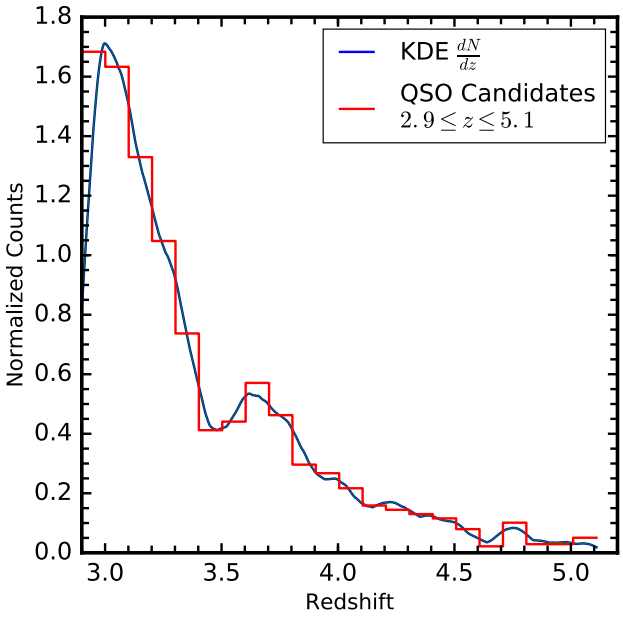

Figure 3.8 Photometric redshift distribution of the quasar candidates. The blue curve was determined from kernel density estimation using the 'epanechnikov' kernel with a bandwidth = 0.1. This curve is used in Limber's equation to estimate the bias by comparing the projected correlation function to the three dimensional dark-matter power spectrum. The red histogram depicts the distribution of the photometric redshifts in the data set.

Solving for ϕ and incorporating into Limber's equation, we get:

$$\omega(\theta) = \int_0^\infty \left(\frac{dN}{dz}\right)^2 \left(\frac{dz}{dr}\right) dz \int_0^\infty \xi(\sqrt{u^2 + r^2\theta^2})\, du \tag{3.15}$$

with the variable r defined as the comoving distance χ (Brewer, 2008). Assuming a flat Universe:

$$dr = d\chi = \frac{c}{H_0 E_z}\, dz \tag{3.16}$$

where $E_z = [\Omega_M(1+z)^3 + \Omega_\Lambda]^{\frac{1}{2}}$. Thus Equation (3.13) transforms to:

$$\omega(\theta) = \int_0^\infty \left(\frac{dN}{dz}\right)^2 \frac{H_0 E_z}{c}\, dz \int_0^\infty \xi(\sqrt{u^2 + r^2\theta^2})\, du \tag{3.17}$$

Using the fact that the correlation function is the Fourier transform of the power spectrum, and since we know that u is small, we can employ the Hankel transformation on the second integral to

obtain Limber's equation in terms of the quasar power spectrum:

$$\omega(\theta) = \frac{H_0 \pi}{c} \iint \left(\frac{dN}{dz}\right)^2 E_z \frac{\Delta_Q^2(k,z)}{k^2} J_0(k\theta\chi(z))\ dk\ dz \tag{3.18}$$

where Δ_Q^2 is the dimensionless quasar power spectrum ($\Delta^2 = \frac{k^3 P(k)}{2\pi^2}$) and J_0 is the zeroth order Bessel Function of the first kind (Bonometto et al. 2002; Myers et al. 2007; Brewer 2008). This formula relates the 3-D quasar power spectrum to the 2-D correlation function.

Equation (3.1) can now be written in a similar fashion by replacing the correlation functions with the dimensionless power spectra of quasars and dark-matter, $\Delta_Q^2 = b^2 \Delta_{DM}^2$. We substitute this relation into Equation (3.18) which allows us to cast this equation as a function of bias directly,

$$\omega(\theta) = \frac{b^2 H_0 \pi}{c} \iint \left(\frac{dN}{dz}\right)^2 E_z \frac{\Delta_{DM}^2(k,z)}{k^2} J_0(k\theta\chi(z))\ dk\ dz \tag{3.19}$$

where we assume that, for our samples of interest, bias does not evolve strongly with redshift or scale (e.g., Myers et al. 2007). Using Equation (3.19), we can fit a bias value using the measurement of the projected correlation function and the 3-D dimensionless dark-matter power spectrum.

To compute the dark-matter power spectrum, we use the Code for Anisotropies in the Microwave Background (CAMB[13]), which is a general cosmology package that creates a model cosmography. CAMB has the functionality to compute the dark-matter power spectrum including the nonlinear corrections from the halo model in Smith et al. (2003). Combining the dark-matter power spectrum (which is a function of wave-number, k, and redshift, z) with the redshift selection function for our candidates (blue curve in Figure 3.8), we Monte Carlo integrate Equation (3.19) and generate a theoretical model for the projected clustering of dark-matter. Finally, we fit the DM clustering model to the measurement from our sample and obtain a bias.

[13] http://camb.info

3.4 Results

3.4.1 Projected Clustering

The measured SpIES/SHELA angular projected 2PCF of the quasars in this sample is shown in Figure 3.9. We estimate the errors on these points using both the Poisson approximation (see Equation 3.27 in Appendix 3.7.4) along with the Jackknife resampling technique (Scranton et al. 2002; Myers et al. 2007; Ross et al. 2009; Eftekharzadeh et al. 2015), where a subset of the data (and the randoms) is removed from the full set, and the clustering analysis is performed on the remaining objects. In this investigation, the data sample was split into ten declination slices, resulting in ten separate clustering measurements, each excluding a different region. Using the ten jackknife clustering measurements and their RR pair counts, we compute the full covariance matrix by:

$$\begin{aligned} C_{ij} = \sum_{L} & \sqrt{\frac{RR_L(\theta_i)}{RR(\theta_i)}}[\omega_L(\theta_i) - \omega(\theta_i)] \\ & \times \sqrt{\frac{RR_L(\theta_j)}{RR(\theta_j)}}[\omega_L(\theta_j) - \omega(\theta_j)], \end{aligned} \tag{3.20}$$

where L denotes the *removal* of one of our 10 regions to form a jackknife sample comprising the other 9 regions, and θ_i, θ_j represent the clustering result at different separation values. The error bars on the orange points in Figure 3.9 show the standard deviations of the full measurement, computed by taking the square root of the main diagonal of the covariance matrix (Myers et al. 2007, Ross et al. 2009, Eftekharzadeh et al. 2015). We take Poisson errors to be the minimum error of the data, therefore we replace any Jackknife error with a value less than the Poisson estimate with the Poisson error value (see Appendix 3.7.4).

The orange curve in Figure 3.9, which is a fit of the DM clustering model to the measured clustering result, incorporates an estimate of stellar contamination in the sample. Following the method in Myers et al. (2006), stellar contamination is modeled using the measured correlation function of known stars in the field (with $g < 17.1$; Myers et al. 2006), as well as the efficiency, e, of the classification algorithms (e =0.86 in this study; see Table 3.3). With an efficiency of $e = 0.86$, we predict 14% contamination from stellar sources in our DM model fit. The correlation function

Table 3.5. Pair Counts Results

θ (arcmin)	DD	DR	RR	$\omega(\theta)$	$\sigma_{JK}(\theta)$	$\sigma_P(\theta)$
0.076	0	7	828	-0.754	0.5038	–
0.116	0	17	1980	-0.781	0.5074	–
0.175	0	37	4370	-0.756	0.2472	–
0.266	2	116	10432	0.756	2.7091	1.7556
0.403	2	268	23350	-0.459	0.8997	0.5406
0.611	8	542	53470	0.507	0.6561	0.7534
0.927	12	1162	120784	0.073	0.5107	0.4381
1.405	32	2652	273444	0.247	0.3623	0.3118
2.131	74	6022	619802	0.269	0.2385	0.2086
3.231	156	13782	1403064	0.158	0.1461	0.1312
4.899	324	30643	3191350	0.100	0.0680	0.0865
7.428	692	69474	7209140	0.034	0.0398	0.0556
11.262	1506	155010	16201178	0.015	0.0295	0.0370
17.075	3226	343452	36027696	-0.014	0.0239	0.0245
25.889	7104	747011	78725580	0.002	0.0116	0.0168
39.253	14932	1581774	166710784	-0.005	0.0088	0.0115
59.516	29674	3141776	330927082	-0.005	0.0058	0.0082
90.237	53028	5579277	583004272	-0.007	0.0067	0.0061
136.818	75010	7795784	815069184	0.006	0.0048	0.0052
207.443	100858	10584579	1113270342	0.002	0.0050	0.0045

Note. — Pair counts and correlation function measurements within increasing separations on the sky. Also recorded are the error estimates from the main diagonal of the covariance matrix (see Equation 3.20) estimated using jackknife resampling, as well as Poisson errors (see Equation 3.27). In this investigation, jackknife errors are replaced with Poisson errors where the ratio of jackknife to Poisson is less than unity (see Appendix 3.7.4). In this table, we report DD and RR as double counted pairs.

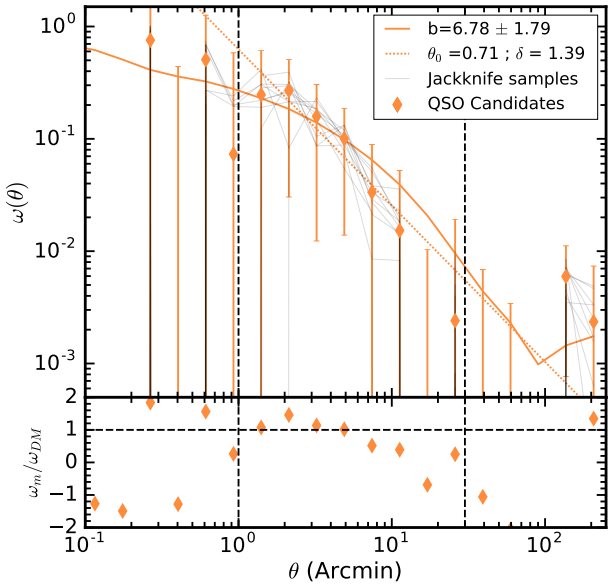

Figure 3.9 Clustering result from Table 3.5 of the 1378 high-redshift ($2.9 \leq z \leq 5.1$) photometric candidates (orange diamonds). Fitting the DM model to the data over the range $1'$ to $30'$ (black dashed vertical lines) produces a best fit bias of $b = 6.78 \pm 1.79$ (orange curve). This model accounts for excess power at large scales by incorporating stellar contamination into the model fit from Equation 3.21. The dotted line indicates the best fit power-law with θ_0=0.71 ± 0.546 arcmin and δ=1.39 ± 0.618. The lower panel shows the ratio between the measured points and the DM model. Error bars were computed using jackknife resampling, where the grey lines represent the correlation function results for each of the jackknife samples.

estimate becomes:

$$\omega(\theta) = e^2\omega_{QQ}(\theta) + (1 - e^2)\omega_{SS}(\theta) + \epsilon(\theta) \tag{3.21}$$

where $\omega_{QQ}(\theta)$ is the model result from Limber's equation, $\omega_{SS}(\theta)$ is the stellar correlation function in the field, and $\epsilon(\theta)$ is the cross correlation between quasars and stars (theoretically zero; Myers et al. 2007) which is insignificant in our study. Following Myers et al. (2006), we estimate the stellar correlation function by performing the clustering analysis of SDSS point sources which have bright g-band magnitudes ($16.9 < g < 17.1$). The stellar correlation function in the footprint of this survey is $\omega_{SS} \simeq 0.1$ at $30'$, slightly less than what Myers et al. (2006) found ($\omega_{SS} \simeq 0.25$) using an expanded version of the KDE-selected sample of Richards et al. (2004). We also fit a single power law to the data of the form:

$$\omega(\theta) = \left(\frac{\theta}{\theta_0}\right)^{-\delta}, \tag{3.22}$$

where θ_0 is the angular separation over which objects are correlated, and δ defines the degree of

clustering as a function of angular scale.

Using the measurement and errors of the 2PCF (see Table 3.5), and the DM model estimated using Limber's Equation, we can determine the bias that best relates the measurement and the theory. Similar to the fit in Myers et al. (2007), the bias was fit on scales with sufficient data-data pairs ($\theta \geq 1'$) and before the stellar correlation function dominates the quasar clustering signal ($\theta \leq 30'$). In principle, stellar contamination does not greatly change the correlation function at small scales (Myers et al., 2006), however, photometrically-selected samples inevitably contain some level of contamination, thus it is imperative that we incorporate an estimation of contamination in our model.

We fit the bias value, b, as well as the cross-correlation, ϵ, over the range of $1'$ to $30'$ (removing the negative value points) using Equation 3.21. The best-fit bias value is $b = 6.78 \pm 1.79$ and $\epsilon = -0.010 \pm 0.018$ for the full sample of 1378 quasar candidates, which have an average redshift of $\langle z \rangle = 3.38$. Using a simple chi-squared, goodness-of-fit test $\chi^2 = 1.73$ over 5 degrees of freedom (DOF), which corresponds to a p-value of $p = 0.885$ on the fitting scales. Our model is also consistent, within error, with the data at larger scales despite fitting over the range of $1'$ to $30'$. This behavior reveals the effect that the stellar contaminants have and suggests that our larger-scale correlation function is contaminated with stellar sources.

Over the same scales ($1'$ to $30'$), we fit the two-dimensional power-law model in Equation 3.22 to the data. The best-fit values from this two parameter model are $\theta_0 = 0.71 \pm 0.546$ and $\delta = 1.39 \pm 0.618$. Using only the best-fit amplitude of the power-law model, we estimate that the significant of this clustering result is $\sim 1.3\sigma$ above the null hypothesis of an unclustered sample (i.e. $\theta_0 = 0$ at all scales). Reducing the error bars inherent to our selection technique is not practical in the near future given the depth of *WISE* and the limited mapping capability of *Spitzer*; however, the combination of other deep and wide-area optical and infrared data in the near future, such as The Dark Energy Survey (DES; Diehl et al. 2014) and *Euclid* (**?**), should allow further progress.

3.4.2 Faint Quasar Clustering

The results in Figure 3.9 show the clustering strength of *all* of our candidate quasars, both bright ($i < 20.2$; 252 objects) and faint ($i \geq 20.2$; 1126 objects). In this analysis, we remove the bright quasars and cluster *only* the 1126 faint objects to directly test the degeneracy in the models of Hopkins et al. (2007a). The computation of the correlation function and bias is the same as the previous section, we simply change the redshift selection function in Limber's equation to match the new distribution. We find a best fit bias of $b = 6.64 \pm 2.23$ and $\epsilon = 0.005 \pm 0.022$ for this faint sample with an average redshift of $\langle z \rangle = 3.39$. The chi-squared test results in $\chi^2 = 0.45$, again over 5 degrees of freedom (DOF), which corresponds to a p-value of $p = 0.994$ on the fitting scales. The results of this analysis are shown in Figure 3.10. The error in this fit is much larger than in the full sample which we attribute to the size of the error bar at ~ 5 arcmin and the difference in value at $\sim$ 30 arcmin, which are both likely due to a smaller number density of objects. Despite this difference, the bias between this sample and the full sample are consistent; however, we focus on the full sample results in the next section.

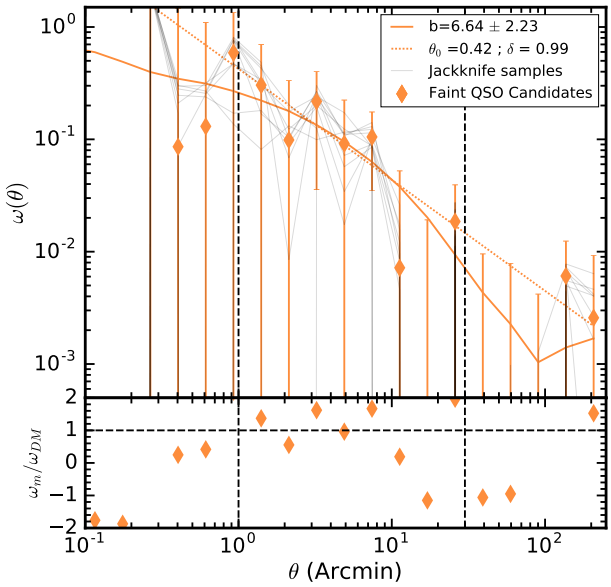

Figure 3.10 Clustering result of the faint sample of high redshift photometric candidates (orange diamonds). Fitting a new DM model to the data over the range $1'$ to $30'$ (black dashed vertical lines), we find a best fit bias of $b = 6.64 \pm 2.23$ (orange curve). Once again, we also model stellar contamination using Equation 3.21 with the new selection function for the faint objects. The dotted line indicates the best fit power-law with θ_0=0.42 ± 0.582 arcmin and δ=0.99 ± 0.502. As in Figure 3.9, we show the ratio of the data to the DM model in the lower panel and errors are computed with jackknife resampling (grey lines).

3.5 Implications

3.5.1 Comparison to Other Observations

This paper presents the first measurements of the autocorrelation function of photometrically-selected high-z quasars; however, there are other measurements of quasar clustering with which we can compare across our redshift range of interest. Here we examine the techniques and results of the surveys in the literature to those in our study.

We first compare our results to the results of the BOSS survey from Eftekharzadeh et al. (2015). This study examined the redshift-space correlation function of spectroscopically-confirmed quasars in the SDSS field in the redshift range of $2.2 \leq z \leq 2.8$. For a more direct comparison with our angular-projected correlation function, we compute the angular correlation function of the BOSS data from Eftekharzadeh et al. (2015), using their NGC CORE sample and a random catalog with five times the data; shown in the top panel of Figure 3.11. Despite spanning slightly disjoint redshift ranges, the two correlation functions agree on scales before contamination dominates the signal ($\sim 25h^{-1}$ Mpc or $\sim 20'$; Eftekharzadeh et al. 2015). Since these correlation functions have similar power in the clustering signal, yet are at different redshifts, the best fit bias values are different (see Figure 3.12).

Next, we compare with the results of He et al. (2018), who computed the quasar cross-correlation function (as opposed our measurement of the auto correlation function; ACF) for photometrically selected quasars in the redshift range $3 \leq z \leq 4$. In the He et al. (2018) investigation, quasars are selected using optical and near-infrared colors from the Hyper Suprime-Cam[14] (HSC). In total, they selected 1023 quasars as candidates across 172 deg^2, 901 of which were both faint ($i \geq 21$) and high-z. Using these candidates, they computed the cross-correlation function (CCF) between their candidates and Lyman-Break Galaxies at $z \sim 4$. Figure 3.11 (middle panel) depicts the results from the CCF analysis compared to our study. Since the measurement is performed with two different statistics, the amplitudes of the two $w(\theta)$ should not be directly compared; however the bias measurements from these two surveys can be compared, despite being computed with different

[14] https://www.naoj.org/Projects/HSC/surveyplan.html

Table 3.6. High-z Bias Measurements

Measurement	z-interval	$\langle z \rangle$	N_{qso}	bias	θ_0 (arcmin)	δ	M_i[z=2] faint, bright	M_{DMH} ($\times 10^{12} h^{-1} M_\odot$)
This work (all)	2.90, 5.10	3.48	1 378	6.78 ± 1.79	0.710 ± 0.546	1.39 ± 0.618	-23.80, -27.50	1.70 – 9.83
This work (faint)	2.90, 5.10	3.49	1 126	6.64 ± 2.23	0.420 ± 0.582	0.99 ± 0.502	-23.80, -26.40	1.04 – 10.6
He et al. (2018)[a]	3.00, 4.00	3.80	901	5.93 ± 1.43	0.148 ± 0.050	0.86[a]	-23.70, -25.86	1.00 – 2.00
Eftekharzadeh et al. (2015)[b]	2.64, 3.40	2.97	24 724	3.57 ± 0.09	–	–	-24.40, -29.31	0.60 – 0.72
Shen et al. (2007)[b,c]	2.90, 3.50	3.20	2 651	7.90 ± 0.80	–	–	-26.00, -30.00	2.00 – 3.00
Shen et al. (2007)[b,c]	3.50, 5.40	4.00	1 775	14.0 ± 2.00	–	–	-26.50, -30.00	4.00 – 6.00

Note. — Bias estimates for selected surveys of comparable redshifts to our study.
[a] Cross-correlation of the faint sample. Power law index held fixed at $\delta = 0.86$ in this study.
[b] Redshift space estimate, thus no angular power law information is given.
[c] Shen et al. (2007) results split into two redshift bins to reflect the bias values shown in Figure 3.12.

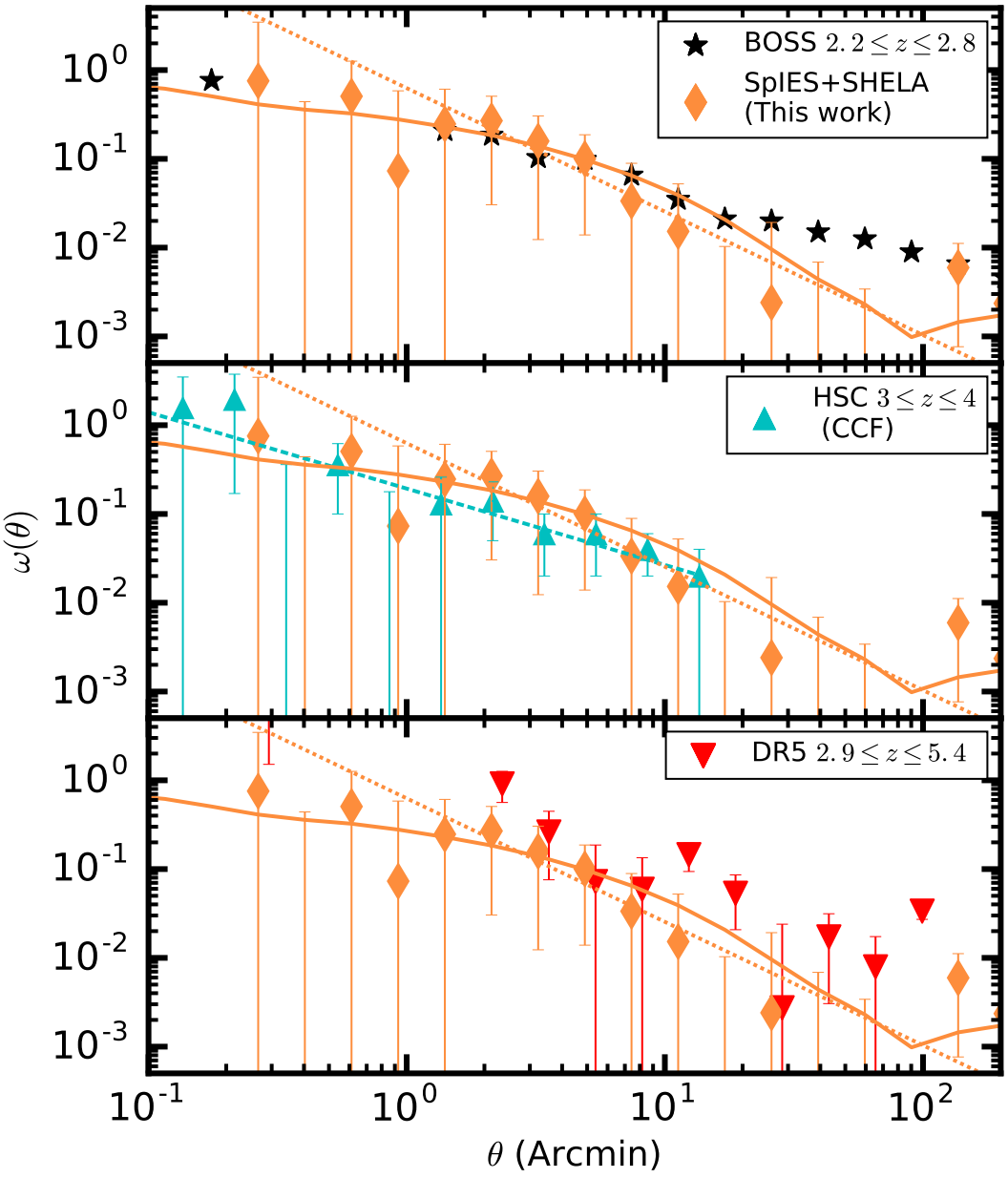

Figure 3.11 Top: Comparison of the clustering measurement from this study (orange diamonds) to the angular correlation function from the BOSS survey (black stars), which was computed using a subset of the data from Eftekharzadeh et al. (2015). Middle: Comparison to the CCF results of He et al. (2018) (light blue triangles). While the ACF (our study) and CCF (He et al., 2018) cannot be directly compared, these results cover approximately the same redshift range, and have slightly different bias values (see Figure 3.12). Bottom: Clustering results of the full redshift range in this study compared to that of spectroscopically-confirmed quasars from SDSS Data Release Five (inverted red triangles; Shen et al. 2007). The two surveys cover the same redshift range; however, the quasars in this study are significantly fainter than those in Shen et al. (2007), as shown in Figure 3.6. Poisson error bars are depicted for the Shen et al. (2007) data, using the data-data pair counts we estimate in our analysis. Points for Shen et al. (2007) are offset by $0.1\times\theta$ arcmin for clarity. We compare with these three surveys since they are closest in redshift range (although not exactly the same), and are consistent on scales before contamination dominates ($\sim 20'$; Eftekharzadeh et al. 2015).

statistics (CCF and ACF). Our ACF measurements find a bias of $b = 6.78 \pm 1.79$, and the bias from the CCF of the less-luminous ($i \geq 21$) quasars in He et al. (2018) is $b = 5.93 \pm 1.43$; both results are displayed in Figure 3.12. The biases of these two studies overlap within their measurement error, and can be interpreted using a similar physical model. We will discuss the physical implications of this model in Section 3.5.3. A larger sample of spectroscopic high-z quasars is needed to reduce the uncertainties in the bias measurement of high-z quasars.

We also compare our study over the full redshift range to the results of Shen et al. (2007), who investigated the clustering properties of spectroscopically-confirmed high-z quasars from SDSS Data Release five (DR5). These DR5 quasars span a redshift range of $2.9 \leq z \leq 5.4$, and are bright ($i \leq 20.2$; see Figure 3.6). With spectroscopic redshifts, Shen et al. (2007) present a measurement of the 3D redshift-space correlation function, so to compare their results to ours, we compute the angular projected correlation function using their data and the DR5 mask from Ross et al. (2009). The results are shown in the bottom panel of Figure 3.11. The correlation function is, in general, higher in amplitude for the objects in DR5 than our candidates over the relevant scales ($\sim 30'$), however we find a slightly smaller bias value than Shen et al. (2007).

The Shen et al. (2007) quasar sample has an i-band limiting magnitude of $M_i = -26.5$ (their Table 6), and is thus only sampling the very bright end of the quasar luminosity function. By contrast, our data as well as the data from He et al. (2018) have an i-band limiting magnitude of $M_i \simeq -24.0$. A direct comparison of the bias values (see Figure 3.12) between Shen et al. (2007), He et al. (2018), and our study hints at a level of luminosity-dependence of clustering for high-z ($z \geq 3$) quasars. This difference in clustering would suggest that, at $z \geq 3$, the mass of the dark-matter (DM) halo hosting bright quasars is larger than the host DM halo masses of low luminosity quasars. Luminosity dependence at high-z would be a fascinating result since, at low-z, it has been shown that clustering is weakly dependent on luminosity, if at all (da Ângela et al. 2008; Shen et al. 2009; Eftekharzadeh et al. 2015; Chehade et al. 2016).

3.5.2 Dark Matter Halo Mass

Using the measured quasar bias in this study, and the hypothesis that quasars are biased tracers of the underlying DM distribution, we can estimate the characteristic mass for a typical DM halo. Here we use the formalism of Tinker et al. (2010), who fit analytic models to the results of simulated clustering of DM halos in a flat ΛCDM cosmology. We adopt the fitting function in Equation (6) of Tinker et al. (2010):

$$b(\nu) = 1 - A\frac{\nu^a}{\nu^a + \delta_c^a} + B\nu^b + C\nu^c, \tag{3.23}$$

where $b(\nu)$ is the measured bias in our study and ν is the "peak height" of the density field defined by $\nu = \delta_c/\sigma(M)$. Here, the peak height is defined in terms of the critical density for collapse of the DM halo (δ_c=1.686) and the linear matter variance at the radial scale of each halo, $R_{halo} = (3M_{halo}/4\pi\bar{\rho}_m)^{1/3}$ ($\bar{\rho}_m = 2.78 \times 10^{11}\Omega_m h^2 M_\odot$; He et al. 2018), defined by:

$$\sigma^2(M) = \frac{1}{2\pi^2}\int P(k,z)\hat{W}^2(k,R)k^2 dk. \tag{3.24}$$

We estimate the matter power spectrum, $P(k,z)$, using CAMB and our adopted cosmology, where $\hat{W}(k,R)$ is the spherical top-hat window function;

$$\hat{W}(k,R) = \frac{3}{(kR)^3}\left(\sin(kR) - kR\cos(kR)\right). \tag{3.25}$$

The parameters A, a, B, b, C, c in Equation (3.23) are adopted from Table 2 of Tinker et al. (2010) for $\Delta = 200$, where Δ is the ratio of mean density to background density (similarly used in

Eftekharzadeh et al. 2015, DiPompeo et al. 2016, He et al. 2018):

$$
\begin{aligned}
y &= \log_{10}(\Delta) \\
A &= 1 + 0.24ye^{-(4/y)^4} \\
a &= 0.44y - 0.88 \\
B &= 0.183 \\
b &= 1.5 \\
C &= 0.019 + 0.107y + 0.19e^{-(4/y)^4} \\
c &= 2.4
\end{aligned}
\tag{3.26}
$$

Using the measured bias values in Equation 3.23, the power spectrum from CAMB, and the parameters defined above, we can solve for the characteristic halo mass (see Table 3.6). For our measured bias over the full redshift range of $b = 6.78 \pm 1.79$, the characteristic halo mass ranges between 1.70–9.83$\times 10^{12} h^{-1} M_{\odot}$. Computing the halo mass from the bias estimated using only the faint quasars, yields 1.04–10.56$\times 10^{12} h^{-1} M_{\odot}$, where the large mass ranges in both estimates are a direct result of the large uncertainty in the bias values.

We compare our estimated halo masses to the masses found in Shen et al. (2007) who computed the minimum halo mass, which is slightly different from our computation in that an estimate of the luminosity function is required. Over the redshift range of $2.9 \leq z \leq 3.5$, Shen et al. (2007) find a minimum halo mass of $\sim$(2-3)$\times 10^{12} h^{-1} M_{\odot}$, and in the redshift range $z \geq 3.5$, Shen et al. (2007) estimates a minimum halo mass of $\sim$(4–6)$\times 10^{12} h^{-1} M_{\odot}$.

The low-z halo mass estimate from Eftekharzadeh et al. (2015) of $\sim$0.66 $\times 10^{12} h^{-1} M_{\odot}$ over the redshift range of $2.64 \leq z \leq 3.4$ (their Table 7), is a factor of ten smaller than our results; however they also report halo masses on the redshift range $2.20 \leq z \leq 2.80$ of $\sim$ 1.2–2.8 $\times 10^{12} h^{-1} M_{\odot}$, which is $\sim 3\times$ smaller than our result. This difference arises from the different redshifts as well as the large difference in bias. The high-z estimate of the He et al. (2018) less-luminous sample is 1–2$\times 10^{12} h^{-1} M_{\odot}$. Again, the difference here is mainly due to the difference in bias between the two

studies. However, if we take these results at face value it does imply that less luminous quasars tend to have smaller halo masses at high-z. A larger sample of spectroscopically confirmed faint, high-z quasars is needed to answer this question with greater certainty. If we could increase the number of pair counts along the fitting scales by 50%, we estimate that the error bars would decrease by $\sim$ 20% (using the Poisson error estimate which scales as $DD^{-0.5}$). More data would reduce the error on the bias which, in turn, leads to a tighter constraint on the DM halo masses.

3.5.3 Implications for Feedback

The measurement of the 2PCF and bias of the faint, high-z quasars in this study is ideal information to constrain the feedback mechanisms presented in Hopkins et al. (2007a). The Hopkins et al. (2007a) study compared the clustering of quasars and galaxies as a function of different intrinsic properties

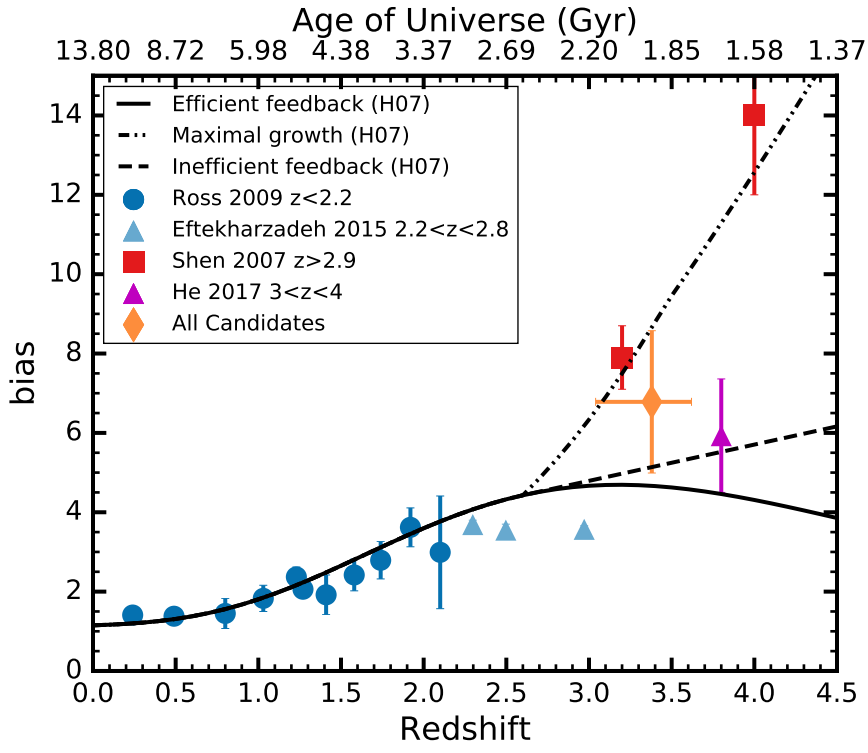

Figure 3.12 The evolution of the bias with redshift. We show the bias result for our full candidate sample (orange diamond). Also displayed are the feedback models from Hopkins et al. (2007a) as well as the low- and intermediate redshift measurements from Ross et al. (2009) (dark blue circles) and Eftekharzadeh et al. (2015) (light blue triangles), respectively. Finally, we show the high-z bias of the bright quasars from Shen et al. (2007) (red squares) and the new HSC study from He et al. (2018) (purple triangle). The bias increases with redshift in our sample and tends to agree with the "inefficient feedback" model, however we cannot rule out the "maximal growth" model.

(e.g., mass, luminosity, redshift) to investigate triggering mechanisms and the growth of the quasar and galaxy populations. Included in this study was an analysis of how different quasar feedback mechanisms affect their clustering strength. These models were designed to fit measured results at low-z (e.g., Croom et al. 2005, among others) which we represent using the Ross et al. (2009) results, yet vary at high-z ($z \geq 3$). This study highlighted three feedback scenarios; *"efficient"* and *"inefficient"* feedback, as well as a *"maximal growth"* model. We depict the clustering predictions from these three models as the black lines in Figure 3.12, and provide a brief explanation below.

The solid line in Figure 3.12 depicts the clustering evolution with redshift if BH growth shuts down after the quasar epoch. This is the "efficient" feedback model in Hopkins et al. (2007a), and assumes that quasars represent a single, short-lived, phase in the growth of the central BH. Here, feedback efficiently terminates the quasar phase, and the central BH ceases its growth. This model assumes that the observed properties of quasars at $z < 2$ are the same as at higher redshifts, thus the predicted clustering strength weakens at high-z to reflect observations at low-z.

Additionally, Hopkins et al. (2007a) presents a model in which quasars, and their central BHs, grow intermittently until $z \sim 2.5$ when "downsizing" begins (the dashed line in Figure 3.12). In this model, the quasar grows with the luminosity function, and the evolution of the luminosity function is dictated by the same objects growing hierarchically. Thus, feedback is "inefficient" since the BH continues to grow over various epochs, as opposed to the first model where, after the initial quasar phase, BH growth ends. This also means that brighter quasars live in very massive DM halos and fainter quasars would live in smaller DM halos at early times.

The "maximal growth" model postulates that the central BHs continue to grow proportionally with the DM halo until $z \sim 2$. This model assumes that quasars are continually accreting at their Eddington rates. Here, feedback is not only inefficient prior to $z \sim 2$ but is not sufficient to stop the BH from growing at its most maximal rate. These quasars live in the highest mass DM halos which accumulates gas unimpeded by the radiation from the central quasar. Therefore, the predicted clustering is very high from this model is shown by the dot-dashed line in Figure 3.12.

At low-z, the three models are designed to match measurements of the 2PCF (Croom et al. 2005;

Myers et al. 2007; Ross et al. 2009), but beyond $z \sim 3$, the models diverge. Additionally, the three models become degenerate for a sample of quasars with $i \leq 20.2$ (Hopkins et al., 2007a); all taking the form of the "maximal growth" model. Our study, however, examines quasars fainter than the limit at high-z, thus breaking the degeneracy between the models in the redshift range $3 \leq z \leq 4$.

Figure 3.12 displays the best fit bias result over all of our candidates over the full redshift range in this study (orange diamond). The bias of the of faint candidates ($i \geq 20.2$) is not depicted, however is consistent with the full result. Also depicted in Figure 3.12 is the error in both the bias, which is a result from fitting the dark-matter model, and in redshift, where, since our redshift distribution is not Gaussian, we depict the first and third quartile of the redshifts (as opposed to the standard deviation).

Within the error of these results, the bias in our study overlaps both the "maximal growth" model and the "inefficient feedback" model, as shown in Figure 3.12, for the full sample of candidates in this analysis. The "maximal growth" model is also consistent with the results of Shen et al. (2007); however, we remind the reader that our investigation clustered a different population of quasars than Shen et al. (2007). We analyzed the clustering of *faint* quasars and are therefore capable of breaking the degeneracy limit noted in Hopkins et al. (2007a). As shown in Figure 3.12, our result deviates from the "maximal growth" model toward the "inefficient feedback" model, which coincides with the result from He et al. (2018) at $z \sim 4$. The "inefficient feedback" model predicts that feedback from the central BH intermittently shuts down the accretion of gas onto the BH at early times. This model also suggests a degree of luminosity dependence of quasar clustering at high-z and that fainter quasars live in less massive DM halos as compared to bright quasars. To better understand these models at $z \sim 3.4$ will likely require a larger sample of spectroscopically-confirmed quasars that are both faint, and high-redshift.

At first glance, it may appear that the findings in Eftekharzadeh et al. (2015) contradict our results; however, a significant difference in the bias measurements between our study and Eftekharzadeh et al. (2015) can be attributed to the difference in the redshift selection functions. While Figure 3.11 shows that our results and the angular correlation function of Eftekharzadeh et al. (2015) have

a similar amplitude, the DM model is strongly dependent on the redshift selection function. Lower redshift ranges result in larger power in the angular correlation function model which, in turn, results in a smaller bias fit (i.e., decreasing redshift in the model shifts the orange curve in Figure 3.11 to the right). As a result, we expect Eftekharzadeh et al. (2015) to have a lower bias than our investigation despite having similar amplitudes in angular correlation space. Taking these bias values at face value shows a rapid change in the bias at $z \sim 3.1$. Understanding this jump in bias at this particular redshift will be the topic of future work.

3.6 Summary

In this investigation, we have determined the two-point autocorrelation function of 1378 photometrically-selected, faint ($i \geq 20.2$), high-z ($2.9 \leq z \leq 5.1$) quasars across $\sim$100 deg^2 on SDSS S82. Details about this catalog as well as our main findings are as follows:

- We combine the deep optical photometry on S82 from SDSS with new, deep MIR information from the SpIES and SHELA surveys to form a comprehensive catalog of photometric objects. Utilizing their optical/MIR colors, and the colors of known high-z quasars from the Richards et al. (2015) composite catalog (see Figure 3.2), we use three machine-learning algorithms to select 1378 faint, high-z quasar candidates.

- We estimate the photometric redshifts of these candidates using Nadaraya-Watson kernel regression. When tested on spectroscopic quasars, this algorithm predicts photometric redshifts within a range of $z_{phot} - z_{spec} = 0.1$ for 93% of the quasars (Figure 3.3). The overlap in color-redshift space between the photometric candidates and the known quasars with which they were selected is presented (Figure 3.4).

- Figure 3.6 demonstrates that our candidates are generally fainter than the objects used in the Shen et al. (2007) study. This aspect of our sample helps to break the degeneracy between the feedback models studied in Hopkins et al. (2007a).

- Utilizing the estimator from Landy & Szalay (1993), we compute the angular 2PCF of our faint high-z quasars, where a random mask is generated using MANGLE (Figure 3.7). The

correlation function result is presented in Figure 3.9

- We estimate a linear bias using the method of Limber (1953) which relates the 3D DM power spectrum to the angular correlation function. We compute the 3D power spectrum using CAMB and our fiducial cosmology. Over the full redshift range of our sample ($\langle z \rangle = 3.38$), the bias is $b = 6.78 \pm 1.79$. The best-fit values from the power law model are $\theta_0 = 0.71 \pm 0.546$ and $\delta = 1.39 \pm 0.618$.

- In Figure 3.10, we remove the bright objects and recompute the correlation function of 1126 faint quasar candidates. We find the faint quasars have a bias of $b = 6.64 \pm 2.23$, similar to the full study. The agreement in bias demonstrates that the bright quasars in the sample do not skew the bias result of the faint objects. We compare the results of our full study with other surveys in Figure 3.11.

- Using the estimates of bias, we compute characteristic DM halo masses using the formalism of Tinker et al. (2010). Our quasars inhabit DM halos with masses of 1.70–9.83$\times 10^{12} h^{-1} M_{\odot}$. This mass estimate covers a wide range due to the large uncertainty in the bias.

- We use our bias estimate to constrain the feedback models of Hopkins et al. (2007a) in Figure 3.12. Our data is consistent with both the "maximal growth" model, which assumes that the central quasar is not powerful enough to shut down accretion of material onto the BH, as well as the "inefficient feedback" model, which suggests that feedback from the central source intermittently shuts down accretion of the central BH. The "inefficient feedback" model, however, also coincides with the bias of faint quasars at $z \sim 4$ found in He et al. (2018). Finally, the "inefficient feedback" model suggests that fainter quasars sit in smaller DM halos.

Further studies of the 2PCF of faint, high-z quasars will benefit from the new optical and infrared surveys on the horizon. Surveys performed with the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009) in the optical and the Wide-Field Infrared Survey Telescope (WFIRST; Spergel et al. 2013) and, to an extent, the James Webb Space Telescope (JWST; Gardner et al. 2006) in the infrared will be able to observe fainter than what we have

now. These surveys will add an immense amount of data to our sample and a significant amount of area which, in turn, increases the significance of the results. Similarly, spectroscopic investigation on the candidates will allow us to add to the high-z training data, as well as make the necessary corrections to our photometric redshifts to compute the redshift-space 2PCF. In this investigation, however, we have demonstrated that, using machine-learning techniques, we can both select faint, high-z quasars cleanly and compute the 2PCF on these samples. These techniques will be crucial in the next phase of astronomy, which will be dominated by photometric data that lacks detailed spectroscopic follow-up.

Acknowledgements

This work is based on observations made with the *Spitzer* Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech. We acknowledge support from NASA-ADAP grant NNX17AF04G. ADM was partially supported by NSF grants 1515404 and 1616168 and NASA grant NNX16AN48G. FEB acknowledges support from CONICYT-Chile (Basal-CATA PFB-06/2007, FONDECYT Regular 1141218), the Ministry of Economy, Development, and Tourism's Millennium Science Initiative through grant IC120009, awarded to The Millennium Institute of Astrophysics, MAS. GTR acknowledges support from NSF grant 1411773.

We make our full datasets, analyses code and methodologies available at `https://github.com/JDTimlin/QSO_Clustering/tree/master/highz_clustering`

For this research, we use the Python language and Astropy[15] (Astropy Collaboration et al., 2013). TOPCAT[16] (Taylor, 2005).

We thank Michael DiPompeo and Ryan Hickox for their correspondence and advice on both our contamination checks and in our interpretations of the feedback models. We also thank Yao-Yuan Mao for the DECaLS image list tool (`https://github.com/yymao/decals-image-list-tool`) used in this study.

[15] astropy.org
[16] starlink.ac.uk/topcat

3.7 Appendix

3.7.1 Contamination Checks

Contamination in any clustering sample can drastically change the correlation function and the resulting bias. We carefully define our sample in this study to avoid contamination as much as possible. As part of this work, we also performed a clustering analysis using the selection results without restricting to the point sources alone. We found that, if we just use color selection and do not check for low-z contamination, we get a bias value of $b \sim 5$ instead of $b \sim 6.5$, which would lead us to different conclusions in Figure 3.12. It is therefore very important that we eliminate as much contamination as possible in this study.

While we explicitly model stellar contamination in this study, there are other forms of contamination that dilute the clustering signal. The two main sources of additional contamination are mis-identification of objects in the classification algorithms, and regions where the angular mask of the random objects is not identical to the data. Here we describe our methods to identify and reduce contamination from galaxies in our analysis.

3.7.2 Extinction Cut

As mentioned in Section 3.2.5, we cut the overlapping region between SpIES and the outskirts of the disk of the Milky Way ($330 \leq \alpha_{J2000} \leq 344.4$ which corresponds to a galactic latitude of $-51.5 \leq b_{gal} \leq -41.5$) to eliminate highly-extincted objects from the analysis which act as contaminants in the clustering signal. Figure 3.13 depicts the clustering result before (green circles) and after (orange diamonds) this extinction cut, as well as their best fit DM models (which have slightly different redshift selection functions). These models are fit as before using an efficiency of $e = 0.86$ which means that 14% of the sample are stellar contaminants. At large scales, the model (green curve) lies below the measured clustering strength, which implies that there are more contaminants than estimated using just stellar contamination. After the extinction cut is performed, however, there is much better agreement between the model and the data (in fact, it appears that the model over-estimates the contamination at large scales). Deep infrared spectra are required to determine the particular type of object contaminating the sample, however it is most likely stars

that were reddened by Galactic dust such as late type M-dwarf stars. These objects would not appear in optically–selected samples, however since we include the infrared colors in our selection, they could be selected as quasars.

While the extinction cut resulted in a loss of ~ 20 deg^2, it also significantly decreased the power of the correlation function at larger scales (see Figure 3.13). There were, however, objects in that field with lower extinction measurements that were also cut. Ideally, we would keep these objects to use in our correlation function measurements, but cutting on the extinction value causes the density to drop significantly in this area, which affects the correlation function if not properly accounted for in the angular mask. Our future work to remedy this problem is to change the density of the random mask in this field to reflect the data.

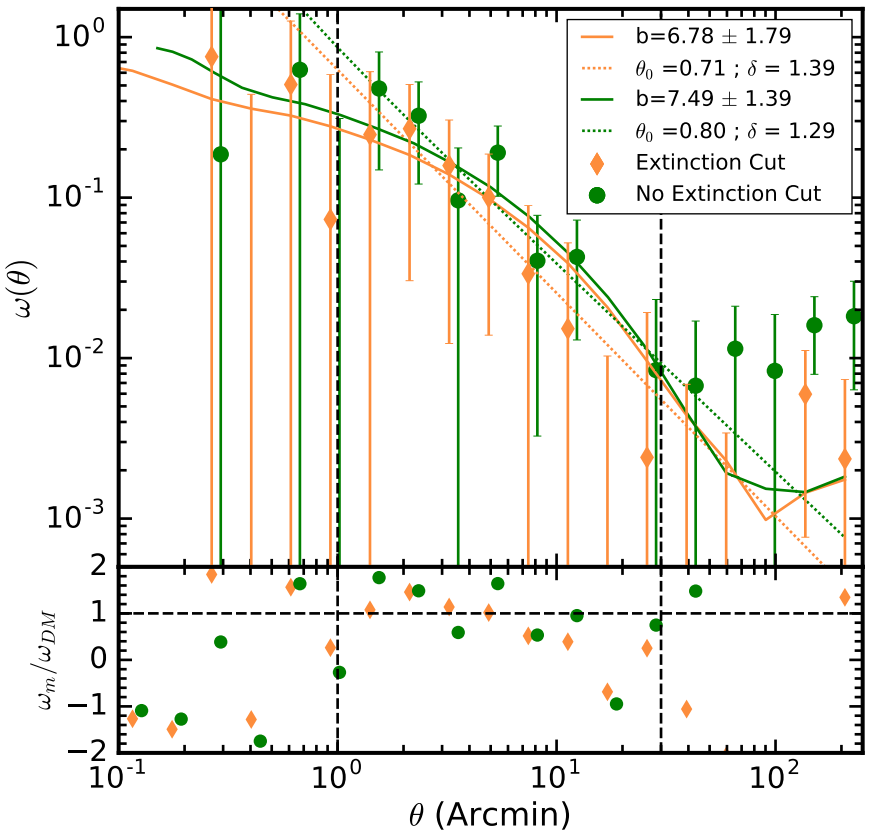

Figure 3.13 Correlation function of the final sample after the extinction cut (orange diamonds) compared to the correlation function of the full sample of objects, including the region $330 \leq \alpha_{J2000} \leq 344.4$ (green circles; offset by 0.1 arcmin for clarity). While the correlation function is similar over the fitting range ($1' \leq \theta \leq 30'$), the power at larger scale is significantly higher for the full study compared to the extinction cut survey. The green circles are offset by $0.1 * \theta$ for clarity.

3.7.3 Visual Inspection

Visual inspection using the DECam Legacy Survey image cutout tool[17] enabled us to examine the classified objects and eliminate obvious sources of contamination. The superior depth and resolution of DECam is crucial for the follow-up visual inspection of the candidates that were selected in each algorithm. This inspection also drove the need to create the point-source metric we used to cut all extended objects in this study. We note that fainter quasars are more likely to be classified as extended emission, thus spectroscopic follow-up is needed on *all* faint candidates, not just the point sources used in this study.

Figure 3.14 Visual inspection examples of contamination using image cutouts from DECam for objects classified as quasars in this study. Left: Obvious galaxy contaminant selected by our algorithm. This object is a low-redshift galaxy which has similar $u-g$ colors to quasars with $z \leq 3$. Center: An object selected by our algorithm that exhibits extended behavior, but is not visually an obvious contaminant like the galaxy in the left panel. Objects like this are removed in our final clustering result; spectroscopic follow-up is needed to classify these objects as galaxies. Right: A known high-z quasar that we also select using our algorithm. This particular object is at $z \sim 3.7$ and is a typical point source commonly associated with quasar activity. The co-added color of this quasar appears to be green, however quasars in this study can have a range of colors in DECam Legacy Survey cutouts, depending on their redshift. Each frame is $\sim 45''$ on a side.

Figure 3.14 depicts three types of objects that passed the high-z quasar selection algorithms (in either redshift range). In the left panel we show local galaxies ($z \sim 0.3$) which, as a result of the 4000Å break in their spectra, can be mistaken for the Lyman-α forest from high-z quasars (at $z \sim 3.5$). This confusion causes the low-z galaxies to pass the machine-learning selection. These are obvious contaminants that were easily detected and removed by hand.

We also selected objects that appeared to have extended emission; an example of which is shown in the center panel of Figure 3.14. While these objects could be galaxies at higher redshift (e.g.,

[17] `https://github.com/yymao/decals-image-list-tool`

Lyman-Break galaxies; He et al. 2018), it is also possible that they could be faint quasars at high-z whose emission from the central engine is not bright enough to outshine the host galaxy. For the faint quasars in our study, this could certainly be the case. These objects did not pass our final point source metric and thus were removed from our final analysis.

Finally, in the right-hand panel of Figure 3.14, we show a known quasar at redshift $z \sim 3.7$ which our machine-learning algorithm also classifies as a high-z quasar. This object passes the point-source metric and is thus included in this study. Most of the objects that we call point like have similar profiles to this object (albeit, some are much fainter). Once again, spectroscopic follow-up is needed on these objects as well for a combination of testing the classification and testing the redshifts estimates from our machine-learning algorithms.

3.7.4 Error Estimates and Fitting Parameters

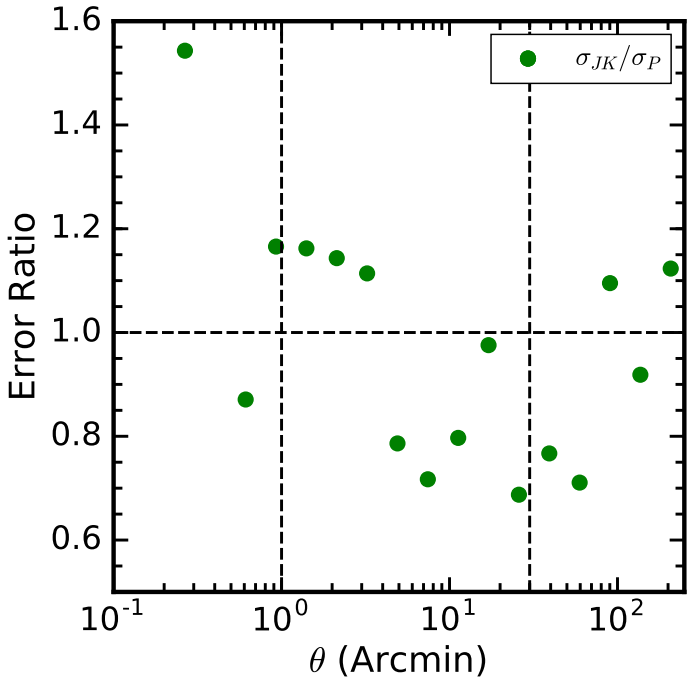

Figure 3.15 Ratio of the Jackknife errors to Poisson errors for the full quasar candidate sample. Poisson errors were computed using the pair counts reported in Table 3.5. In this investigation, we replace the Jackknife errors with Poisson errors wherever the ratio is less than unity.

To ensure that we obtain reasonable jackknife errors, we compare our errors in Table 3.5 to the

Poisson errors (Peebles, 1973) defined as:

$$\sigma_{Poisson} = \frac{1 + \omega(\theta)}{\sqrt{DD}}. \tag{3.27}$$

Poisson error is a measure of the noise due to the number of pairs in the sample (Ross et al., 2009), and is most valid at smaller scales where pairs of objects are independent of each other (**?**). We depict the ratio of our Jackknife errors to the Poisson errors in Figure 3.15. Poisson errors represent a minimum standard deviation in a clustering measurement, particularly on the smallest scales, thus the ratio of the Jackknife to Poisson errors should be of order unity. In this investigation, we replace the Jackknife errors with Poisson errors wherever the ratio of the two in Figure 3.15 is less than one.

We also test the best fit parameters from both the power law model and the dark-matter model by generating χ^2 maps for each space. We compute the χ^2 as:

$$\chi^2 = \sum \frac{(\omega_{\rm measured}(\theta) - \omega_{\rm model}(\theta))^2}{\sigma^2} \tag{3.28}$$

For our power law model, we iterate the power law index over the range $0.5 \leq \delta \leq 2.2$ in 300 steps and the correlation angle $0 \leq \theta_0 \leq 1.2$ in 400 steps and compute the χ^2 value. Figure 3.16 depicts the results of this analysis for the full sample of quasar candidates (left) and the faint sample (right). In both cases, we find that the best-fit parameters given in Table 3.6 (represented by the black 'x') lie in the region of the minimum χ^2. We also plot the 1σ region and find that it is consistent with the ranges given in Table 3.6.

Figure 3.17 depicts the χ^2 map of the dark-matter model, and is computed in a similar manner as before. Here, we iterate both the bias values over a range of $3 \leq$ b ≤ 9 in 300 steps, and the cross-correlation term over a range $-0.03 \leq \epsilon \leq 0.01$ in 600 steps. Again, we find that the values reported in Table 3.6 are consistent with the minimum χ^2 value, and the errors span an appropriate range.

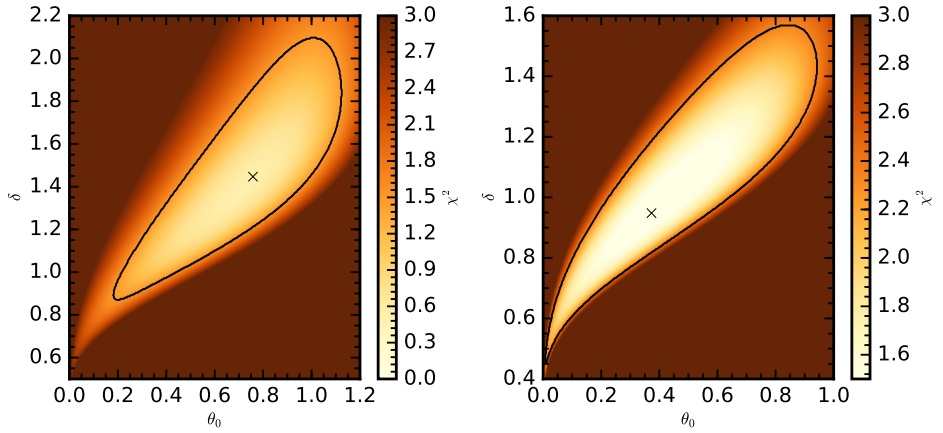

Figure 3.16 χ^2 map of the free parameters in the power law model the full (left) and faint (right) samples of photometrically selected quasars. The black point depicts the location of the minimum value of χ^2 which corresponds to the best fit values in Table 3.6. The black contour outlines the 1σ region in this space. The range in sigma in both dimensions is representative of the range recorded in Table 3.6.

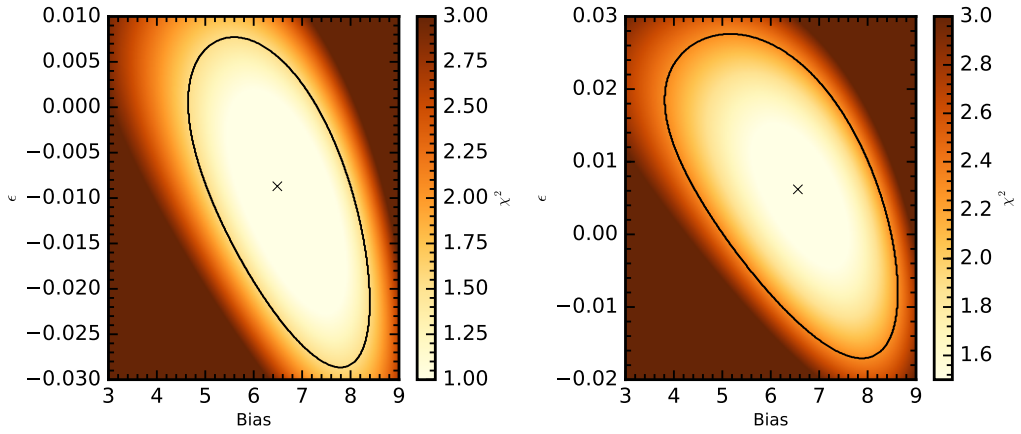

Figure 3.17 χ^2 map of the free parameters in the DM model for the full (left) and faint (right) samples of photometrically selected quasars. As in Figure 3.16, the black point depicts the location of the minimum χ^2 corresponding to the values in Table 3.6. The black contour outlines the 1σ region in this space, and reflects the errors presented in Table 3.6.

Chapter 4: Conclusion

4.1 The *Spitzer* IRAC Equatorial Survey

In this work, we have presented the stacked images and catalogs from the *Spitzer* IRAC Equatorial Survey. Spanning ∼115 deg^2 on SDSS Stripe 82, SpIES is the largest area *Spitzer* survey to date. The equatorial Stripe 82 field was specifically chosen for this survey because of the deep photometric and spectroscopic optical data that already exists, along with a plethora of multi-wavelength observations which, when combined, provides a complete picture of the objects in the field. SpIES was also designed to adjoin with the SHELA survey, which added ∼20 deg^2 of continuous mid-infrared coverage on Stripe 82.

The SpIES images were taken in a 3.6 micron and a 4.5 micron filter, and consisted of over 70,000 individual pointings of the *Spitzer* telescope. All of these images were cleaned of artifacts such as latent images and column pulldown and were stacked together. The observations were split into segments of 8×28 sequential pointings of the telescope called AORs (shown in Figure 2.2), where each AOR was observed over two epochs to identify transient objects, and each epoch consisted of two dithers, or small offsets, which eliminated the effects of bad pixels in the images. For each filter, all of the pointings in each AOR were stacked to generate the final, full depth, SpIES images that were used to generate source catalogs.

This work presents a single-band catalog for each observed wavelength, as well as a dual-band catalog which contains objects that were detected in both bands. There are a total of 5.4 million objects in the dual-band catalog, with ∼3.9 million being flagged as 'highly reliable' and having a signal-to-noise greater than three. We measured the positional accuracy of these objects against matching point sources in the SDSS catalogs, and have implemented slight corrections to the SpIES positions. Additionally, we report the measured photometry from six circular apertures as well as their respective photometric errors, which are estimates of the combined background error and Poisson error. Finally, we include several flags, such as a bright star mask, which indicates which

objects are close to bright stars, and coverage flag, which reports how many pointings covered each object, as a service to the user.

The completeness of the survey, which reports the fraction of objects recovered in the extraction process as a function of magnitude, was estimated using a Monte Carlo approach. Simulated point sources of varying magnitude were randomly placed in our stacked AOR images that were then subject to the source extraction routine. In Figure 2.12, we showed that SpIES is 90% complete at AB magnitudes of 21.75 in the 3.6 micron detector and 21.90 in the 4.5 micron detector, similar to other surveys of similar depth. These completeness values can also be used as an estimate of the depth of the survey, however we further estimate depth using the magnitude errors reported in the dual-band catalogs. A 5σ error in flux corresponds to a magnitude error of $\sim$0.2, which we find to be at an AB magnitude of 21.93 in the 3.6 micron images and 22.00 in the 4.5 micron images, as depicted in Figure 2.14.

Finally, this work demonstrated that the SpIES images were deep enough to fulfill the mission goal of detecting high-z quasars. The SpIES dual-band catalog was matched to the composite catalog of Richards et al. (2015) which contains known quasars that are both faint and high-z. SpIES recovered $\sim$ 98% of the high-z ($z \geq 3.5$) quasars in the field and matched to $\sim$ 94% of the faint ($i \geq 20.2$) and high-z quasars in this field. Figure 2.17 showed that SpIES has both the sensitivity and the resolution to detect quasars at redshift $\sim$ 6 and quasars as faint as 22nd magnitude in the SDSS i-band. Therefore, the SpIES catalogs are capable of detecting and selecting high-z quasars for a clustering measurement. All images and catalogs were made public to the broader astronomy community in 2016.

4.2 High-Redshift Quasar Clustering

The deep infrared data from SpIES and SHELA was combined with the deep optical data from SDSS to classify high-z quasars photometrically. Using the optical and infrared information for the known high-z quasars (depicted in Figure 3.2), we trained three machine-learning algorithms to photometrically select faint, high-z quasars from the SpIES and SHELA footprint. To reduce contamination from galaxies in this selection, we tested the robustness of the machine-learning

algorithms at different redshift ranges. We found that selection from the Bootstrap Aggregation of K-Nearest Neighbors was the most robust in a lower redshift range ($2.9 \leq z \leq 3.4$) whereas all three algorithms performed equally well at higher redshifts ($3.4 \leq z \leq 5.2$; see Table 3.3). We therefore split the selection into these two redshift ranges to reduce contamination. Also, to be considered a quasar, we limited the sample to point sources only. Our final sample consisted of 652 new, high-z quasar candidates which we combined with 726 spectroscopically-confirmed or previously selected quasars, resulting in a data set of 1378 objects in the SpIES+SHELA field with which to compute the correlation function.

Photometric redshifts were then estimated for each of these candidates using the Nadaraya-Watson regression algorithm, which is a form of basis function regression. Once again, the photometric information from the known quasars was used to train the photo-z algorithm. When tested on quasars with spectroscopic redshifts, the estimated redshift from photometric colors and the spectroscopic redshift do not differ by more than $\Delta z \sim 0.1$ for approximately 93% of these known high-z quasars, as shown in Figure 3.3. Using these redshift estimates, we compute the absolute magnitude of the selected quasar candidates, and show in Figure 3.6 that they are fainter than the quasars from Shen et al. (2007), which measured the high-z quasar correlation function for the brightest quasar per square degree on the sky. In fact, the selected candidates delve deep enough to break degeneracy in the feedback models of Hopkins et al. (2007a) between $3 \leq z \leq 4$.

To compute the angular correlation function, we employ the estimator from Landy & Szalay (1993) which is calculated by comparing the angular distribution of the data to that of a random field. To avoid noise in the correlation function, however, it is imperative that the random points cover an identical field to the data, so to ensure that the field boundaries are the same we input the corner positions of the SpIES and SHELA AORs into the MANGLE software. Using these positions, MANGLE draws and snaps together spherical caps to form a continuous polygon on the sky. Within this polygon, we generate a random distribution of points which we compare the to data to estimate the correlation function.

We then fit a dark-matter model to the clustering result to obtain a linear bias factor. The dark-

matter model used in this work is defined by Limber's Equation, which is the Fourier Transform of the three-dimensional dark-matter power spectrum for a ΛCDM universe, projected into two-dimensions. We used a combination of CAMB, to compute the dark-matter power spectrum, and Monte Carlo integration to generate a model to fit to the data. Using a least-squares fit, we found a linear bias of $b = 6.78 \pm 1.79$. Using this value of bias and the formalism of Tinker et al. (2010), we estimate the characteristic dark matter halo mass of the quasars in this sample to be 1.70–$9.83 \times 10^{12} h^{-1} M_{\odot}$, which is consistent with other results. We also fit a single power-law model to the correlation function and found an index of $\delta = 1.39 \pm 0.618$ and amplitude of $\theta_0 = 0.71 \pm 0.546$ arcmin. The correlation function and best-fit models are depicted in Figure 3.9.

Finally, in Figure 3.12 we compare our measured bias result with the predicted bias from each of the feedback models in Hopkins et al. (2007a). Our measurement is most consistent with both the 'inefficient feedback' model across our redshift range, however, due to the uncertainty in our result, we also overlap the 'maximal growth' model. The 'inefficient feedback' model implies that the quasar is intermittently shutting down at early times in the universe. This means that the radiation from the central source is powerful enough to push gas away from the SMBH, however the gas does not escape the gravitational potential. Therefore, the gas will eventually fall back into the central region which, in turn, reactivates the quasar. This model also hints at a level of luminosity dependence of quasar clustering at high-z; a trend which has been observed to weakly exist, if at all, at lower redshift. We have a higher confidence in the 'inefficient feedback' model due to the work of He et al. (2018), who computed the cross-correlation function of faint, high-z quasars and galaxies and recovered a bias that agrees with ours.

In summary, we have presented both the images and source detection catalogs for the SpIES survey which is the largest *Spitzer* survey to date. We have discussed why SpIES is ideally placed to detect and classify faint, high-z quasars, and we have demonstrated that SpIES is sensitive enough to recover known, high-z quasars as faint as $i \sim 22$ and as distant as $z \sim 6$. After matching the source catalogs with deep optical data from SDSS, we trained three machine-learning algorithms on known quasars to classify quasars at high-z. Applying these trained algorithms to our matched

objects, we have found 652 new quasar candidates which we combine with 726 known or previously selected quasars to generate a catalog of 1378 quasar candidates. We then computed the angular two-point autocorrelation function of these candidates, and fit a dark matter model to the results which gives an estimate of the bias. This bias is most consistent with the 'inefficient feedback' model in Hopkins et al. (2007a) which predicts intermittent quasar activity in the early universe. Finally, this result also hints at a luminosity dependence of quasar clustering which has been shown to not exist at lower redshifts.

4.3 Future Work

This work inspires a few different follow-up projects. First, and foremost, we need spectroscopic measurements of the 652 quasar candidates that were selected in our algorithm, as well as the candidates selected in Richards et al. (2015) and Peters et al. (2015) that we use in this investigation. This will help confirm that the majority of these objects are, in fact, quasars, and it will also provide true redshift measurements as opposed to the estimates from the Nadaraya-Watson algorithm. With accurate measurements of the redshift, the three-dimensional correlation function can be accurately computed which provides another estimate of the linear bias of the quasars at high redshift. This will help to confirm the result presented here and reduce the uncertainty in this result.

Along these lines, we can reduce the uncertainty by adding data to the analysis. Until recently, the Stripe 82 field has been home to some of the largest-area, deep optical and infrared data crucial for the selection of faint, high$z$ quasars. New optical data from the Dark Energy Survey (DES; Diehl et al. 2014) in the South Pole Deep Field can be combined with existing *Spitzer* data from the *Spitzer*-South Pole Telescope Deep Field (SSDF; Ashby et al. 2013). The SSDF field probes to approximately the same depth as SpIES, and would effectively double the area that could be used to measure clustering. Combining the infrared and optical data in this field would allow for photometric selection, similar to our work, however the optical data does not contain the u-band, so some testing would be required here, particularly when selecting in the $2.9 \leq z \leq 3.4$ redshift range. With some testing, we are confident that another measurement of the clustering of faint, high-z quasars can be performed with better accuracy using a combination of the SpIES+SHELA+SSDF

fields.

With or without spectroscopic follow-up on the candidates, an estimate of the quasar luminosity function (QLF) can be made with the data in hand. The candidates in our analysis are faint enough to find the turn-over of the QLF at high-z. This measurement will be very useful to test the known X-ray QLF results of Glikman et al. (2011) and Giallongo et al. (2015). To perform this measurement, one will need to generate a sample of simulated quasars which can be passed through our selection results to more accurately model the completeness of the algorithms used to select the candidates. With that completeness function, however, we can use the formalism of Ross et al. (2013) and Peters et al. (2015) to estimate the QLF of these faint quasars.

Bibliography

Aird, J., Coil, A. L., & Georgakakis, A. 2018, MNRAS, 474, 1225

Alam, S., Ata, M., Bailey, S., et al. 2016, ArXiv e-prints, arXiv:1607.03155

Alexandroff, R., Strauss, M. A., Greene, J. E., et al. 2013, MNRAS, 435, 3306

Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120

Antonucci, R. 1993, ARA&A, 31, 473

Arendt, R. G., Odegard, N., Weiland, J. L., et al. 1998, ApJ, 508, 74

Ashby, M. L. N., Stern, D., Brodwin, M., et al. 2009, ApJ, 701, 428

Ashby, M. L. N., Stanford, S. A., Brodwin, M., et al. 2013, ApJS, 209, 22

Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, ApJ, 713, 970

Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26

Assef, R. J., Eisenhardt, P. R. M., Stern, D., et al. 2015, ApJ, 804, 27

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, Astron. & Astrophys., 558, A33

Ata, M., Baumgarten, F., Bautista, J., et al. 2017, ArXiv e-prints, arXiv:1705.06373

Bautista, J. E., Busca, N. G., Guy, J., et al. 2017, ArXiv e-prints, arXiv:1702.00176

Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559

Bell, E. F. 2008, ApJ, 682, 355

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

Bonometto, S., Gorini, V., & Moschella, U., eds. 2002, Modern cosmology

Bovy, J., Hennawi, J. F., Hogg, D. W., Myers, A. D., & Ross, N. P. 2011, in Bulletin of the American Astronomical Society, Vol. 43, American Astronomical Society Meeting Abstracts #217, 222.05

Bower, R. G., Schaye, J., Frenk, C. S., et al. 2017, MNRAS, 465, 32

Boyle, B. J., Shanks, T., Croom, S. M., et al. 2000, MNRAS, 317, 1014

Brewer, J. 2008, PhD thesis, University of Pittsburgh

Bundy, K., Leauthaud, A., Saito, S., et al. 2015, ArXiv e-prints, arXiv:1509.01276

Chehade, B., Shanks, T., Findlay, J., et al. 2016, MNRAS, 459, 1179

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

Condon, J. J. 1974, ApJ, 188, 279

Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2004, MNRAS, 349, 1397

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

Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 392, 19

da Ângela, J., Shanks, T., Croom, S. M., et al. 2008, MNRAS, 383, 565

Daddi, E., Alexander, D. M., Dickinson, M., et al. 2007, ApJ, 670, 173

Das, S., Louis, T., Nolta, M. R., et al. 2014, Journal of Cosmology and Astroparticle Physics, 2014, 014

Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJ Lett., 660, L1

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

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44

de la Torre, S., Guzzo, L., Peacock, J. A., et al. 2013, Astron. & Astrophys., 557, A54

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

Diehl, H. T., Abbott, T. M. C., Annis, J., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9149, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0

DiPompeo, M. A., Hickox, R. C., & Myers, A. D. 2016, MNRAS, 456, 924

DiPompeo, M. A., Myers, A. D., Hickox, R. C., Geach, J. E., & Hainline, K. N. 2014, MNRAS, 442, 3443

DiPompeo, M. A., Myers, A. D., Hickox, R. C., et al. 2015, MNRAS, 446, 3492

Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, ApJ, 748, 142

Donoso, E., Yan, L., Stern, D., & Assef, R. J. 2014, ApJ, 789, 44

Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429

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

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

Fabricant, D., Fata, R., Roll, J., et al. 2005, PASP, 117, 1411

Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10

Feigelson, E. D., & Babu, G. J. 2012, Modern Statistical Methods for Astronomy

Font-Ribera, A., Arnau, E., Miralda-Escudé, J., et al. 2013, **??**jnlJ. Cosmology Astropart. Phys., 5, 018

Fox, O. D., Johansson, J., Kasliwal, M., et al. 2015, ArXiv e-prints, arXiv:1510.08070 [astro-ph.HE]

Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Science Reviews, 123, 485

Geach, J. E., Hickox, R. C., Bleem, L. E., et al. 2013, ApJ Lett., 776, L41

Giallongo, E., Grazian, A., Fiore, F., et al. 2015, Astron. & Astrophys., 578, A83

Glikman, E., Djorgovski, S. G., Stern, D., et al. 2011, ApJ Lett., 728, L26

Glikman, E., Urrutia, T., Lacy, M., et al. 2013, ApJ, 778, 127

Gralla, M. B., Crichton, D., Marriage, T. A., et al. 2014, MNRAS, 445, 460

He, W., Akiyama, M., Bosch, J., et al. 2018, PASJ, 70, S33

Helfand, D. J., White, R. L., & Becker, R. H. 2015, ApJ, 801, 26

Hickox, R. C., Jones, C., Forman, W. R., et al. 2007, ApJ, 671, 1365

—. 2009, ApJ, 696, 891

Hickox, R. C., Myers, A. D., Brodwin, M., et al. 2011, ApJ, 731, 117

Hill, G. J., MacQueen, P. J., Palunas, P., Barnes, S. I., & Shetrone, M. D. 2008, in Proc. SPIE, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, 701406

Hodge, J. A., Becker, R. H., White, R. L., Richards, G. T., & Zeimann, G. R. 2011, AJ, 142, 3

Hogg, D. W. 2001, AJ, 121, 1207

Hopkins, P. F., Lidz, A., Hernquist, L., et al. 2007a, ApJ, 662, 110

Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007b, ApJ, 654, 731

Houck, J. R., Roellig, T. L., van Cleve, J., et al. 2004, ApJS, 154, 18

Ivezić, Ž., Connelly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learningin Astronomy

Jarvis, M. J., Bhatnagar, S., Bruggen, M., et al. 2014, ArXiv e-prints, arXiv:1401.4018 [astro-ph.CO]

Jiang, L., Fan, X., Bian, F., et al. 2014, ApJS, 213, 12

Kaiser, N. 1984, ApJ Lett., 284, L9

Krolewski, A. G., & Eisenstein, D. J. 2015, ApJ, 803, 4

Labbe, I., Oesch, P. A., Illingworth, G. D., et al. 2015, ArXiv e-prints, arXiv:1507.08313

Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, ApJS, 154, 166

LaMassa, S. M., Urry, C. M., Glikman, E., et al. 2013a, MNRAS, 432, 1351

LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2013b, MNRAS, 436, 3581

—. 2015, ArXiv e-prints, arXiv:1510.00852

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

Laurent, P., Eftekharzadeh, S., Le Goff, J.-M., et al. 2017, ArXiv e-prints, arXiv:1705.04718

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, Astron. & Astrophys., 439, 845

Liddle, A. R., & Lyth, D. H. 2000, Cosmological Inflation and Large-Scale Structure, 414

Limber, D. N. 1953, ApJ, 117, 134

Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, PASP, 115, 897

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201 [astro-ph.IM]

Lucas, P. W., Tinney, C. G., Burningham, B., et al. 2010, MNRAS, 408, L56

Lynden-Bell, D. 1969, Nat, 223, 690

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ Lett., 619, L1

Martinez, V., & Saar, E. 2002, in Proc. SPIE, Vol. 4847, Astronomical Data Analysis II, ed. J.-L. Starck & F. D. Murtagh, 86

Mauduit, J.-C., Lacy, M., Farrah, D., et al. 2012, PASP, 124, 714

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, ApJ, 768, 105

McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, The Messenger, 154, 35

Mendez, A. J., Coil, A. L., Aird, J., et al. 2016, ApJ, 821, 55

Mooley, K. P., Myers, S. T., Hallinan, G., et al. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts 223, 236.02

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

Myers, A. D., Brunner, R. J., Richards, G. T., et al. 2006, ApJ, 638, 622

Myers, A. D., Palanque-Delabrouille, N., Prakash, A., et al. 2015, ApJS, 221, 27

Oke, J. B., & Gunn, J. E. 1983a, ApJ, 266, 713

—. 1983b, ApJ, 266, 713

Oliver, S. J., Bock, J., Altieri, B., et al. 2012, MNRAS, 424, 1614

Papovich, C., Shipley, H. V., Mehrtens, N., et al. 2016, ArXiv e-prints, arXiv:1603.05660

Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, Astron. & Astrophys., 563, A54

Peacock, J. A. 1999, Cosmological Physics, 704

Peebles, P. J. E. 1973, ApJ, 185, 413

—. 1980, The large-scale structure of the universe

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

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, Astron. & Astrophys., 594, A13

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

Rees, M. J. 1984, ARA&A, 22, 471

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

Richards, G. T., Nichol, R. C., Gray, A. G., et al. 2004, ApJS, 155, 257

Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, ApJS, 166, 470

Richards, G. T., Deo, R. P., Lacy, M., et al. 2009, AJ, 137, 3884

Richards, G. T., Myers, A. D., Peters, C. M., et al. 2015, ApJS, 219, 39

Richstone, D., Ajhar, E. A., Bender, R., et al. 1998, Nat, 395, A14

Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, ApJS, 154, 25

Rodríguez-Torres, S. A., Chuang, C.-H., Prada, F., et al. 2016, MNRAS, 460, 1173

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

Ross, N. P., McGreer, I. D., White, M., et al. 2013, ApJ, 773, 14

Salpeter, E. E. 1964, ApJ, 140, 796

Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

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

Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, AJ, 110, 68

Scranton, R., Johnston, D., Dodelson, S., et al. 2002, ApJ, 579, 48

Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, AJ, 133, 2222

Shen, Y., Strauss, M. A., Ross, N. P., et al. 2009, ApJ, 697, 1656

Shen, Y., McBride, C. K., White, M., et al. 2013, ApJ, 778, 98

Sherwin, B. D., Das, S., Hajian, A., et al. 2012, Phys. Rev. D, 86, 083006

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Simon, P. 2007, Astron. & Astrophys., 473, 711

Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311

Soltan, A. 1982, MNRAS, 200, 115

Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, ArXiv e-prints, arXiv:1305.5422 [astro-ph.IM]

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nat, 435, 629

Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, ApJ Lett., 791, L25

Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, ApJ, 631, 163

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485

Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782

Surace, J. A., Shupe, D. L., Fang, F., et al. 2005, in Bulletin of the American Astronomical Society, Vol. 37, American Astronomical Society Meeting Abstracts, 1246

Swanson, M. E. C., Tegmark, M., Hamilton, A. J. S., & Hill, J. C. 2008, MNRAS, 387, 1391

Taylor, M. B. 2005, TOPCAT & STIL: Starlink Table/VOTable Processing Software

Timlin, J. D., Ross, N. P., Richards, G. T., et al. 2016, ApJS, 225, 1

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Totsuji, H., & Kihara, T. 1969, PASJ, 21, 221

Viero, M. P., Asboth, V., Roseboom, I. G., et al. 2014, ApJS, 210, 22

Wainscoat, R. J., Cohen, M., Volk, K., Walker, H. J., & Schwartz, D. E. 1992, ApJS, 83, 111

Wang, W.-H., Cowie, L. L., & Barger, A. J. 2006, ApJ, 647, 74

Weinstein, M. A., Richards, G. T., Schneider, D. P., et al. 2004, ApJS, 155, 243

Werner, M. W., Gallagher, D. B., & Irace, W. R. 2004, Advances in Space Research, 34, 600

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

Worseck, G., & Prochaska, J. X. 2011, ApJ, 728, 23

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59

Vita

John David Timlin III

Education

Drexel University, Philadelphia, Pennsylvania USA

Ph.D., Physics, March 2018

Advisor: Professor Gordon Richards

M.S., Physics, June 2014

Millersville University, Millersville, Pennsylvania USA

B.A., Physics, May 2012, *Magna Cum Laude*

Publications

The Astrophysical Journal: Supplements

SpIES: The *Spitzer* IRAC Equatorial Survey, *July 2016*

The Astrophysical Journal

The Clustering of High-Redshift ($2.9 \leq z \leq 5.1$) Quasars in SDSS Stripe 82, *February 2018*

Presentations & Conferences

American Astronomical Society

Poster Presentations, 2013-2016, 2018

Dissertation Talk, 2017

Hidden Monsters

Poster Presentation, 2016-2017

Awards & Honors

Millersville University

Faraday Physics Scholarship, 2008-2010

Harry A. '65 and Carolyn J. Lohss Physics Scholarship, 2010-2012

Teaching Experience

Teaching Assistant, Drexel University (2012-2014)

Tutor, Millersville University (2009-2012)