EFFICIENT PHOTOMETRIC SELECTION OF QUASARS FROM THE SLOAN DIGITAL SKY SURVEY. II. ~1,000,000 QUASARS FROM DATA RELEASE 6

GORDON T. RICHARDS[1,2,8], ADAM D. MYERS[3], ALEXANDER G. GRAY[4], RYAN N. RIEGEL[4], ROBERT C. NICHOL[5], ROBERT J. BRUNNER[3], ALEXANDER S. SZALAY[2], DONALD P. SCHNEIDER[6], AND SCOTT F. ANDERSON[7]

[1] Department of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA
[2] Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218-2686, USA
[3] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801-3080, USA
[4] Center for Experimental Research in Computer Systems, Georgia Institute of Technology, 240 Technology Square Research Building, 85 5th St. NW, Atlanta, GA 30318, USA
[5] Institute of Cosmology and Gravitation, Mercantile House, Hampshire Terrace, University of Portsmouth, Portsmouth, PO1 2EG, UK
[6] Department of Astronomy and Astrophysics, The Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802, USA
[7] Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA
Received 2008 May 1; accepted 2008 September 4; published 2008 December 23

ABSTRACT

We present a catalog of 1,172,157 quasar candidates selected from the photometric imaging data of the Sloan Digital Sky Survey (SDSS). The objects are all point sources to a limiting magnitude of $i = 21.3$ from 8417 deg^2 of imaging from SDSS Data Release 6 (DR6). This sample extends our previous catalog by using the latest SDSS public release data and probing both ultraviolet (UV)-excess and high-redshift quasars. While the addition of high-redshift candidates reduces the overall efficiency (quasars:quasar candidates) of the catalog to ~80%, it is expected to contain no fewer than 850,000 bona fide quasars, which is ~8 times the number of our previous sample and ~10 times the size of the largest spectroscopic quasar catalog. Cross-matching between our photometric catalog and spectroscopic quasar catalogs from both the SDSS and 2dF survey yields 88,879 *spectroscopically* confirmed quasars. For judicious selection of the most robust UV-excess sources (~500,000 objects in all), the efficiency is nearly 97%—more than sufficient for detailed statistical analyses. The catalog's completeness to type 1 (broad-line) quasars is expected to be no worse than 70%, with most missing objects occurring at $z < 0.7$ and $2.5 < z < 3.0$. In addition to classification information, we provide photometric redshift estimates (typically good to $\Delta z \pm 0.3$ [2σ]) and cross-matching with radio, X-ray, and proper-motion catalogs. Finally, we consider the catalog's utility for determining the optical luminosity function of quasars and are able to confirm the flattening of the bright-end slope of the quasar luminosity function at $z \sim 4$ as compared to $z \sim 2$.

Key words: catalogs – quasars: general

Online-only material: color figures, machine-readable tables

1. INTRODUCTION

The number of known quasars has grown exponentially since their discovery by Maarten Schmidt in 1963 (Figure 1). There have been relatively frequent compilations of heterogeneous catalogs over the years and the 100, 1000, and 10,000 quasar marks were reached in 1967, 1977, and 1998, respectively (see Hewitt & Burbidge 1993; Véron-Cetty & Véron 2006, and references therein). Early quasar discoveries were often based on heterogeneous samples and/or previously existing photometric surveys, so the exact lineage of the growth of homogeneous samples is more difficult to trace. However, the number of spectroscopically-confirmed, optically-selected quasars in a single homogeneous survey had certainly reached 100 by 1977 (e.g., MacAlpine et al. 1977). The 1000 quasar mark was broken during the Large Bright Quasar Survey (LBQS) in 1991 (Morris et al. 1991; Hewett et al. 1995). The 2dF Quasar Redshift Survey (2QZ; Boyle et al. 2000) first cataloged 10,000 quasars by 2001 (Croom et al. 2001), soon followed by the Sloan Digital Sky Survey (SDSS; York et al. 2000) Quasar Survey (Schneider et al. 2007).

While the number of known quasars continues to grow at a rapid pace (e.g., Schneider et al. 2007), the 100,000 object mark was broken years ahead of the extrapolated trend (see Figure 1)

by this group's *photometric* sample in 2004 (Richards et al. 2004; hereafter Paper I). Quasar catalogs, used for meaningful statistical analyses, are almost always spectroscopic. This is in contrast to galaxies, for which a wealth of major statistical studies utilized purely photometric catalogs (e.g., Maddox et al. 1990). Historically, this has been due to an inability to obtain ~90% or greater star–quasar separation efficiency to match the typical star–galaxy separation, readily obtainable from morphology. For instance, standard ultraviolet (UV)-excess (UVX) quasar selection (e.g., Croom et al. 2001) is ~50% efficient and the SDSS's official quasar targeting efficiency is ~80% (at best) for bright ($i = 19.1$) UVX sources (Richards et al. 2002). The ~95% efficiency (Richards et al. 2004; Myers et al. 2006) of our catalog thus heralded the era of statistically useful photometric star–quasar separation, opening up a new avenue for quasar studies.

Using the most recent SDSS data release (Adelman-McCarthy et al. 2008), this paper marks the next milestone by presenting a homogeneous photometric catalog of nearly one *million* quasars. Unfortunately, with our current approach, this trend will likely moderate in the near future, as this sample covers 8417 deg^2 to $i = 21.3$ and there are only 41253 deg^2 in our sky. On the other hand, large-scale synoptic surveys such as the Large Synoptic Survey Telescope (LSST; Tyson 2002), the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Kaiser et al. 2002), and the Dark Energy Survey

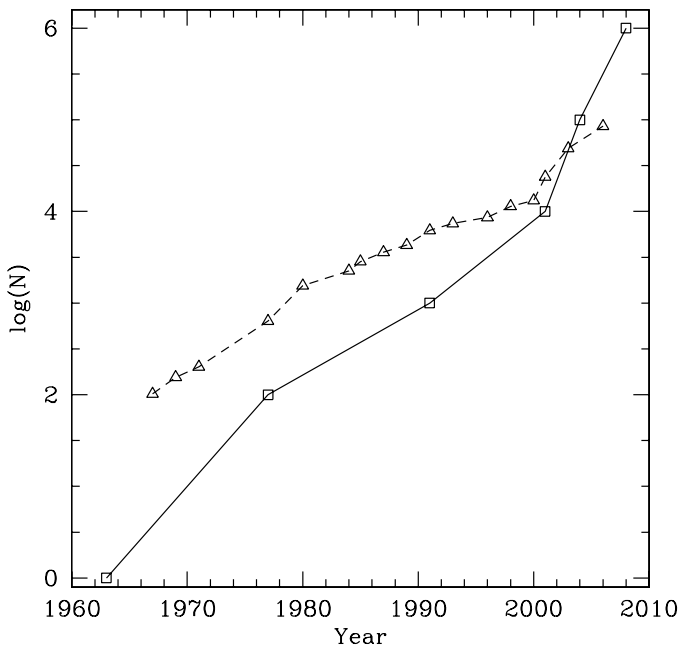---

[8] Alfred P. Sloan Research Fellow.

Figure 1. Growth in the number of known quasars in the largest homogeneous (solid) and heterogeneous (dashed) quasar catalogs as a function of time. See Hewitt & Burbidge (1993), Véron-Cetty & Véron (2006), and references therein.

(DES; The Dark Energy Survey Collaboration 2005) will, in the next decade, enable another order of magnitude gain by taking advantage of fainter photometric limits and quasar variability. In the meantime, an alternative path allows us to anticipate an explosion in the number of obscured (so-called type 2) quasars (Antonucci 1993), which are expected to outnumber the type 1 quasars cataloged herein by up to a factor of 4-to-1 (e.g., Lacy et al. 2004; Treister et al. 2004; Brandt & Hasinger 2005; Polletta et al. 2008; Reyes et al. 2008) and whose numbers will increase as the *Spitzer Space Telescope* (*SST*) maps ever larger areas of sky during its warm mission.

The need for robust photometric classification has rapidly become apparent and will be an absolute necessity by the time LSST and Pan-STARRS are fully underway. Even with multiobject spectrographs observing thousands of objects per square degree at a time, the small fields and relatively long exposure times mean that it will simply never be possible to obtain spectra of all of objects identified. In addition, new science goals nearly always demand an increased sample size. Indeed, this has been aptly demonstrated by previous work on the far smaller versions of this catalog. Much of the new science that used our catalogs detected subtle cosmological effects that were previously impossible without a large quasar catalog, but also highlighted the need for more extensive samples with which to study elusive aspects of cosmology and the quasar population.

For example, Myers et al. (2006) explored quasar clustering using the Paper I catalog—the first such study of quasar evolution in a photometric catalog—and found results consistent with spectroscopic surveys. This study was expanded in Myers et al. (2007a), providing a luminosity baseline large enough to uniquely constrain topical models of quasar activity, but still with too few objects with which to constrain any luminosity dependence to quasar clustering. Hennawi et al. (2006) used the catalog to enhance their study of binary quasars, and detected the first definitive evidence for excess quasar clustering on small scales. In Myers et al. (2007b), we further examined small-scale quasar clustering, providing a homogeneous catalog of binary

quasar candidates. Myers et al. (2008) present spectroscopic observations of pairs of photometric quasar candidates and are able to place only weak constraints on any redshift dependence to small-scale quasar clustering at $z < 2$, providing yet more impetus to produce a larger catalog over a wider redshift range. These papers on the clustering of our photometric quasars provided critical input to the clustering analysis done by Hopkins et al. (2007a). Cross-correlating with the cosmic microwave background, Giannantonio et al. (2006, 2008) used the large number of photometric quasars to constrain dark energy using the Integrated Sachs–Wolfe (ISW) effect (Sachs & Wolfe 1967), the first detection of the ISW effect using optically-selected quasars. These measurements represent one of the most robust measurements of dark energy at high redshift and are found to be consistent with predictions for flat Λ-dominated cold dark matter (Λ CDM) models (see Giannantonio et al. 2008). Finally, after many years of contradictory results in the field, Scranton et al. (2005) used photometric quasars to categorically measure cosmic magnification bias, detecting the effect of gravitational lensing by foreground galaxies on quasar source counts at $\sim 8\sigma$.

This paper is laid out as follows. Section 2 briefly describes the data. Section 3 reviews the Bayesian selection algorithm, discusses the changes from Richards et al. (2004), and describes the construction of the training and test data sets. The catalog itself (in Tables 1–3) is presented in Section 4. Various catalog properties and diagnostics of the efficiency and completeness are described as is our prescription for limiting the catalog to particularly robust subsamples. We also discuss matching of the catalog to nonoptical object catalogs and the determination of photometric redshifts. Finally, a rough analysis of the number counts and luminosity function are given in Section 5.

2. THE DATA

The photometric imaging data that this catalog is based upon are from SDSS Data Release 6 (DR6; Adelman-McCarthy et al. 2008). We specifically used the SQL interface to the Catalog Archive Server (CAS) to extract point sources (type=6) with i-band magnitudes between 14.5 and (de-reddened) 21.3 (psfmag$_i$ >14.5 && psfmag$_i$ −extinction$_i$ <21.3). Note that the bright limit uses magnitudes uncorrected for Galactic extinction since the purpose of this limit is to reject objects that may be saturated in the imaging. Throughout this paper, we utilize über-calibrated point-spread-function (PSF) magnitudes, which are now available in the SDSS database.[9] The über-calibrated magnitudes (Padmanabhan et al. 2008) represent the most robust photometric measurements as they are calibrated across SDSS "stripes" to a single uniform photometric system for the entire SDSS area. The SDSS photometric system is described in Fukugita et al. (1996) and Smith et al. (2002). The SDSS photometric measurements are expressed in asinh magnitudes (Lupton et al. 1999). All magnitudes reported herein have been corrected for Galactic extinction using the Schlegel et al. (1998) dust maps.

We specifically queried the **photoObjAll** table, requiring mode=1 in order to limit the sample to "primary" detections (see Stoughton et al. 2002 for the details of SDSS database flags). The DR6 primary imaging data cover an area of 8417 deg². As the SDSS databases are designed to be maximally inclusive, one must carefully cull the object lists for false positive detections.

---

[9] Objects with $i < 21.3$ prior to über-calibration were also included in our sample for the sake of completeness.

We thus exclude objects using criteria similar to those described on the SDSS Web site[10]; also see Table 2 of Bramich et al. (2008) for similar criteria. As we include a cut on certain bad objects in SDSS run numbers 2189 and 2190, the total effective area covered by this catalog should be reduced by ∼75 deg^2.

Further details regarding the SDSS data set and the first six data releases (DRx) can be found in the series of SDSS technical papers (e.g., Adelman-McCarthy et al. 2008, and references therein). Familiarity with those papers will assist in optimal use of the catalog presented herein. Details of the camera and telescope systems are given by Gunn et al. (1998, 2006). Photometric processing details are discussed by Hogg et al. (2001), Lupton et al. (2001), Pier et al. (2003), Ivezić et al. (2004), and Tucker et al. (2006). Given that we match the catalog to objects with spectroscopy, details of the tiling (Blanton et al. 2003) and (point-source) target selection algorithms (Richards et al. 2002; Stoughton et al. 2002) may also be of interest.

3. OBJECT CLASSIFICATION

3.1. Overview

Paper I describes the details of our Bayesian classification algorithm. Herein, we make a few changes to the procedure, but, overall, the concepts are the same, so we present only a brief review of the most relevant aspects. Our goal is simply to take an unknown data set and assign one of two distinct classes to each object based on the colors of that object: quasar or star (or more specifically nonquasar). To accomplish this, we first build *training sets* of quasars and stars that serve as classification templates. Then, for each object in the *test set* of unknown objects that we wish to classify, we compute the probability of each object being a quasar or star.

The probability of belonging to a certain class of given parameter(s), x, is the likelihood of x under the probability density function (pdf), which describes that class, i.e., $p(x|C)$, where C is the class of object. Rather than describing the pdf with a histogram of discrete bins whose centers are pre-ordained, we instead use a *kernel density estimate* (KDE; Silverman 1986) of the pdf. The KDE defines each bin by its center point and the extent of the bin by a continuous *kernel function*. In our case, this kernel function will be either Gaussian or Epanechnikov (like a truncated Gaussian).

As we are not completely ignorant with regard to the most likely classification (e.g., the vast majority of objects in our initial test set are stars), we take a Bayesian (Bayes 1763) approach and factor in our prior belief regarding the class of each object (at least in the ensemble average), denoted as $P(C)$. Thus, the posterior probability, $P(C|x)$, of an object belonging to Class 1, C_1, will be

$$P(C_1|x) = \frac{p(x|C_1)P(C_1)}{p(x|C_1)P(C_1) + p(x|C_2)P(C_2)}, \quad (1)$$

where C_2 indicates Class 2. A class is then assigned to each object according to whether $P(C|x)$ is greater or less than 0.5. We refer to the resulting overall classifier as a nonparametric Bayes classifier (NBC); it is sometimes also called kernel discriminant analysis (KDA) or kernel density classification.

3.2. The Training Sets

The parameters, x, that we use for classification are simply the four primary SDSS colors ($u − g$, $g − r$, $r − i$, $i − z$). Thus,

we attempt a classification in four-dimensional color space as compared with the more traditional two-dimensional color-space selection or even the three-dimensional algorithms used by the formal SDSS quasar targeting algorithm (Richards et al. 2002). We define *training sets* of stars and quasars as discussed below and will use their four-dimensional SDSS colors as the basis of our classification. All objects in the training set are equally weighted in the classification. Photometric errors are not currently considered explicitly, but they are implicitly accounted for by the distributions of the training sets.

3.2.1. Quasars

For the quasar training set, we start with the 77,429 hand-vetted SDSS-DR5 quasars with spectra as cataloged by Schneider et al. (2007), which is based upon the SDSS-DR5 data (Adelman-McCarthy et al. 2007). These quasars span a redshift range of $0.08 \leqslant z \leqslant 5.4$. Initially, no additional cuts based on luminosity, morphology, selection method, photometric errors, etc. are applied. However, after the initial classification, we realized that at the faintest limits of our photometric catalog, there is some level of galaxy contamination (see Section 4.5.1), so for the final training set we chose to exclude all of the known quasars that are extended. This decision reduces our completeness to $z \lesssim 0.7$ quasars (see Section 4.4), but improves the overall efficiency of the algorithm.

As one of the goals of this paper is to extend the catalog in Paper I to higher redshifts, we supplement the DR5 quasar catalog with three other data sets. This is perhaps less necessary than it might have been for Paper I as the initial training set is now more than a factor of 4 larger and has correspondingly more high-redshift quasars. Nevertheless, high-redshift quasars are rare and the SDSS algorithm is known to be incomplete in certain redshift regions (Richards et al. 2006), thus, we include three additional sources of high-redshift quasars.

We first supplement the SDSS-DR5 quasar catalog with quasars discovered during the first observing season (2006) of the AAOmega-UKIDSS-SDSS (AUS) QSO Survey. This program targets $2.8 < z < 5.5$, $i < 21.6$ quasars with the AAOmega spectrograph on the Anglo-Australian Telescope (AAT) in order to fill a crucial gap in the redshift (and magnitude) coverage of quasars. This data set adds another 304 spectroscopically-confirmed quasars (of which 121 have $z > 2.2$). In addition, 131 confirmed nonquasars are added to the stars training set. While the numbers are small in comparison with the SDSS-DR5 sample, these objects span an important range of parameter space.

Next, we include all of the $z > 5.7$ quasars discovered by the SDSS to date; see Fan et al. (2006). This addition expands the upper-redshift limit of our training set from $z = 5.4$ to $z \sim 6.3$. Note that the $5.4 < z < 5.7$ region is under-represented by the main SDSS quasar survey and subsequent work, but these objects have colors sufficiently similar to $z \sim 5.4$ and $z \sim 5.7$ quasars and sufficiently different colors from most stars that they should still be identified as photometric quasar candidates (albeit with contamination from L/T dwarfs).

Finally, we included 920 objects that were selected as highly likely quasar candidates from cross-comparison of SDSS and *Spitzer* data. These are objects that meet the two-dimensional mid-infrared (MIR) color ("wedge") selection criteria of both Lacy et al. (2004) and Stern et al. (2005) in addition to our own three-dimensional Bayesian criteria using MIR colors from *Spitzer*–IRAC (Richards et al. 2008). They are also unresolved point sources in the SDSS imaging, have red MIR colors

---

[10] http://www.sdss.org/dr6/products/catalogs/flags.html

(whereas stars are blue in the MIR), are limited to $i < 20.2$ (while SDSS goes to $i = 21.3$), and are brighter than $S_{8\,\mu m} > 100\ \mu$Jy. Although these objects are photometrically selected, they are relatively bright point sources selected as quasars by four separate methods and are expected to unambiguously be type 1 quasars. Inclusion of such objects provides a crucial vector for multidimensional photometric selection of quasars at redshifts where traditional optical methods have difficulty (e.g., Richards et al. 2002).

The final quasar training set includes 75,382 confirmed quasars.

Note that our quasar training set is largely limited to $i < 19.1$ at redshifts less than 3 and $i < 20.2$ at redshifts higher than 3, yet we attempt to classify quasars to $i < 21.3$. Typically, it is inadvisable to extrapolate the results of a classification algorithm beyond the parameter space represented by the training set. However, there is no strong evidence for significant color changes in quasars (apparent or absolute) save brighter quasars tending to be slightly bluer (e.g., Vanden Berk et al. 2004). Therefore, modulo larger photometric errors for fainter objects, the parameter space of our training set should remain representative of all $i < 21.3$ quasars that we attempt to classify.

3.2.2. Stars

For the stars training set, we have roughly two classes of objects to consider. First are those stars with colors that are quite different from quasars. Second are objects that are more easily confused with quasars.

To account for the general population of stars, we extracted a random sample of ∼1% of all reliable point sources in the SDSS-DR6 imaging area with $14.5 < g < 21.3$, totaling 441,335 objects; see Section 2. As discussed in Paper I, unlike for quasars, we do not have a fully representative spectroscopic sample of stars to use as our training set. Thus, this sample of "stars" is really a point source sample and will include quasars as a contaminant. As a result, we first clean the stars training set of objects that are most likely to be quasars by running the stars training set through the classification algorithm. For this step, we took a star prior of 0.8 (roughly consistent with the fraction of stars in the initial training set) and removed any objects initially classified as quasars by our algorithm. In this step, we also removed objects that are known radio or X-ray sources (since point-like radio/X-ray sources are likely to be quasars) and with existing quasar spectral classification. This process removes ∼10,000 objects from the stellar training set. Spectroscopically-confirmed stars are retained.

In addition, past experience has shown that H II regions in galaxies can sometimes have colors that can be confused with quasars (either intrinsically or due to deblending problems). To help remove such sources, we inspected the images of all (a few hundred) pairs with $\leqslant 6''$ image separation, previously classified as quasars by an initial pass of our algorithm (see, e.g., the discussion in Myers et al. 2007b). The 317 galactic H II regions that were thus detected are included in the stars training set.

The final stars training set, including the 1% sparse-sampling of point sources (cleaned of likely quasars) and the H II regions, comes to a total of 429,908 objects.

Note that, unlike for quasars, the colors of stars *do* change appreciably with apparent magnitude—largely as a result of changing metallicity. As the fainter stars tend to be somewhat bluer, one expects a higher degree of stellar contamination with fainter catalog magnitudes. This effect will be even more important to account for in any future attempts at a deeper quasar catalog (even considering deeper photometry with reduced photometric errors). See Figure 3 in Jiang et al. (2006) for an illustration of how stellar colors change as a function of magnitude in SDSS color space.

3.3. The Test Set

The test set is simply the same data set as used for the initial stars training set, but without the random sampling to 1%. As described in Section 2, we limit the sample to point sources that are considered to be reliable and have $14.5 < i < 21.3$. The test set for Paper I was selected in the g-band as it was meant to be a UVX catalog. Here, we switch to i, consistent with the SDSS spectroscopic quasar sample, in order to minimize the effects of the Ly-α forest at high redshift. The full test set includes 44,449,609 objects to be classified.

3.4. Fast Kernel Density Estimation

Once the training and test sets are defined, we compute the likelihood of each object x in the test set with respect to each training set (or, equivalently, the density at x under the stars and under the quasars), using the nonparametric (i.e., distribution-free) *kernel density estimator* (Silverman 1986):

$$\hat{p}(x) = \frac{1}{N} \sum_i^N K_h(\|x - x_i\|), \qquad (2)$$

where N is the number of training set data points, $K_h(z)$ is called the kernel function and satisfies $\int_{-\infty}^{\infty} K_h(z)\,dz = 1$, h is a scaling factor called the bandwidth, and z is the "distance" between a point in the test set and a point in the training set (in our case, these distances are four-dimensional Euclidean color differences, $\|x - x_i\|$). Initial classification uses an Epanechnikov (truncated Gaussian) kernel, which improves the classification speed (as a result of a lack of infinitely long tails) without any loss of robustness in terms of binary classification.

Formally, this process is an N^2 one. Thus, the tractability of our approach relies on the use of space-partitioning trees (e.g., Gray & Moore 2003) and the fact that we only require binary classification. As a result, it is not necessary to explicitly compute the density under each of the training sets; rather we are satisfied with knowing only the upper and lower bounds on the density for each class. The code stops when the bounds no longer overlap. Nevertheless, the algorithm is exact, i.e., it computes the classification labels as if the kernel density estimates had been computed exactly. Full details of the algorithm are given by Gray & Moore (2003), Gray & Riegel (2006), and Riegel et al. (2008).

One improvement over the algorithm used in Paper I is the implementation of code to aid in the (fast) determination of the optimal bandwidth for classification. Finding the optimal KDE bandwidth is similar to the choice of bin size when constructing a histogram. Bins that are too large cause information to be lost. Bins that are too small result in unphysically large small number statistical fluctuations. An initial broad search of possible bandwidths is first attempted. Then a narrower search around the most optimal bandwidth is executed. The criteria used for best bandwidth was the completeness of the quasar training set under self-classification. Efficiency or the product of efficiency and completeness are also viable choices. The final bandwidths were 0.11 mag for stars and 0.12 mag for quasars,

which resulted in an accuracy (completeness) of 92.6% for the quasar training set.

3.5. Priors and Secondary Classification

The algorithm used for Paper I used a flat prior (i.e., a prior that was not a function of magnitude, spatial location, etc.). However, the probability of a given point source being a star is a function of various parameters that are measured by the SDSS photometric pipeline and are included in the database. For example, the probability of an object being a star decreases with fainter magnitudes (since the Galaxy has a finite size) and with increasing Galactic latitude (since the stellar density is higher in the plane of the Galaxy). Thus, we have included the ability in the new algorithm for assigning a parameter-dependent prior. However, in the end, we have not implemented this capability, essentially because the complicated priors we analyzed only provided very modest improvements in the classification. For example, the stellar prior is already 0.95; making the prior a function of Galactic latitude only spreads out the prior over a small range of values and has relatively little effect.

That said, we recognize the value of added information in the catalog beyond the initial binary classification. We, therefore, include other pieces of classification information that can be used to cull interlopers from the catalog and/or to select particular regions of parameter space for further consideration.

Our initial classification used a stellar prior of 0.95 (i.e., ~95% of objects in the test set are expected to be stars). These objects are flagged in the catalog with qsots = 1 (see Section 4). We have also classified all of the objects in the test set after restricting the quasar training set to three narrower redshift ranges (moving the quasars outside of these ranges to the "stars" training set). We classified objects as low redshift ($z \leqslant 2.2$), mid redshift ($2.2 < z < 3.5$), and high redshift ($z \geqslant 3.5$). The rationale for this process is that the distribution of quasar colors considerably changes with redshift, sometimes being more consistent with the stellar locus than others. Thus, subclassification by redshift can improve the robustness of the sample. The priors for these subsamples were set to a somewhat more conservative value of 0.98 rather than 0.95. The bandwidth optimizing algorithm was also rerun for these subclassifications, and the paired (star, quasar) bandwidth values were (0.16, 0.13), (0.12, 0.12), and (0.185, 0.195) for low-z, mid-z, and high-z, respectively, as compared to (0.11, 0.12) for the full sample. Small changes (of order of the range quoted here) in these values would have relatively little impact on our results. The redshift-dependent selected entries in the catalog are flagged as lowzts = 1, midzts = 1, and highzts = 1, respectively.

In addition, for backwards compatibility with the catalog from Paper I (and our unpublished DR3 and DR4 catalogs), we have also provided a flag that indicates whether each object would be selected by that algorithm as well. See Paper I for more details on this selection. These entries in the catalog are flagged as uvxts = 1.

In the end, we catalog all 1,172,157 objects that were classified as quasars by one or more of the above five methods (all redshifts, $p = 0.95$; low-redshift, mid-redshift, high-redshift $p = 0.98$; UVX, $p = 0.88$). This number is 2.6% of the objects in the test set—roughly consistent with the stellar priors of 95–98% and amounting to nearly 140 quasar candidates per square degree. Paper I had a density of only ~48 quasar candidates per square degree over 2099 deg^2. Most of this increase

comes from the deeper i-band cut (21.3 instead of 21.0) and the move from g to i itself as our i-band limit of 21.3 roughly corresponds to $g = 21.55$. The remainder comes from the additional redshift coverage and from contamination (which we will explore how to minimize in Section 4.2).

Finally, as in Paper I, in addition to nonparametric classification, we also provide the parametric quasar and star densities (likelihoods). As discussed above, these values are intractable to determine for the entire test of more than 44 million objects. However, for the smaller sample of objects classified as quasars using any of the above five criteria, it is possible to determine the exact values in addition to the binary classification. In Paper I, we showed how this information can be used to clean the quasar candidate list of the most obvious sources of contamination; also see Section 4.2.

4. THE QUASAR CATALOG

After applying our algorithm to the test set as described above, we are left with 1,172,157 quasar candidates that define this catalog. The next sections describe the efficiency and completeness of the catalog in addition to prescriptions for making more robust subsets of the whole catalog. Table 1 lists the most robust quasar candidates, while Table 2 provides a description of each column in the machine-readable table. Table 3 is a listing of objects that were culled (see Section 4.2) from Table 1 as known or likely contaminants, but are included as a separate table for the sake of completeness. Table 3 has the same format as Table 1.

4.1. Known Quasar Cross-Matching

Each object in the catalog was cross-matched to the DR5 quasar catalog (Schneider et al. 2007), the 2QZ quasar catalog (Croom et al. 2004), the SDSS-2dF LRG and QSO Survey (2SLAQ) Early Data Release quasar catalog (Croom et al. 2008), and the SDSS-DR6 spectroscopic database (Adelman-McCarthy et al. 2008). The matching was done in the above order. Once a match was found, no further matches were allowed for that object as this hierarchy represents the most effective path to robust identifications. Objects from the DR6 spectroscopic database were required to have a high confidence zStatus flag.

In all, 88,879 spectroscopically-confirmed quasars, 4962 stars, and 891 "other" objects (e.g., normal and narrow emission line galaxies) were identified. As such, our *photometric* quasar catalog is also one of the largest single catalogs of *spectroscopically confirmed* quasars to date even though we only include known quasars from three sources. However, it is clearly spatially (and otherwise) biased to locations (and reasons) where follow-up spectroscopic surveys have been carried out. While ~16,000 of these have not been vetted by eye as is done for the SDSS spectroscopic quasar catalogs (Schneider et al. 2007), we have only included those objects that pass relatively robust flag checking diagnostics. Comparison with the heterogeneous catalog of Véron-Cetty & Véron (2006), which generally includes automatically identified quasars from the SDSS database rather than the more carefully vetted sample from Schneider et al. (2005), suggests that most of these objects should be robust. Of the 36,948 quasars in Véron-Cetty & Véron (2006) that were taken directly from the SDSS database, 85 were not included in Schneider et al. (2005) and 43 had redshifts corrected by Schneider et al. (2005). Among the redshift errors is SDSS J205644.53−005904.2, which is listed by Véron-Cetty & Véron (2006) as a $z = 5.989$ quasar (though the SDSS database

Table 1
NBCKDE Quasar Candidate Catalog

Number (1)	Name (SDSS J) (2)	R.A. (deg) (3)	Decl. (deg) (4)	Obj ID (5)	z_{phot} (6)	z_{low} (7)	z_{high} (8)	z_{prob} (9)	u (10)	g (11)	r (12)	i (13)
1	000000.70+160540.6	0.0029420	16.0946121	587727223561060668	2.685	2.180	2.890	0.402	22.734	22.068	21.706	21.296
2	000000.98+144518.1	0.0041090	14.7550374	587727221950382615	2.115	1.660	2.220	0.546	21.128	20.951	21.004	20.788
3	000001.10+011037.1	0.0045944	1.1769856	587731187814498541	0.825	0.670	1.040	0.602	20.911	20.863	20.919	21.185
4	000001.38−010852.2	0.0057816	−1.1478427	588015507658768592	2.225	2.130	2.650	0.299	21.584	21.180	20.787	20.702
6	000001.88−094652.0	0.0078461	−9.7811385	587727179523227759	0.975	0.770	1.420	0.921	19.563	19.396	19.232	19.312

(This table is available in its entirety in machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

Table 2
NBC Quasar Candidate Catalog Format

Column	Format	Description
1	I7	Unique catalog number
2	A18	Name: SDSS J$hhmmss.ss + ddmmss.s$ (J2000.0)
3	F12.7	Right ascension in decimal degrees (J2000.0)
4	F11.7	Declination in decimal degrees (J2000.0)
5	A19	SDSS Object ID
6	F7.3	zphot; photometric redshift (see Weinstein et al. 2004)
7	F6.3	Lower limit of the photometric redshift range
8	F6.3	Upper limit of the photometric redshift range
9	F6.3	zphotprob; photometric redshift range probability
10	F7.3	u PSF übercalibrated asinh magnitude (corrected for Galactic extinction)
11	F6.3	g PSF übercalibrated asinh magnitude (corrected for Galactic extinction)
12	F6.3	r PSF übercalibrated asinh magnitude (corrected for Galactic extinction)
13	F6.3	i PSF übercalibrated asinh magnitude (corrected for Galactic extinction)
14	F6.3	z PSF übercalibrated asinh magnitude (corrected for Galactic extinction)
15	F6.3	Error in PSF u asinh magnitude
16	F5.3	Error in PSF g asinh magnitude
17	F5.3	Error in PSF r asinh magnitude
18	F5.3	Error in PSF i asinh magnitude
19	F5.3	Error in PSF z asinh magnitude
20	F7.3	$E(B - V)$ (mag); $A_u/A_g/A_r/A_i/A_z = 5.155/3.793/2.751/2.086/1.479 \times E(B - V)$
21	F7.3	c; concentration (=PSFMag_i−modelMag_i) for star/galaxy separation
22	F8.2	radio; 20 cm flux density (mJy) (−1 for not detected or not covered)
23	F7.4	xray; RASS full-band count rate (−9 for not detected or not covered)
24	F8.2	pm; proper motion (mas/year)
25	I2	moved; an addition flag to indicate possible moving objects (=1 if moving)
26	I1	qsots; selection Flag; full redshift range, 95% star prior
27	I1	lowzts; selection Flag; low-redshift range ($z \leqslant 2.2$), 98% star prior
28	I1	midzts; selection Flag; mid-redshift range ($2.2 < z < 3.5$), 98% star prior
29	I1	highzts; selection Flag; high-redshift range ($z \geqslant 3.5$), 98% star prior
30	I1	uvxts; selection Flag; UV-excess, 88% star prior (see Paper I)
31	F9.3	qsodens; log KDE quasar density
32	F8.3	stardens; log KDE star density
33	I1	good; quality flag (6 = most robust; −6 = least robust)
34	A16	Previous catalog object classification
35	F6.3	Previous catalog object redshift

Table 3
Rejected Quasar Candidates

Number (1)	Name (SDSS J) (2)	R.A. (deg) (3)	Decl. (deg) (4)	Obj ID (5)	z_{phot} (6)	z_{low} (7)	z_{high} (8)	z_{prob} (9)	u (10)	g (11)	r (12)	i (13)
5	000001.81+141150.5	0.0075587	14.1973842	587730773351858843	3.495	3.180	4.320	0.885	25.335	21.597	20.502	20.503
10	000002.27−085640.9	0.0094825	−8.9447047	587727180596969488	3.515	3.220	4.470	0.814	25.037	21.031	20.103	19.876
12	000003.67−095452.9	0.0153217	−9.9146988	587727179523228066	3.135	2.910	3.360	0.206	24.054	21.485	21.211	20.984
13	000003.73−003705.5	0.0155724	−0.6182073	587731185667080833	4.615	4.190	4.830	0.317	24.677	23.928	22.203	20.931
24	000006.00−085014.3	0.0250328	−8.8373328	587727227837612402	2.875	2.680	3.010	0.141	22.910	21.657	21.451	21.146

(This table is available in its entirety in machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

has a warning flag), but is cataloged by Trump et al. (2006) as a $z = 2.48$ iron-dominated, low-ionization, broad absorption-line quasar. On the other hand, there are, in fact, objects in our catalog classified as nonquasars that are actually quasars. For example, most of the objects with $z > 1$ and marked in the catalog as "DR6_GALAXY" are indeed quasars for which the spectroscopic classification templates failed for some reason; such objects are recovered during the careful review process used to construct the published spectroscopic sample of SDSS quasars (Schneider et al. 2007). However, we maintain their galaxy classifications here since complete double-checking of the SDSS's automated identifications is better left for the careful construction of the next installment in the SDSS's spectroscopic quasar catalog series.

4.2. Culling

For Paper I, after running the "NBC-KDE" algorithm, we made an additional cut on the stellar density to remove the most likely contaminants. For this version of the catalog, we have chosen instead to tabulate all of the objects that passed the NBC criterion and flag the sample of the most likely contaminants after the fact.

The table includes a parameter "good," which is meant to be indicative of how likely we feel that the object is truly a quasar. This column is an integer value that spans the range $[-6, 6]$. More positive values indicate greater confidence in the quasar classification, and we generally recommend using objects with good $\geqslant 0$ for statistical analysis (with the possible exception of mid- and high-z candidates, see below). As such, objects with good < 0 and/or that are known contaminants have been removed from Table 1 and are included separately in Table 3.

The value of good starts at 0 for each object. It is incremented by 2 if the object is a spectroscopically-confirmed quasar. It is decremented by 2 if it is a known nonquasar. The following conditions cause the good flag to be incremented by 1 (see Table 2 for an explanation of the parameters):

1. qsodens > 1.0
2. radio > 0 (i.e., radio-detected)
3. xray > 0 (i.e., X-ray-detected)
4. (lowzts > 0 || uvxts > 0) && zphot < 2.25 && zphotprob > 0.5 (i.e., consistent photo-z and class)
5. midzts > 0 && zphot > 2.15 && zphot < 3.5 && zphotprob > 0.75 (i.e., consistent photo-z and class)

Note that there is no criterion for consistent photo-z and class for high-z candidates as the contaminants generally have "correct" photo-z's.

The following conditions cause the good flag to be decremented by 1:

1. pm > 20.0 || ($i < 18$ && pm > 10.0) (high proper motion)
2. moved $= 1$ (likely moving source)
3. $E(B-V) > 0.1438$ (*i*-band reddening more than 0.3 mag)
4. uvxts $= 1$ && lowzts $+$ midzts $+$ highzts $= 0$ && ($\sigma_{ug} > 0.25$ || zphot > 3.6 && zphotprob > 0.8)) (UVX-selected object that otherwise appears high-z)
5. (lowzts $= 1$ || midzts $= 1$ || highzts $= 1$) && qsodens < -1.3 (quasar likelihood too low)
6. midzts $= 1$ && qsots $+$ lowzts $+$ highzts $+$ uvxts $= 0$ && zphot > 2.90 && zphot < 2.91 (likely mid-z interlopers)

7. (highzts $= 1$ && $\sigma_r > 0.15$) || ((midzts $= 1$ || highzts $= 1$) && $\sigma_i > 0.25$) (drop-out objects with an insufficient signal-to-noise ratio (S/N))
8. $i < 17$ && $u - g > 1.0$ && midzts $= 1$ && qsots $= 0$ (bright mid-z interlopers)
9. $i < 17$ && $u - g > 1.0$ && highzts $= 0$ && (qsots $= 0$ || $g - r > 1.0$) (bright high-z interlopers)
10. $b < 18$ (Galactic latitude (not given in tables) too low)

Note that we have also capped the photometric redshift probability (see Section 4.6) at 0.499 for objects that are likely to be extended, yet have redshifts inconsistent with an extended morphology (specifically, $c > 0.1$ && zphot > 0.8 && zphotprob $\geqslant 0.5$), and that are high-z candidates but are not u-band dropouts (zphot > 3.6) or g-band dropouts (zphot > 4.5). These modified values come into play for some of the above criteria.

In the end, there are 80,404, 136,232, 292,800, 505,646, 129,246, 19,632, and 8197 with good flags of greater than 2, 2, 1, 0, −1, −1, and less than −2, respectively. The maximum and minimum values are 6 and −6, respectively. Known quasars and nonquasars are not set to the extreme values so that their relative quasar likelihood in the absence of spectroscopic confirmation can be used to assess the relative likelihood of unknown objects.

4.3. Properties

Figure 2 shows the magnitude distributions of the catalog. Known interlopers are included, in part, to show their effect on the distribution at the bright end. The *i*-band distribution is thus given with (solid black) and without (dashed black) cuts on the good parameter. The $i < 21.3$ limit is not sharp as objects with $i < 21.3$, either before or after über-calibration were included. The colored histograms indicate the magnitude distributions in the other bands as this is important for assessing the color completeness of the catalog at the faint end. Note, however, that SDSS's use of asinh magnitudes (Lupton et al. 1999) means that there is no hard magnitude limit and that all objects detected to our chosen *i*-band limit will have meaningful measurements in the other four bands.

The spatial distribution of the catalog is given in Figure 3. As one generally expects more quasars at higher Galactic latitude as a result of lower dust (Schlegel et al. 1998) and fewer Galactic stars blocking the light from distant sources, we show the distribution of sources as a function of Galactic latitude in Figure 4. At low Galactic latitudes, stars masquerading as quasars in our catalog show a spike in the distribution due to the increase in stellar density toward the Galactic plane; thus, in Section 4.2, we decremented the good flag for the lowest Galactic latitude objects in our sample.

While these quasars have their photometry corrected for Galactic extinction according to the Schlegel et al. (1998) prescription, one obviously cannot correct undetected objects for extinction. As the limit of our sample is $i < 21.3$ and the 95% completeness limits of SDSS is $i = 21.3$, our catalog will fail to include quasars (for example) with *i*-band extinction, A_i, larger than 0.3 at $i = 21$ (equivalently, $E(B-V) = 0.144$). The distribution of $E(B-V)$ values in our sample is shown in Figure 5. Myers et al. (2006) showed that the selection efficiency of the DR1 catalog was improved by making a more rigorous cut of $A_g < 0.18$ ($A_i < 0.099$; $E(B-V) < 0.0475$). The two cuts are shown in Figure 5 and account for roughly 1% and 20% of the sample, respectively.

The colors of the quasars and stars in the training sets are given by Figure 6, while Figure 7 shows the color distribution
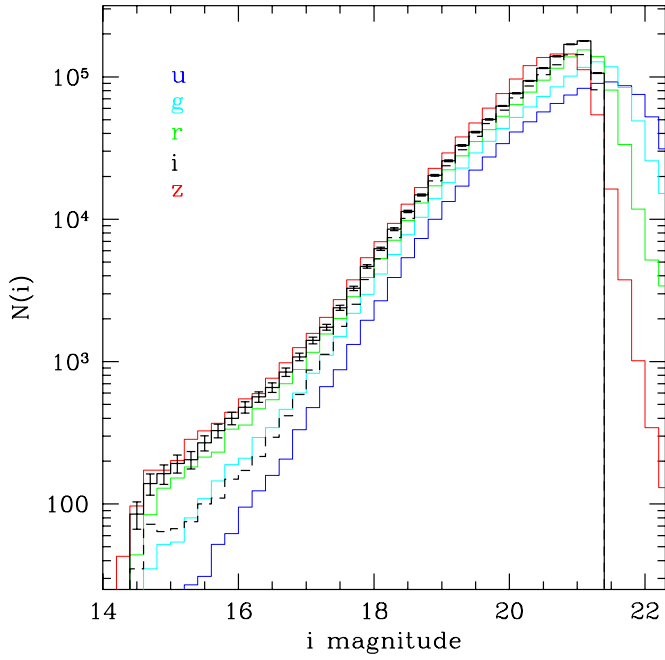
Figure 2. *i*-Band magnitude distribution of the 1,172,157 quasar candidates (i.e., Tables 1 and 3 combined) in the catalog (solid black line). Colors show the magnitude distributions in the other bands to indicate where the relative limits are. The dashed black line is the *i*-band histogram for the most robust sources in the catalog, that is, limited to the good ⩾ 0 objects in Table 1.

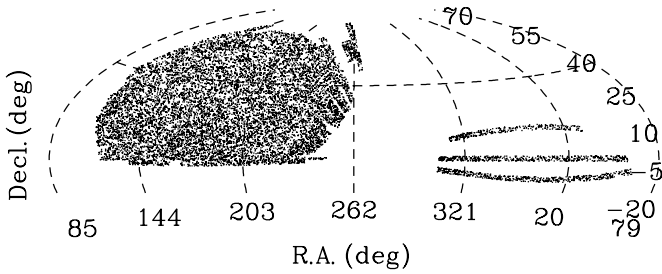(A color version of this figure is available in the online journal.)

Figure 3. Spatial distribution of quasar candidates in an Aitoff projection. For the sake of clarity, only one in every 100 objects is shown.

of test set objects that were classified as quasars (i.e., the objects in this catalog). By comparing the location of likely interlopers (magenta) in Figure 7 with the relative location of stars/quasars in the training sets from Figure 6, it is possible to identify the most likely contaminants in the catalog.

In Paper I, we explicitly culled objects with star probability in excess of 0.01. For this sample, no such cut is applied (with the exception of the initial selection of UVX candidates using the same algorithm as in Paper I). However, it may be useful for additional culling to know the distribution of star and quasar probabilities. Thus, we show them in Figure 8 for the entire sample, and broken down by the redshift-selected subsamples. Examination of this figure can help determine optimal cuts for statistical subsamples. For example, a very robust subsample could be made by making a cut requiring a high value for QSO density, but Figure 8 shows that this comes with the trade-off of cutting most mid- and high-*z* quasars in addition to some of the UVX sources.
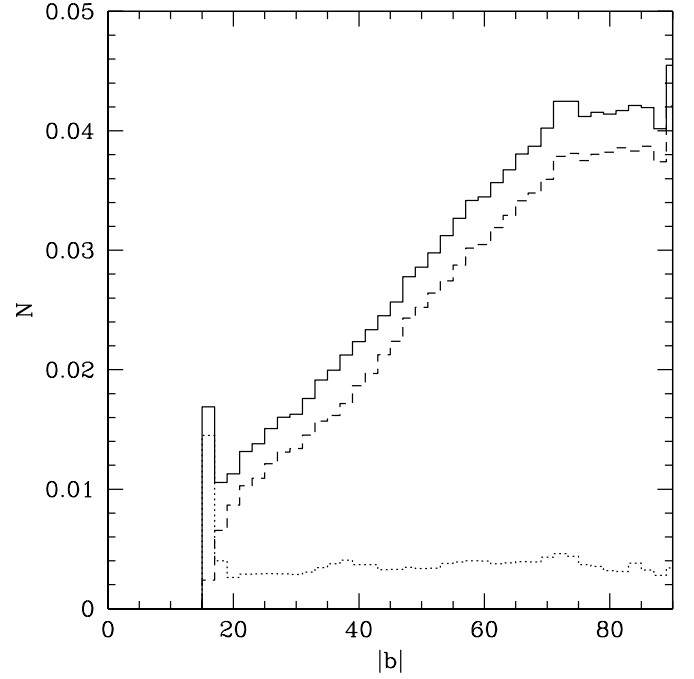
Figure 4. Ratio of quasar candidates in the catalog to all point sources as a function of Galactic latitude (*b*). The full sample (solid line); the most likely quasars, having good ⩾ 0 (dashed); and the least likely quasars, having good < 0 (dotted) are plotted. The sharp increase at the lowest *b* values is indicative of increased stellar contamination near the Galactic plane.
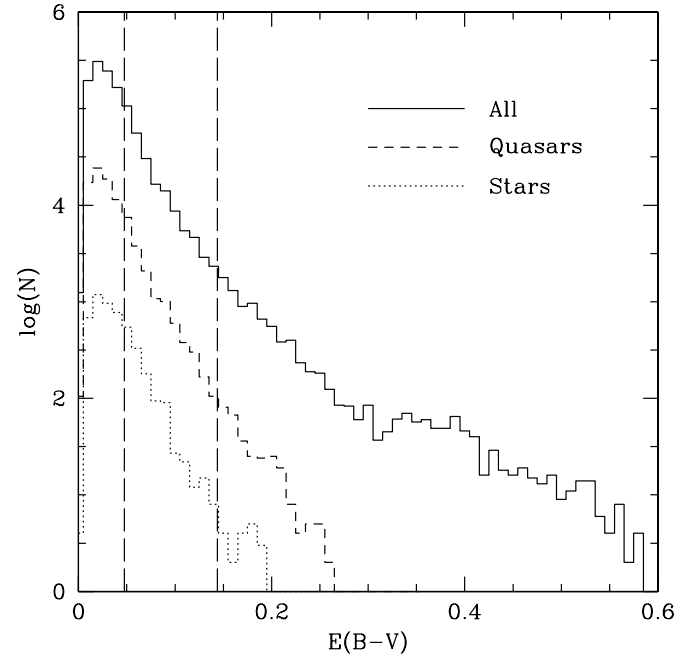
Figure 5. $E(B-V)$ distribtion. The top (solid) histogram represents the whole sample. The middle (dashed) histogram is for spectroscopically-confirmed quasars in the sample. The bottom (dotted) histogram shows spectroscopically-confirmed stars. The long dashed vertical lines indicate the $A_i < 0.3$ and $A_i < 0.099$ completeness limits.

4.4. Completeness

It is difficult to quantify the completeness of the catalog since it extends to deeper magnitudes and higher redshifts than most existing spectroscopic quasar catalogs. Yet, we can do some simple tests to get an idea of the completeness. We
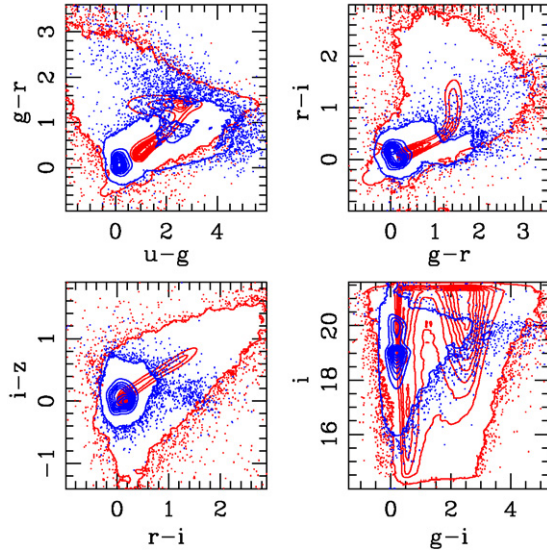
Figure 6. Color–color and color–magnitude distribution of objects in the training sets. Quasars are given in blue (75,382 objects). "Stars" are given in red (429,908 objects). The (linear) contour levels are relative to the peak in each sample.

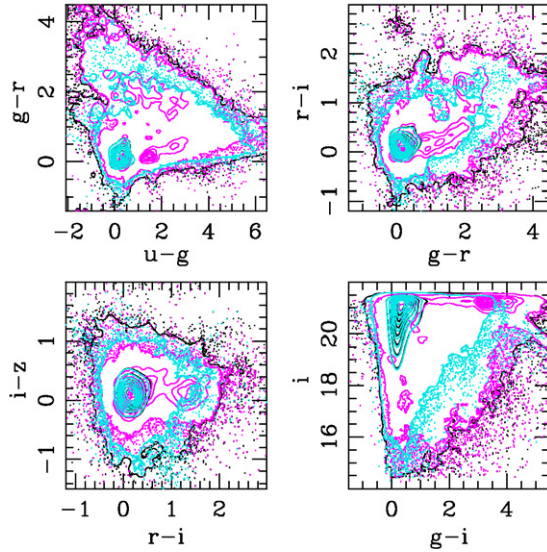(A color version of this figure is available in the online journal.)

Figure 7. Color–color and color–magnitude distribution of all quasar candidates in the catalog (black). Cyan contours indicate the most likely quasars good $\geqslant 0$, while magenta contours represent the most likely interlopers good $\leqslant -2$.

(A color version of this figure is available in the online journal.)

Figure 8. Distribution of KDE star and quasar probability densities for all objects classified as quasars by one or more of the NBC methods. Black points and contours give the full sample (repeated in each panel). Low-z quasars are shown in blue, UVX in cyan, mid-z in green, and high-z in red. Note that the NBC selection by definition rejects objects with star probability greater than quasar probability, but the KDE values were determined for objects selected by any of the NBC methods, not only the overall NBC selection, so some objects appear above the diagonal.

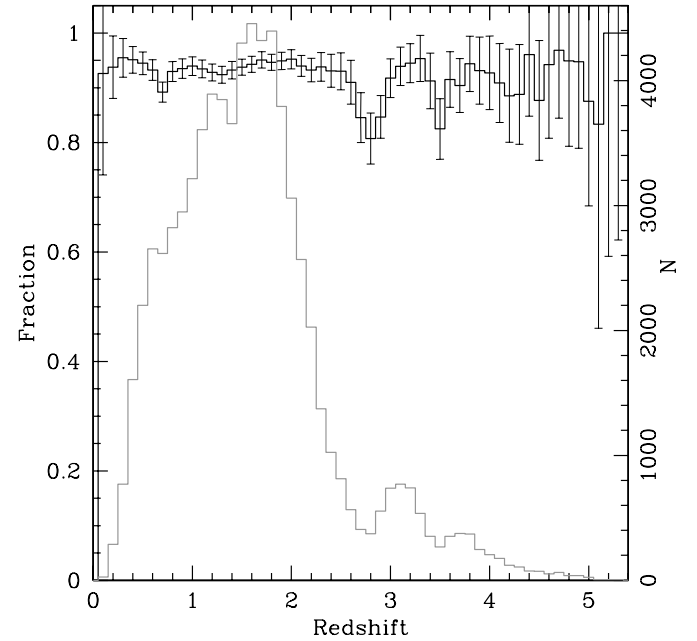(A color version of this figure is available in the online journal.)

Figure 9. Fraction of training set quasars recovered as a function of magnitude. The overall recovered fraction (completeness) is 93.4%. Somewhat higher levels of incompleteness are found at $z \sim 2.8$ and $z \sim 3.5$, where it is particularly difficult to cleanly separate stars from quasars in SDSS color space. The gray histogram and right-hand axis give the redshift distribution of the quasar training set.

first compare to the SDSS-DR5 quasar catalog. While this sample is the basis of our quasar training set, it is instructive to explore the completeness of this sample to see if there are any redshift regions where the selection algorithm is particularly incomplete. Of the 77,429 quasars in the SDSS-DR5 catalog, 73,924 of these are point sources with $i < 21.3$, thus meeting our initial selection requirements. Our algorithm recovers 69,031 of these for an overall completeness of 93.4%. Note that the true completeness to $z \lesssim 1$ quasars will be lower as a result of our point source requirement.

Figure 9 shows the completeness distribution as a function of redshift. The gray histogram and right-hand axis give the redshift distribution of the input sample. Note the relatively incomplete regions near $z \sim 2.8$ and $z \sim 3.5$ in both the input and output

samples, respectively. These occur where quasars and stars have very similar colors in SDSS color space and quasars are difficult to separate cleanly. For these regions, the completeness is not well constrained given that the quasar training set was

initially incomplete in these regions. It is not clear whether the photometric catalog completeness is likely to be higher or lower; however, the construction of the training sets is such that the completeness is hoped to be higher than for the main SDSS quasar sample. An additional region with a slightly lower completeness is found near $z \sim 0.675$, where white dwarfs are a source of contamination.

It must be emphasized that our catalog is limited to optically-selected type 1 quasars. This is primarily a limitation due to the nature of the SDSS data rather than to our actual technique. Other methods/datasets, including radio, IR, and X-ray, can and do find quasars (and less-luminous active galactic nuclei, AGNs) that will not be found by our method/data, particularly type 2 quasars (e.g., Lacy et al. 2004; Treister et al. 2004; Martínez-Sansigre et al. 2006). The completeness numbers herein do not consider such objects even though the size of the obscured population is substantial (e.g., Polletta et al. 2008).

Another source of incompleteness is due to extra-Galactic reddening (whether by the AGNs' dusty torus, the host galaxy, or another galaxy along the line of sight). Richards et al. (2003) estimated that the fraction of quasars reddened out of the optically-selected SDSS sample (but still detected as broad-line quasars) is $\sim 15\%$, whereas some radio and near-IR selected samples (e.g., Glikman et al. 2007) argue for up to $\sim 60\%$ incompleteness of optically-selected samples (albeit with small number statistics). Recent work by Maddox et al. (2008) estimated the fraction as 30% based on a K-band selected sample. Thus, we expect that our i-band selected sample will be incomplete at a comparable level due to dust extinction that occurs outside of the Milky Way.

A more detailed analysis of the effects of dust extinction is beyond the scope of this paper; however, for guidance, we refer the reader to Ménard et al. (2008). While that paper specifically discusses the effects of dust from intervening galaxies, the conclusions regarding completeness at a given $E(B-V)$ are generic. In short, the majority of quasars are expected to be recovered at $E(B-V) = 0.1$, but we expect neglible completeness above $E(B-V) = 0.4$. Further empirical assessment of the completeness of our catalog will come from current and future spectroscopic samples that were selected with complementary selection methods. For example, the catalog includes the NOAO Deep Wide-Field Survey (NDWFS; Jannuzi & Dey 1999) area, which includes extensive spectroscopic coverage from the AGN and Galaxy Evolution Survey (AGES; e.g., Cool 2006) that will be suitable for such analysis once the AGES data are published.

As a simple check on our completeness versus nonoptical quasar selection, we cross-match the multiwavelength-selected spectroscopic sample (Trump et al. 2007) from the COSMOS (Scoville et al. 2007) field with our photometric sample. We find 45 matches to within $1''$; most of these are indeed type 1 (broad-line) quasars. In all, the Trump et al. (2007) sample includes 47 type 1 objects with $i < 21.0$, which, in principle, should have been recovered by our algorithm (allowing for a slightly brighter magnitude limit to mitigate any differences in the magnitudes used). We recover 33 of these 47 (70%). Six of the missing objects have $z < 0.7$, which we preferentially select against due to the point source nature of our catalog. Three have $2.5 < z < 3.0$, where optical selection is notoriously inefficient. This leaves three objects at $z \sim 1$ and two objects at $z \sim 2$ that we might have otherwise expected to find. We find that three of these are rejected due to our strict photometric flag cuts as described above, while the remaining two are likely lost because of dust reddening.
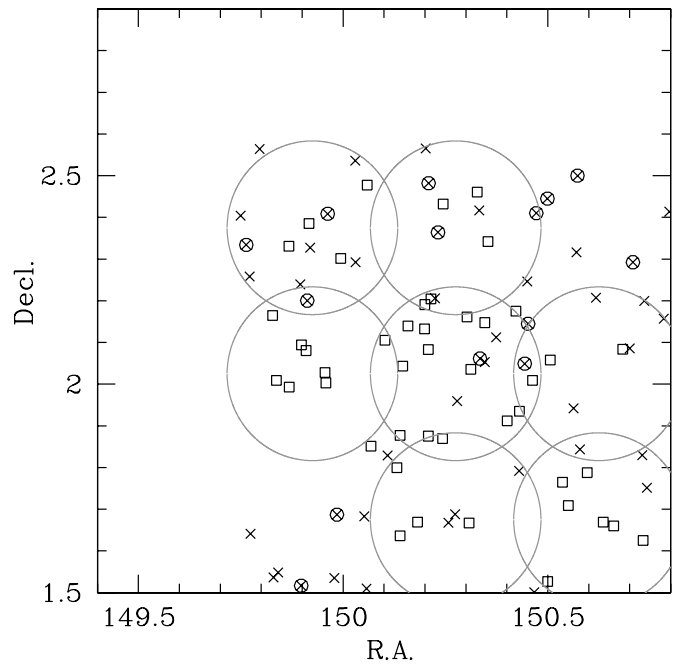

Figure 10. Type 1 quasars in the COSMOS field. Open squares indicate objects that were spectroscopically confirmed by Trump et al. (2007) and are matched to objects in our photometric catalog. Large circles roughly indicate the area of maximal coverage by Trump et al. (2007). Crosses denote 51 photometric quasar candidates that were not cataloged by Trump et al. (2007). The 14 most robust (good $\geqslant 1$ in this case) of these 51 candidates are additionally circled. Roughly, half are in regions covered by Trump et al. (2007) and, in principle, should have been found. Three of these are not in the COSMOS X-ray catalogs (Hasinger et al. 2007) and may be X-ray and radio-weak broad absorption line quasars.

However, our catalog also includes 51 previously unconfirmed objects in the COSMOS field that were not cataloged by Trump et al. (2007); out of these, we consider 14 to be particularly robust candidates (good $\geqslant 1$). Figure 10 shows the distribution of these sources in comparison with the coverage of Trump et al. (2007). Some of these objects may be among those to which the Trump et al. (2007) investigation is incomplete ($\sim 10\%$ at $i < 22$ and $\sim 25\%$ of the X-ray targets, whether due to tiling collisions or low S/N spectra). Even by considering this incompleteness, many of these 14 candidates should have been recovered. Three have no match within $3''$ in the COSMOS X-ray catalog (Hasinger et al. 2007) and may be broad absorption line quasars (BALQSOs) given that BALQSOs are known to be X-ray weak (Green et al. 2001; Gallagher et al. 2002) and are generally not strong radio sources (Stocke et al. 1992), and thus are the most likely type 1 quasars to be missed by Trump et al. (2007). These missing objects serve to illustrate the importance of combining multiple selection methods when attempting a truly complete AGN census. Matching the full set of 51 objects to the catalog of Hasinger et al. (2007) reveals 22 objects with X-ray matches to within $2''$, which suggests that no less than 43% of the 51 previously unconfirmed/uncataloged candidates are indeed quasars.

As our primary science motivations for this work thus far have largely been statistical analysis of clustering, our emphasis has been on creating clean samples of photometric quasars as opposed to a complete sample. Thus, we have not considered the completeness of the sample in more detail here. As such we caution that some investigations, such as a full bolometric quasar luminosity function, will require more detailed understanding
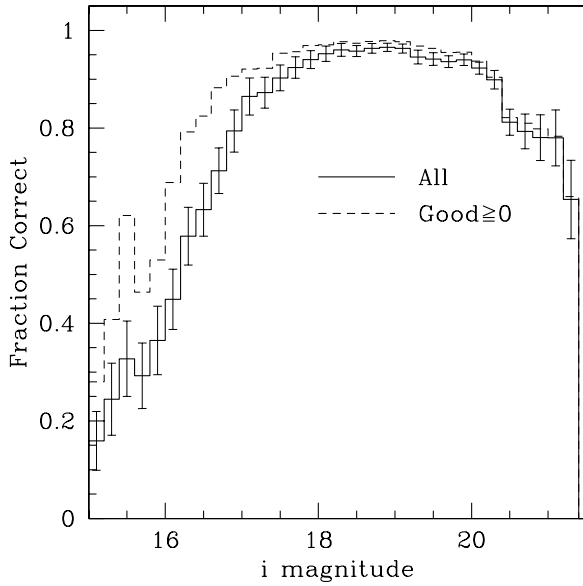
Figure 11. Efficiency as a function of magnitude. The dashed line gives the efficiency for those quasar candidates that we consider most robust (good $\geqslant$ 0). While the efficiency is low at the bright end, so are the absolute numbers of objects (see Figure 2); thus, the overall contamination from bright objects is relatively small.

Table 4
Estimated Catalog Efficiency

Sample	Overall Efficiency	good >= 0 Efficiency
All	71.5 ± 3.5	79.5 ± 2.6
UVX		96.4 ± 1.4
Low-z		91.7 ± 1.3
UVX $\|\|\|$ Low-z		92.7 ± 1.7
UVX && Low-z		96.3 ± 1.2
Mid-z		46.4 ± 5.8
High-z		40.1 ± 7.9

of the completeness of this sample with respect to both dust-reddened sources and completely optically obscured (type 2) sources.

4.5. Efficiency

A naive test of the efficiency of the algorithm is simply to determine the fraction of known quasars amongst the total sample of known objects. This value is 88879/(88879 + 4962 + 891) = 93.8%. Considering only sources with good $\geqslant$ 0, the expected efficiency based on known objects is 95.6%.

We can also compute the efficiency as a function of magnitude. This is shown in Figure 11 for both the full sample and the good $\geqslant$ 0 candidates. The efficiency measured in this manner exceeds 95% for $17 < i < 20.4$ objects that are flagged as "good." At bright magnitudes, the efficiency drops off due to interlopers such as white dwarfs and faint low-metallicity F-stars (e.g., compare Figures 3 and 4 in Ivezić et al. 2007) in addition to mid- and high-z interlopers. The latter can be seen in Figure 7 at $u - g \sim 1.5$ and $g - i \sim 1.5$ (also see Section 5). Overall, this population is small, but is relatively larger for $i < 17$ where the number of real quasars is also small. Restricting the sample to good $\geqslant$ 0 removes some but not all of the contamination. However, there are relatively few bright objects in the catalog, so this contamination has little effect on the catalog as a whole. At the faint end, the efficiency is also lower, here largely due to increasing photometric errors. Convolving our estimate of the efficiency as a function of magnitude with the magnitude distribution shown in Figure 2 results in an expected number of bona fide quasars in the catalog between 850,000 and 990,000.

Furthermore, as shown by Myers et al. (2006), it is possible to use the auto-correlation of the photometric quasar sample to estimate its efficiency since angular scales that are large by clustering standards correspond to relatively small physical scales at Galactic distances, and stars will have a residual clustering signal. As this method is independent of any biases in previous spectroscopic identifications, it is expected to be

more robust than our above crude estimates. Table 4 shows the efficiencies that result for this clustering analysis (at a size scale of 5°) for the whole catalog and various subsamples. The overall efficiency of the catalog is only expected to be ~72%. However, it is nearly 97% for certain subclasses of objects. Users of the catalog should pay particular attention to this table and the flags that are represented when attempting to do any sort of statistical analysis that is sensitive to interlopers.

4.5.1. Star–Galaxy Separation

One caveat with regard to the above efficiency estimates has to do with SDSS star–galaxy separation. Technically, the clustering-based efficiency estimates from Table 4 should not be viewed as the *quasar efficiency* but should rather tell us the rate of *stellar contamination*. As galaxies cluster more like quasars than stars, we must be aware that the clustering results will not uncover non-AGN galaxy interlopers.

In detail, the primary method used by the SDSS pipeline to differentiate between unresolved and resolved sources (i.e., stars and galaxies) is to examine the difference between the PSF magnitudes and so-called model magnitudes (de Vaucouleurs or exponential). For extended sources, like galaxies, PSF magnitudes over-resolve the source and yield fluxes that are smaller (magnitudes that are larger) than for magnitudes, which model the distribution of light better. Thus, it is possible to use the difference between the PSF and model magnitudes to determine the morphology of SDSS sources. Specifically, objects are considered to be extended if psfMag − modelMag > 0.145, where the magnitudes are summed over all bands in which the object is detected (Stoughton et al. 2002).

However, at fainter magnitudes, large photometric errors can make this star–galaxy classification algorithm less effective. In general, the limiting behavior is to classify all faint objects as being stellar. Thus, our catalog of "point sources" will have some degree of contamination from galaxies and this contamination will be a function of magnitude. While it is not possible to make explicit corrections for this contamination, it is possible to estimate the level of its effect as a function of magnitude. We specifically make use of the Bayesian star–galaxy classification algorithm developed by Scranton et al. (2002), which assigned a Bayesian galaxy probability to each object rather than a binary classification.

Figure 12 shows the fraction of SDSS-classified point sources as a function of magnitude that have less than a 10% chance of being galaxies according to the Scranton et al. (2002) method. Values below unity are indicative of the fraction of galaxies that the SDSS has erroneously classified as point sources. At $i \sim 20.2$, the fraction of contamination is only ~5%, but at the limit of our survey, it may be as high as 15%. Thus,
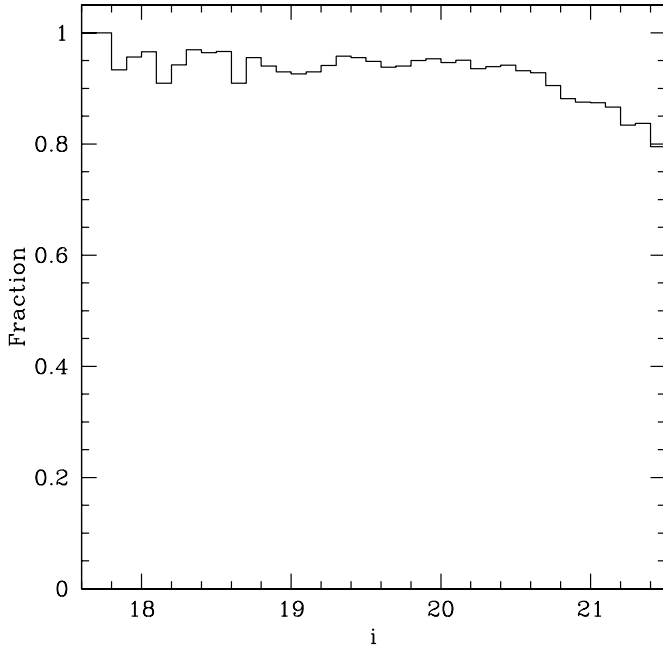
Figure 12. Fraction of objects classified as point sources in single-epoch SDSS photometry that are indeed point sources according to the Bayesian star–galaxy classification algorithm (Scranton et al. 2002). At the limit of our survey, contamination from galaxies may be as high as ∼15%. Brighter than $i \sim 20$, contamination should be lower than ∼5% indicated here, since this plot uses a rather strict cut on galaxy probability that is more appropriate at faint magnitudes than bright.

considerable caution is needed to prevent a significant amount of contamination from galaxies; indeed, much of the contamination at the faint end may arise from galaxies. This issue is particularly important when using the catalog for clustering studies as quasars and galaxies have similar clustering properties.

4.6. Photometric Redshifts

It is possible to estimate redshifts of astrophysical sources using only broadband photometry by identifying the signature of distinct spectral features on the colors of objects. For galaxies, such "photometric" redshifts have a long history (e.g., Connolly et al. 1995, and references therein). Similarly, robust photometric redshift for quasars can be derived for high-redshift quasars where the strong Lyman-α forest decrement produces a relatively sharp change in color. However, robust photometric redshifts for low-z quasars using the smaller broadband color changes induced by emission lines had to wait until the use of many filters (e.g., Wolf et al. 2001) and sensitive photometric calibration over large-area surveys (e.g., Richards et al. 2001; Budavári et al. 2001).

For each object in the catalog, we report photometric redshifts that were determined via the method described in Weinstein et al. (2004). This algorithm minimizes the difference between the measured colors of each object and the median colors of quasars as a function of redshift. We used the colors of all of the unresolved point source quasars in the DR5 quasar catalog of Schneider et al. (2007) as our color-redshift template. For each object, we catalog the most likely photometric redshift (to the nearest 0.01), a redshift range, and the probability that the redshift is within that range; see Weinstein et al. (2004) for more details.

The left panel of Figure 13 shows the spectroscopic versus photometric redshifts of the 88,879 confirmed quasars in the catalog, revealing those redshifts where the algorithm has the largest error rate (either due to degeneracy between distinct redshifts or smearing of nearby redshifts). However, one can see from the highly zero-peaked distribution in the right panel that, overall, the quasar photo-z algorithm performs quite well, with 73,761 (83%) of the redshifts being correct to within ±0.3.

We compare the distribution of photometric and spectroscopic redshifts in Figure 14, which shows that the photo-z's match the spectroscopic redshifts reasonably well in the ensemble average on smoothing scales slightly larger than the photo-z bins, which is important for statistical analysis. Figure 14 also quantifies the fractional accuracy (to $\Delta z \pm 0.3$; gray squares) in each photo-z bin, which was seen more qualitatively in Figure 13. In general, the photo-z accuracy is best where the most training data exist ($1 < z < 2$), which helps explain the 83% overall photo-z accuracy of the catalog. It is lower for $z < 0.5$ in part due to host galaxy contamination, at $z \sim 2.7$ where relatively little training
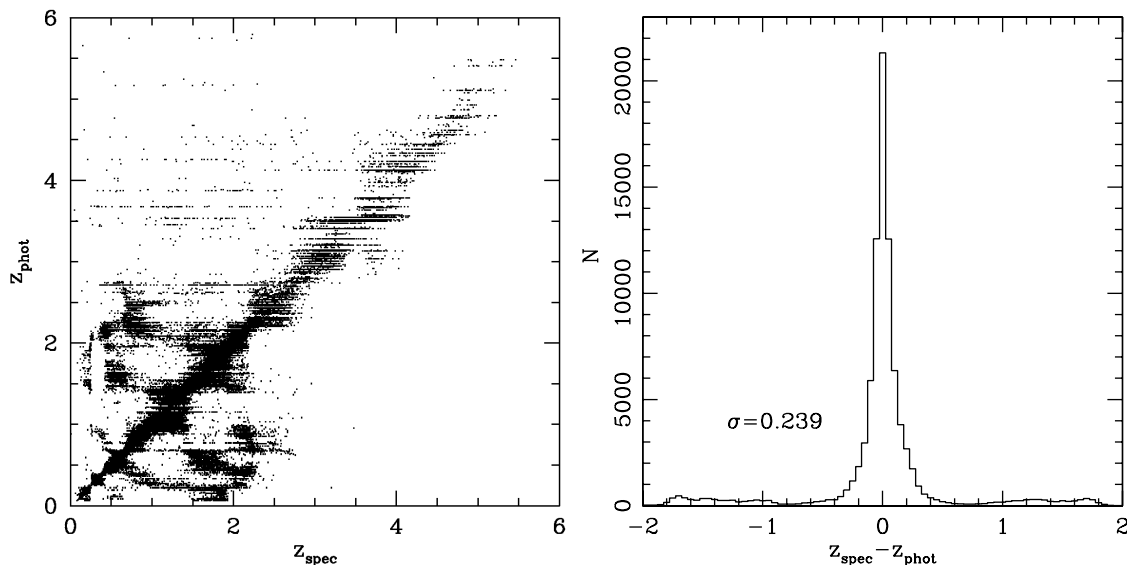

Figure 13. Left: spectroscopic vs. photometric redshifts for all spectroscopically-confirmed quasars in the catalog. Right: histogram of the difference between spectroscopic and photometric redshifts. After rejecting outliers, the width of the distribution is $\sigma = 0.239$.
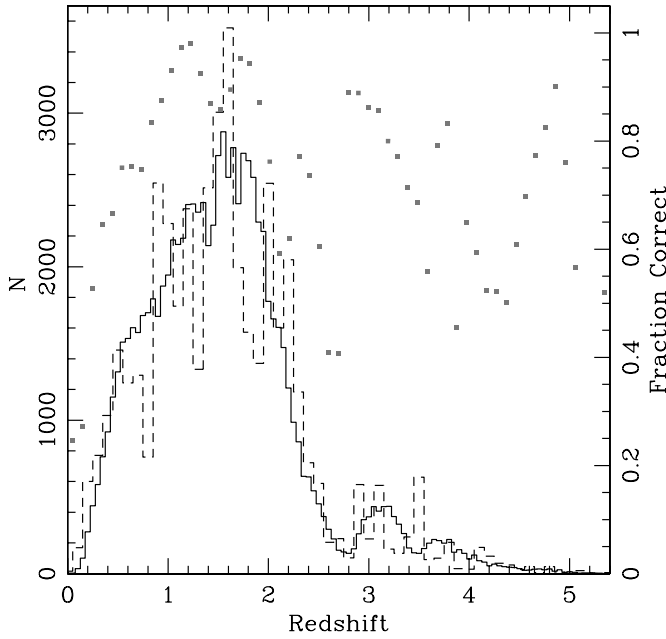
Figure 14. Distribution of spectroscopic redshifts for confirmed quasars in the sample (solid line). The dashed line shows the photometric redshift distribution of the spectroscopically-confirmed quasars. The photometric redshifts are only as accurate as the size of the redshift bins that can be used to define the color–redshift relation, which coarsely quantizes the z_{phot} distribution. Gray squares indicate the fraction of photo-z's that are correct to within ±0.3 for each z_{phot} bin. These are most accurate where the most data exist ($1 < z < 2$).
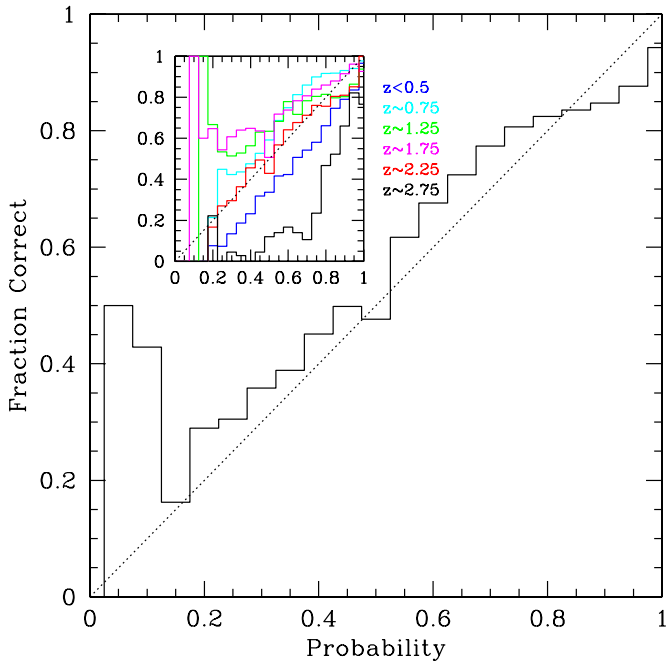
Figure 15. Actual fraction of quasars with correct redshift as a function of the quoted probability that the redshift (actually the redshift range) is correct (solid line: $\Delta z \pm 0.3$). The inset shows the distribution as a function of redshift. Over $0.5 < z < 2.5$, the photo-z probabilities are quite accurate (if not underestimates).

(A color version of this figure is available in the online journal.)

data exist, and in some high-z bins where the errors are larger, but are generally not catastrophic. The redshift dependence of this accuracy should be taken into account for any statistical use of the catalog.
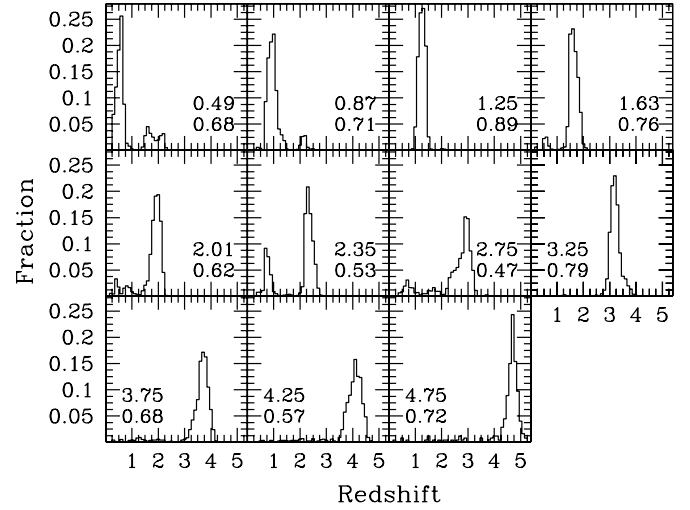
Figure 16. Spectroscopic redshift distribution of known quasars in 11 different bins of photometric redshift. Bins are chosen to match those of the Richards et al. (2006) quasar luminosity function. Some photometric redshift bins are quite robust (e.g., $1.06 < z_{phot} < 1.44$), while others have large spreads or catastrophic errors (e.g., $2.5 < z_{phot} < 3.0$). The mean redshift of each bin is given in each panel along with the fraction of objects within the redshift range explored (top and bottom numbers, respectively).

The photo-z code also gives a probability of an object being in a given redshift range (where the size of that range can vary considerably). That is, we give not only the most likely redshift but also the probability that the redshift is between some minimum and maximum values, which is crucial for dealing with catastrophic failures. Figure 15 plots the estimated probability of the photometric redshift being in the given range versus the actual fraction of those objects with accurate photometric redshifts, demonstrating that these probabilities are accurate in the ensemble average. The inset shows a breakdown as a function of the photometric redshift. Judicious use of the predicted redshifts, the range given, and the probability of the object having a redshift in that range allow these photometric redshift estimates to be very useful for a number of science applications.

One can get a better idea of where the catastrophic photometric redshift failures occur by looking at the distribution of true redshifts within a given photometric redshift bin, as shown in Figure 16. The photometric redshift bins were chosen to match those of the Richards et al. (2006) quasar luminosity function as it is necessary to correct for such photometric redshift errors before determining the quasar luminosity function from our sample (Section 5). The bins edges are at (0.3, 0.68, 1.06, 1.44, 1.82, 2.2, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0). We find that objects with photometric redshifts of $z \sim 1.25$, $z \sim 3.25$, and $z \sim 4.75$ are particularly robust (but note that this robustness is independent of the robustness of the initial quasar classification, which may be worse (e.g., at $z \sim 4.75$)).

4.7. Matching to Radio, X-ray, and Proper-Motion Catalogs

Three additional sources of information that we have used in determining the legitimacy of quasar candidates are their radio and X-ray flux densities and their proper motions. While not all radio and X-ray sources are quasars, the likelihood of a given object that otherwise appears to be a quasar goes up considerably if the source is also detected in the radio or X-ray. However, objects with large proper motions (and small errors) cannot be distant quasars. Compilation of this multiwavelength and

proper-motion information is done within the SDSS database and is described by Stoughton et al. (2002), so we describe them only briefly here.

Objects in the SDSS database are matched (with a 1″.5 radius) to the FIRST (Becker et al. 1995) VLA 20 cm catalog, and resulting radio fluxes are included in the catalog. Column 22 of Table 1 indicates the peak 20 cm flux densities (in mJy) for those quasars with FIRST matches. Entries of "−1" indicate no radio detection (or no coverage of that position). In all, we catalog 18,377 radio detections. As this is considerably lower than what one expects from the fraction of radio-loud quasars (e.g., Ivezić et al. 2002), it is clear that deeper radio surveys are needed. The FIRST survey would need to be about ten times deeper to detect all of the radio-loud quasars in our catalog.

We have also included the results of the cross-correlation of SDSS sources with the X-ray sources listed in the Bright and Faint Source catalogs of the *ROSAT* All-Sky Survey (RASS; Voges et al. 1999, 2000). Positional accuracies for RASS X-ray sources vary with the count rate, but typically have an uncertainty of $\sim$10–30″. Among the SDSS quasar candidates presented here, there are 11,965 objects whose optical positions fall within 60″ of a RASS X-ray source; for these sources, Column 23 of Table 1 gives the broadband (0.1–2.4 kev) count rate (counts s^{-1}) corrected for vignetting. Entries of "−1" indicate no RASS X-ray detection. Note that the large *ROSAT* error circle means that $\sim$28% of these X-ray matches will be spurious; this fraction reduces to $\sim$11% for a 30″ matching radius. A total of 1413 objects have both radio and X-ray matches.

Objects with large proper motions can be rejected as quasar candidates. Thus, we include USNO-B+SDSS proper-motion information in this catalog as it is tabulated in the SDSS database; see Munn et al. (2004).[11] As in Paper I, some constraints are applied in this matching to ensure that the proper-motion measurements are as reliable as possible. Specifically, there must only be one match between SDSS and USNO-B, the number of epochs of observations must be six or more (one SDSS and five USNO), the distance to the next nearest object with $g < 22$ must be larger than 7 arcsec, and the rms proper-motion residuals must be less than 1000 mas per year in both right ascension and declination. In all, 142,271 objects meet these criteria (and have nonzero pm entries in the catalog). However, since quasars will have measured "proper motions" comparable to the typical errors in the proper motions, we must impose a limit on the proper motion to identify objects that are most likely to be stars. As in Paper I, we adopted a conservative limit of 20 mas/year as the threshold for moving objects. Such a cut rejects only 0.2% of the known quasars, while identifying 6.2% of known stars, yielding 3631 moving objects in the catalog that are unlikely to be quasars. Figure 17 shows the distribution of proper motions in the catalog. As the proper-motion catalog from Munn et al. (2004) had a faint limit of $g \sim 19.7$, it is useful to attempt identification of potentially moving objects to fainter limits. We accomplish this by identifying any objects (as moved in the catalog) whose row or column velocities (on the CCD, as measure by the SDSS photometric pipeline) exceed three times the errors in those quantities. This criterion identifies another 21,321 potentially moving objects that are statistically unlikely to be quasars.
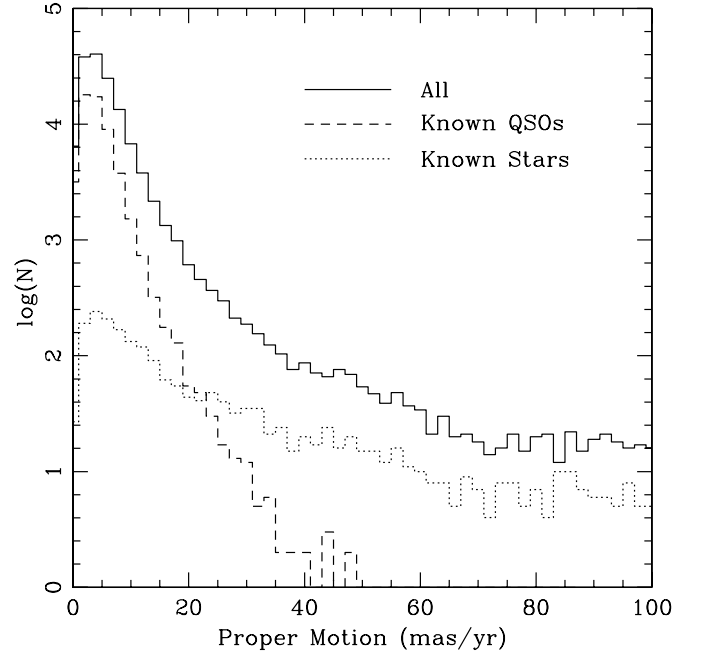

Figure 17. Histogram of measured proper motions for the entire catalog (solid), known quasars (dashed), and known stars (dotted). Due to measurement errors, stationary objects can have nonzero proper motion. Thus, we adopt a value of 20 mas/year as the cutoff for "moving" objects. For bright objects, a less conservative cutoff can be used.

5. NUMBER COUNTS AND THE LUMINOSITY FUNCTION

While the efficiency and completeness of a photometrically-selected quasar sample are perhaps not ideal for determining the number counts distribution and luminosity function, here we examine what we can learn about them from our sample.

Crudely taking our good $\geqslant 0$ quasar candidates as 100% efficient and complete, in Figure 18, we compare our catalog to the number counts of SDSS-DR3 quasars from Richards et al. (2006) and 2QZ/6QZ quasars from Croom et al. (2004). As no corrections for incompleteness or inefficiency in the photometric sample have been applied, this comparison is merely qualitative. However, the general agreement at both low- and high-z is reassuring and the excess at bright magnitudes is completely consistent with our estimate of the (low) efficiency of the brightest objects in our sample, and it should be possible to identify parameters to reduce this contamination.

Similarly, computation of the luminosity function from this catalog requires considerable care in terms of correcting for completeness and efficiency. Such analysis is beyond the scope of this paper. However, we can perform some relative comparisons of the quasar luminosity function (QLF) slopes with redshift that are independent of the overall normalization.

In particular, Richards et al. (2006) had confirmed previous indications of flattening of the slope of the QLF at high ($z \sim 4$) redshift (e.g., Fan et al. 2001). However, two lines of evidence have recently called that flattening into question. Fontanot et al. (2007), in their analysis, found no such flattening and attributed the Richards et al. (2006) flattening to completeness correction effects. Jiang et al. (2008), on the other hand, have not called the $z \sim 4$ result into question, but did show that the $z \sim 6$ slope is steeper and more consistent with $z \lesssim 2$ results, which may implicitly imply that the flattening of the $z \sim 4$ QLF is erroneous.

---

[11] Note that we have used corrected proper motions from this catalog (J. Munn 2008, private communication) that will also be available as part of SDSS-DR7.
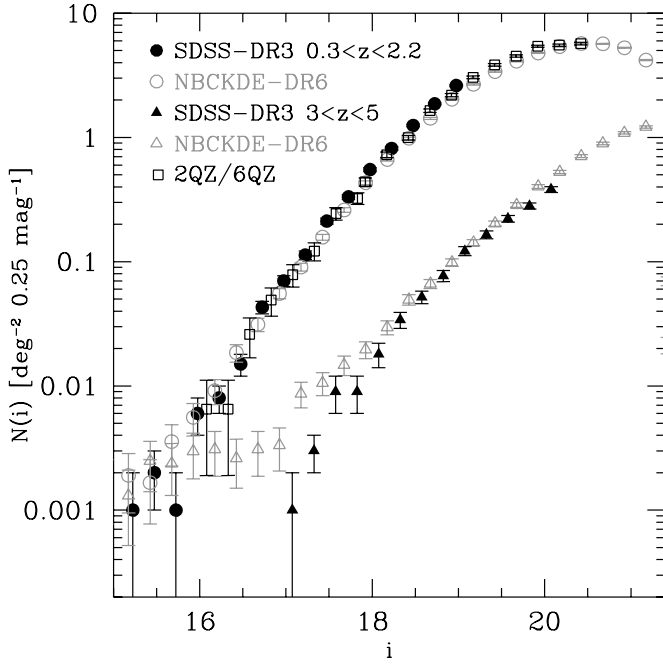
Figure 18. Number counts of quasars in the SDSS *i*-band. Solid circles and triangles show the SDSS-DR3 number counts for $0.3 < z < 2.2$ and $3 < z < 5$, respectively. Open circles and triangles give the values from this catalog (restricted to good $\geqslant 0$). The 2QZ/6QZ number counts are given by open squares. The photometric samples are highly contaminated at bright magnitudes. No corrections for efficiency or completeness have been applied; thus, this comparison is not ideal. Also note that the log–log nature of this plot means that even large discrepancies can appear quite small, but the general agreement is nevertheless reassuring.

Figure 19. Comparison of $z = 2.01$ and $z = 4.25$ quasar luminosity functions between the SDSS-DR3 spectroscopic sample and our DR6 photometric quasar sample. The photometric quasar sample has been corrected for the magnitude dependent on the catalog's efficiency; however, it has *not* been corrected for the overall efficiency or completeness. Thus, the scaling of the DR6phot points is completely arbitrary. We have simply matched the curves near $M_i = -29$ to the DR3 sample. $z \sim 2$ quasars are given as squares, closed and open for the spectroscopic and photometric samples, respectively. There is excellent agreement between the $z \sim 2$ photometric and spectroscopic samples. $z \sim 4$ quasars are given as triangles, closed and open for the spectroscopic and photometric samples, respectively. For the $z \sim 4.25$ photometric sample, gray open triangles are objects with good $\geqslant 0$, while the black open triangles are more conservatively restricted to good $\geqslant 1$. Even for the more conservative sample, statistically significant flattening of the $z \sim 4$ QLF is evident in our data photometric data set.

Here, we address this issue by comparing the $z \sim 2$ QLF to the $z \sim 4.25$ QLF that we derive from the catalog herein. No attempt has been made to correct for the overall efficiency and completeness of the catalog as we merely attempt to compare the slopes. We have, however, corrected for the magnitude dependence of the efficiency. Figure 19 shows the results of this comparison. Including all photometric quasar candidates with $z_{phot} \sim 4.25$ having good $\geqslant 0$, we find a slope similar to that of Richards et al. (2006). Restricting the sample with a more conservative good $\geqslant 1$ limitation yields a steeper slope, but still flatter than for $z \sim 2$. Adopting an even more restricted sample with good $\geqslant 2$ has no effect on the slope. The $z \sim 2$ slope is independent of our choice of good (for good $\geqslant 0$). While this sample cannot be considered completely independent of the Richards et al. (2006) sample (as it was used as the training set for our algorithm), we find statistically significant flattening that cannot be due to the completeness corrections used by Richards et al. (2006). Indeed, one does not necessarily expect the slopes to be similar since at high redshift, the quasar activity is expected to follow the growth of dark matter halos, while at $z \sim 2$–3 feedback mechanisms become dominant (e.g., Hopkins et al. 2007b).

6. CONCLUSIONS

Using a novel Bayesian algorithm, we identify 1,172,157 quasar candidates from a sample of over 40 million SDSS point sources. The overall efficiency of the catalog is $\sim 80\%$ and the catalog is expected to contain a minimum of 850,000 bona fide quasars. A UVX subsample, in excess of 500,000 objects, has an expected efficiency of over 97%. Additional information (redshift-dependent selection and radio, X-ray, and proper-
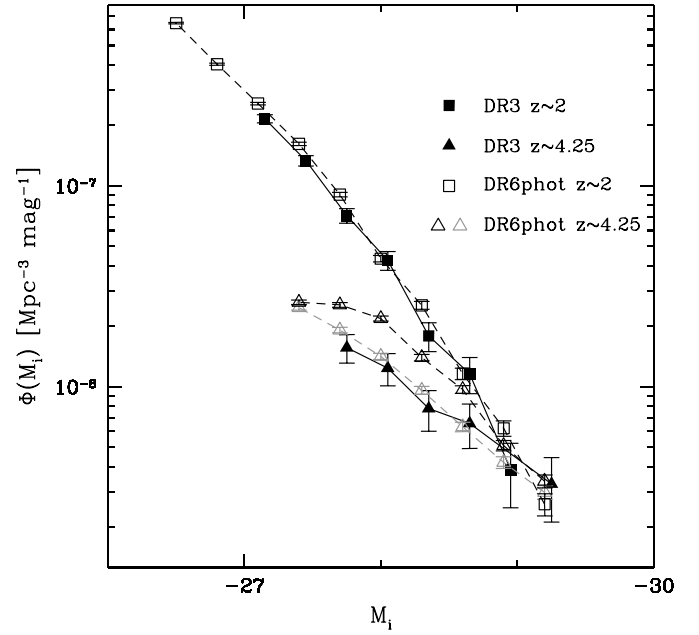
motion catalog matching) is provided in the catalog so that users can select subsamples that are optimal for any particular follow-up investigation. Photometric redshifts are estimated for the full sample and are expected to be accurate to ± 0.3 roughly 80% of the time, with outliers being statistically well defined. Cross-comparison with spectroscopically-confirmed type 1 quasars in the COSMOS field suggests that the sample is at least 70% complete and may recover additional objects missed by X-ray and radio selection methods. Careful analysis of the catalog could be used to create the deepest yet optical quasar luminosity function; simple arguments herein confirm the flattening of the QLF slope at $z \sim 4.25$ as compared with $z \sim 2$. A final installment of this catalog will come after the seventh SDSS data release in the fall of 2008 and should bring the total number of quasars over the one million mark.

Facilities: SDSS.

REFERENCES

Adelman-McCarthy, J. K., et al. 2007, ApJS, 172, 634
Adelman-McCarthy, J. K., et al. 2008, ApJS, 175, 297
Antonucci, R. 1993, ARA&A, 31, 473
Bayes, T. 1763, Phil. Trans. R. Soc. London, 53, 370
Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559
Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003, AJ, 125, 2276
Boyle, B. J., Shanks, T., Croom, S. M., Smith, R. J., Miller, L., Loaring, N., & Heymans, C. 2000, MNRAS, 317, 1014
Bramich, D. M., et al. 2008, MNRAS, 386, 887
Brandt, W. N., & Hasinger, G. 2005, ARA&A, 43, 827
Budavári, T., et al. 2001, AJ, 122, 1163
Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., & Munn, J. A. 1995, AJ, 110, 2655
Cool, R. J. 2006, BAAS, 38, 1170
Croom, S. M., Smith, R. J., Boyle, B. J., Shanks, T., Loaring, N. S., Miller, L., & Lewis, I. J. 2001, MNRAS, 322, L29
Croom, S. M., Smith, R. J., Boyle, B. J., Shanks, T., Miller, L., Outram, P. J., & Loaring, N. S. 2004, MNRAS, 349, 1397
Croom, S. M., et al. 2008, MNRAS, in press
Fan, X., et al. 2001, AJ, 121, 54
Fan, X., et al. 2006, AJ, 131, 1203
Fontanot, F., Cristiani, S., Monaco, P., Nonino, M., Vanzella, E., Brandt, W. N., Grazian, A., & Mao, J. 2007, A&A, 461, 39
Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748
Gallagher, S. C., Brandt, W. N., Chartas, G., & Garmire, G. P. 2002, ApJ, 567, 37
Giannantonio, T., et al. 2006, Phys. Rev. D, 74, 063520
Giannantonio, T., Scranton, R., Crittenden, R. G., Nichol, R. C., Boughn, S. P., Myers, A. D., & Richards, G. T. 2008, Phys. Rev. D, 77, 123520
Glikman, E., Helfand, D. J., White, R. L., Becker, R. H., Gregg, M. D., & Lacy, M. 2007, ApJ, 667, 673
Gray, A., & Riegel, R. 2006, in Proc. in Computational Statistics, ed. A. Rizzi & M. Vichi (Heildelberg: Physica-Verlag, Springer), 845
Gray, A. G., & Moore, A. W. 2003, in Proc. Third SIAM International Conf. on Data Mining, ed. D. Barbara & C. Kamath (Philadelphia, PA: SIAM), 203
Green, P. J., Aldcroft, T. L., Mathur, S., Wilkes, B. J., & Elvis, M. 2001, ApJ, 558, 109
Gunn, J. E., et al. 1998, AJ, 116, 3040
Gunn, J. E., et al. 2006, AJ, 131, 2332
Hasinger, G., et al. 2007, ApJS, 172, 29

Hennawi, J. F., et al. 2006, AJ, 131, 1
Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1995, AJ, 109, 1498
Hewitt, A., & Burbidge, G. 1993, ApJS, 87, 451
Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, AJ, 122, 2129
Hopkins, P. F., Lidz, A., Hernquist, L., Coil, A. L., Myers, A. D., Cox, T. J., & Spergel, D. N. 2007a, ApJ, 662, 110
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007b, ApJ, 654, 731
Ivezić, Ž., et al. 2002, AJ, 124, 2364
Ivezić, Ž., et al. 2004, Astron. Nachr., 325, 583
Ivezić, Ž., et al. 2007, AJ, 134, 973
Jannuzi, B. T., & Dey, A. 1999, in ASP Conf. Ser. 191, Photometric Redshifts and the Detection of High Redshift Galaxies Vol. 191, ed. R. Waymann et al. (San Francisco: ASP), 111
Jiang, L., et al. 2006, AJ, 131, 2788
Jiang, L., et al. 2008, AJ, 135, 1057
Kaiser, N., et al. 2002, in Proc. SPIE 4836, Survey and Other Telescope Technologies and Discoveries, ed. J. A. Tyson & S. Wolff (Bellingham, WA: SPIE), 154
Lacy, M., et al. 2004, ApJS, 154, 166
Lupton, R. H., Gunn, J. E., Ivezić, Z., Knapp, G. R., Kent, S., & Yasuda, N. 2001, in ASP Conf. Ser. 238, Astronomical Data Analysis Software and Systems Vol. 10, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco, CA: ASP), 269
Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406
MacAlpine, G. M., Lewis, D. W., & Smith, S. B. 1977, ApJS, 35, 203
Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, MNRAS, 242, 43P
Maddox, N., Hewett, P. C., Warren, S. J., & Croom, S. M. 2008, MNRAS, 386, 1605
Martínez-Sansigre, A., Rawlings, S., Lacy, M., Fadda, D., Jarvis, M. J., Marleau, F. R., Simpson, C., & Willott, C. J. 2006, MNRAS, 370, 1479
Ménard, B., Nestor, D., Turnshek, D., Quider, A., Richards, G., Chelouche, D., & Rao, S. 2008, MNRAS, 385, 1053
Morris, S. L., Weymann, R. J., Anderson, S. F., Hewett, P. C., Francis, P. J., Foltz, C. B., Chaffee, F. H., & MacAlpine, G. M. 1991, AJ, 102, 1627
Munn, J. A., et al. 2004, AJ, 127, 3034
Myers, A. D., Brunner, R. J., Nichol, R. C., Richards, G. T., Schneider, D. P., & Bahcall, N. A. 2007a, ApJ, 658, 85
Myers, A. D., Brunner, R. J., Richards, G. T., Nichol, R. C., Schneider, D. P., & Bahcall, N. A. 2007b, ApJ, 658, 99
Myers, A. D., Richards, G. T., Brunner, R. J., Schneider, D. P., Strand, N. E., Hall, P. B., Blomquist, J. A., & York, D. G. 2008, ApJ, 678, 635
Myers, A. D., et al. 2006, ApJ, 638, 622
Padmanabhan, N., et al. 2008, ApJ, 674, 1217
Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, AJ, 125, 1559
Polletta, M., Weedman, D., Hönig, S., Lonsdale, C. J., Smith, H. E., & Houck, J. 2008, ApJ, 675, 960
Reyes, R., et al. 2008, accepted (arXiv:0801.1115)
Richards, G. T., et al. 2001, AJ, 122, 1151
Richards, G. T., et al. 2002, AJ, 123, 2945
Richards, G. T., et al. 2003, AJ, 126, 1131
Richards, G. T., et al. 2004, ApJS, 155, 257 (Paper I)
Richards, G. T., et al. 2006, AJ, 131, 2766 (arXiv:astro-ph/0601434)
Richards, G. T., et al. 2008, AJ, submitted (arXiv:0810.3567)
Riegel, R., Gray, A., & Richards, G. 2008, in Proc. SIAM International Conference on Data Mining (SDM), ed. N. Abe et al. (Philadelphia, PA: SIAM), 208
Sachs, R. K., & Wolfe, A. M. 1967, ApJ, 147, 73
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Schmidt, M. 1963, Nature, 197, 1040
Schneider, D. P., et al. 2005, AJ, 130, 367
Schneider, D. P., et al. 2007, AJ, 134, 102 (arXiv:0704.0806)
Scoville, N., et al. 2007, ApJS, 172, 1 (arXiv:astro-ph/0612305)
Scranton, R., et al. 2002, ApJ, 579, 48
Scranton, R., et al. 2005, ApJ, 633, 589 (arXiv:astro-ph/0504510)
Silverman, B. W. 1986, Density Estimation for Statistics and Data Analysis, (London: Chapman and Hall)
Smith, J. A., et al. 2002, AJ, 123, 2121
Stern, D., et al. 2005, ApJ, 631, 163
Stocke, J. T., Morris, S. L., Weymann, R. J., & Foltz, C. B. 1992, ApJ, 396, 487
Stoughton, C., et al. 2002, AJ, 123, 485
The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346
Treister, E., et al. 2004, ApJ, 616, 123

Trump, J. R., et al. 2006, ApJS, 165, 1
Trump, J. R., et al. 2007, ApJS, 172, 383
Tucker, D. L., et al. 2006, Astron. Nachr., 327, 821
Tyson, J. A. 2002, in Proc. SPIE 4836, Survey and Other Telescope Technologies
    and Discoveries, ed. J. A. Tyson & S. Wolff (Bellingham, WA: SPIE), 10
Vanden Berk, D. E., et al. 2004, ApJ, 601, 692
Véron-Cetty, M.-P., & Véron, P. 2006, A&A, 455, 773
Weinstein, M. A., et al. 2004, ApJS, 155, 243
Wolf, C., Meisenheimer, K., & Röser, H.-J. 2001, A&A, 365, 660
York, D. G., et al. 2000, AJ, 120, 1579