FERMILAB-PUB-07-784-CD

The Fifth Data Release of the Sloan Digital Sky Survey

Jennifer K. Adelman-McCarthy[1], Marcel A. Agüeros[2], Sahar S. Allam[1,3], Kurt S. J. Anderson[4], Scott F. Anderson[2], James Annis[1], Neta A. Bahcall[5], Coryn A. L. Bailer-Jones[6], Ivan K. Baldry[7,8], J. C. Barentine[4], Timothy C. Beers[9], V. Belokurov[10], Andreas Berlind[11], Mariangela Bernardi[12], Michael R. Blanton[11], John J. Bochanski[2], William N. Boroski[1], D. M. Bramich[10], Howard J. Brewington[4], Jarle Brinchmann[13], J. Brinkmann[4], Robert J. Brunner[14], Tamás Budavári[8], Larry N. Carey[2], Samuel Carliles[8], Michael A. Carr[5], Francisco J. Castander[15], A. J. Connolly[16], R. J. Cool,[17], Carlos E. Cunha,[18,19], István Csabai[20,8], Julianne J. Dalcanton[2], Mamoru Doi[21], Daniel J. Eisenstein[17], Michael L. Evans[2], N. W. Evans[10], Xiaohui Fan[17], Douglas P. Finkbeiner[5], Scott D. Friedman[22], Joshua A. Frieman[1,18,19], Masataka Fukugita[23], Bruce Gillespie[4], G. Gilmore[10], Karl Glazebrook[8], Jim Gray[24], Eva K. Grebel[25], James E. Gunn[5], Ernst de Haas[5], Patrick B. Hall[26], Michael Harvanek[4], Suzanne L. Hawley[2], Jeffrey Hayes[27], Timothy M. Heckman[8], John S. Hendry[1], Gregory S. Hennessy[28], Robert B. Hindsley[29], Christopher M. Hirata[30], Craig J. Hogan[2], David W. Hogg[11], Jon A. Holtzman[31], Shin-ichi Ichikawa[32], Takashi Ichikawa[33], Željko Ivezić[2], Sebastian Jester[34], David E. Johnston[35,36], Anders M. Jorgensen[37], Mario Jurić[5,30], Guinevere Kauffmann[38], Stephen M. Kent[1], S. J. Kleinman[39], G. R. Knapp[5], Alexei Yu. Kniazev[6], Richard G. Kron[18,1], Jurek Krzesinski[4,40], Nikolay Kuropatkin[1], Donald Q. Lamb[18,41], Hubert Lampeitl[22], Brian C. Lee[42,43], R. French Leger[1], Marcos Lima,[44,19], Huan Lin[1], Daniel C. Long[4], Jon Loveday[45], Robert H. Lupton[5], Rachel Mandelbaum[30], Bruce Margon[46], David Martínez-Delgado[48], Takahiko Matsubara[49], Peregrine M. McGehee[50], Timothy A. McKay[51], Avery Meiksin[52], Jeffrey A. Munn[53], Reiko Nakajima[12], Thomas Nash[1], Eric H. Neilsen, Jr.[1], Heidi Jo Newberg[54], Robert C. Nichol[55], Maria Nieto-Santisteban[8], Atsuko Nitta[56], Hiroaki Oyaizu,[18,19], Sadanori Okamura[57], Jeremiah P. Ostriker[5], Nikhil Padmanabhan[47,42], Changbom Park[58], John Peoples Jr.[1], Jeffrey R. Pier[53], Adrian C. Pope[8], Dimitri Pourbaix[5,59], Thomas R. Quinn[2], M. Jordan Raddick[8], Paola Re Fiorentin[6], Gordon T. Richards[8,60], Michael W. Richmond[61], Hans-Walter Rix[6], Constance M. Rockosi[62], David J. Schlegel[42], Donald P. Schneider[63], Ryan Scranton[16], Uroš Seljak[47,5], Erin Sheldon[18,19], Kazu Shimasaku[57], Nicole M. Silvestri[2], J. Allyn Smith[37,64], Vernesa Smolčić[6], Stephanie A. Snedden[4], Albert Stebbins[1] Chris Stoughton[1], Michael A. Strauss[5], Mark SubbaRao[18,65], Yasushi Suto[66], Alexander S. Szalay[8], István Szapudi[67], Paula Szkody[2], Max Tegmark[68], Aniruddha R. Thakar[8], Christy A. Tremonti[17], Douglas L. Tucker[1], Alan Uomoto[8,69], Daniel E. Vanden Berk[63], Jan Vandenberg[8], S. Vidrih[10], Michael S. Vogeley[60], Wolfgang Voges[70], Nicole P. Vogt[31], David H. Weinberg[71], Andrew A. West[72], Simon D.M. White[38], Brian Wilhite[14], Brian Yanny[1], D. R. Yocum[1], Donald G. York[18,41], Idit Zehavi[73], Stefano Zibetti[70], Daniel B. Zucker[6,10]

[1] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510.

[2] Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195.

[3] Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071.

[4] Apache Point Observatory, P.O. Box 59, Sunspot, NM 88349.

[5] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544.

[6] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany.

[7] Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, UK

[8] Center for Astrophysical Sciences, Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218.

[9] Department of Physics & Astronomy and Joint Institute for Nuclear Astrophysics, Michigan State University, East Lansing, MI 48824-1116

[10] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[11] Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003.

[12] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104.

[13] Centro de Astrofísica da Universidade do Porto, Rua das Estrelas - 4150-762 Porto, Portugal.

[14] Department of Astronomy University of Illinois 1002 West Green Street, Urbana, IL 61801.

[15]Institut d'Estudis Espacials de Catalunya/CSIC, Gran Capitá 2-4, E-08034 Barcelona, Spain.

[16] Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260.

[17] Steward Observatory, 933 North Cherry Avenue, Tucson, AZ 85721.

[18] Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

[19] Kavli Institute for Cosmological Physics, The University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

[20] Department of Physics of Complex Systems, Eötvös Loránd University, Pf. 32, H-1518 Budapest, Hungary.

[21]Institute of Astronomy, School of Science, University of Tokyo, Osawa 2-21-1, Mitaka, 181-0015, Japan.

[22] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218.

[23]Institute for Cosmic Ray Research, University of Tokyo, 5-1-5 Kashiwa, Kashiwa City, Chiba 277-8582, Japan.

[24] Microsoft Research, 455 Market Street, Suite 1690, San Francisco, CA 94105.

[25] Astronomical Institute of the University of Basel, Department of Physics and Astronomy, Venusstrasse 7, CH-4102 Basel, Switzerland

[26] Dept. of Physics & Astronomy, York University, 4700 Keele St., Toronto, ON, M3J 1P3, Canada

[27] Institute for Astronomy and Computational Sciences Physics Department Catholic University of America Washington DC 20064

[28] US Naval Observatory, 3540 Massachusetts Avenue NW, Washington, DC 20392.

[29] Code 7215, Remote Sensing Division Naval Research Laboratory 4555 Overlook Avenue SW Washington, DC 20392.

[30] Institute for Advanced Study Einstein Drive Princeton, NJ 08540

[31] Department of Astronomy, MSC 4500, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003.

[32]National Astronomical Observatory, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan.

[33] Astronomical Institute, Tohoku University, Aoba, Sendai 980-8578, Japan

[34] School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, United Kingdom

[35] Jet Propulsion Laboratory, 4800 Oak Drive, Pasadena, CA 91109

[36] California Institute of Technology, 1200 East California Blvd, Pasadena, CA 91125

[37] ISR-4, MS D448, Los Alamos National Laboratory, P.O.Box 1663, Los Alamos, NM 87545.

[38] Max Planck Institut für Astrophysik, Postfach 1, D-85748 Garching, Germany.

[39] Subaru Telescope, 650 N. A'ohoku Place, Hilo, HI 96720, USA

[40] Obserwatorium Astronomiczne na Suhorze, Akademia Pedogogiczna w Krakowie, ulica Podchorążych 2, PL-30-084 Kraców, Poland.

[41] Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

[42] Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley CA 94720.

[43] Gatan Inc., Pleasanton, CA 94588

[44] Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

[45] Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QJ, UK.

[46] Department of Astronomy & Astrophysics, University of California, Santa Cruz, CA 95064

[47] Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544.

[48] Instituto de Astrofisica de Canarias, La Laguna, Spain

[49] Department of Physics and Astrophysics, Nagoya University, Chikusa, Nagoya 464-8602, Japan.

[50] AOT-IC, MS H820, Los Alamos National Laboratory, P.O.Box 1663, Los Alamos, NM 87545.

[51] Department of Physics, University of Michigan, 500 East University Avenue, Ann Arbor, MI 48109.

[52] SUPA; Institute for Astronomy, Royal Observatory, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK.

[53] US Naval Observatory, Flagstaff Station, 10391 W. Naval Observatory Road, Flagstaff, AZ 86001-8521.

[54] Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180.

[55] Institute of Cosmology and Gravitation (ICG), Mercantile House, Hampshire Terrace, Univ. of Portsmouth, Portsmouth, PO1 2EG, UK.

[56] Gemini Observatory, 670 N. A'ohoku Place, Hilo, HI 96720, USA

[57]Department of Astronomy and Research Center for the Early Universe, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan.

[58] Korea Institute for Advanced Study, 207-43 Cheong-Yang-Ni, Dong-Dae-Mun-Gu Seoul 130-722, Korea

[59] FNRS Institut d'Astronomie et d'Astrophysique, Université Libre de Bruxelles, CP. 226, Boulevard du Triomphe, B-1050 Bruxelles, Belgium.

[60] Department of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104.

[61] Department of Physics, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623-5603.

[62] UCO/Lick Observatory, University of California, Santa Cruz, CA 95064.

[63] Department of Astronomy and Astrophysics, 525 Davey Laboratory, Pennsylvania State University, University Park, PA 16802.

[64] Department of Physics and Astronomy, Austin Peay State University, P.O. Box 4608, Clarksville, TN 37040

[65] Adler Planetarium and Astronomy Museum, 1300 Lake Shore Drive, Chicago, IL 60605.

[66] Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan

[67] Institute for Astronomy, 2680 Woodlawn Road, Honolulu, HI 96822.

[68] Dept. of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139.

[69] Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101.

[70] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85741 Garching, Germany.

[71] Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210.

[72] Astronomy Department, 601 Campbell Hall, University of California, Berkeley, CA 94720-3411

ABSTRACT

This paper describes the Fifth Data Release (DR5) of the Sloan Digital Sky Survey (SDSS). DR5 includes all survey quality data taken through June 2005 and represents the completion of the SDSS-I project (whose successor, SDSS-II will continue through mid-2008). It includes five-band photometric data for 217 million objects selected over 8000 deg^2, and 1,048,960 spectra of galaxies, quasars, and stars selected from 5713 deg^2 of that imaging data. These numbers represent a roughly 20% increment over those of the Fourth Data Release; all the data from previous data releases are included in the present release. In addition to "standard" SDSS observations, DR5 includes repeat scans of the southern equatorial stripe, imaging scans across M31 and the core of the Perseus cluster of galaxies, and the first spectroscopic data from SEGUE, a survey to explore the kinematics and chemical evolution of the Galaxy. The catalog database incorporates several new features, including photometric redshifts of galaxies, tables of matched objects in overlap regions of the imaging survey, and tools that allow precise computations of survey geometry for statistical investigations.

Subject headings: Atlases—Catalogs—Surveys

Submitted to The Astrophysical Journal Supplement Series, October 12, 2006

1. Introduction

The primary goals of the Sloan Digital Sky Survey (SDSS) are: a large-area, well-calibrated imaging survey of the north Galactic cap, repeat imaging of an equatorial stripe in the south Galactic cap to allow variability studies and deeper co-added imaging, and spectroscopic surveys of well-defined samples of roughly 10^6 galaxies and 10^5 quasars (York et al. 2000). The survey uses a dedicated, wide-field, 2.5m telescope (Gunn et al. 2006) at Apache Point Observatory, New Mexico. Imaging is carried out in drift-scan mode using a 142 mega-pixel camera (Gunn et al. 1998) that gathers data in five broad bands, $u\,g\,r\,i\,z$, spanning the range from 3000 to 10,000 Å (Fukugita et al. 1996), with an effective exposure time of 54.1 seconds per band. The images are processed using specialized software (Lupton et al. 2001; Stoughton et al. 2002; Lupton 2005), and are astrometrically (Pier et al. 2003)

[73] Department of Astronomy, Case Western Reserve University, Cleveland, OH 44106

and photometrically (Hogg et al. 2001; Tucker et al. 2006) calibrated using observations of a set of primary standard stars (Smith et al. 2002) observed on a neighboring 20-inch telescope.

Objects are selected from the imaging data for spectroscopy using a variety of algorithms, including a complete sample of galaxies with Petrosian (1976) r magnitudes brighter than 17.77 (Strauss et al. 2002), a deeper sample of color- and magnitude-selected luminous red galaxies (LRGs) from redshift 0.15 to beyond 0.5 (Eisenstein et al. 2001), a color-selected sample of quasars with $0 < z < 5.5$ (Richards et al. 2002), optical counterparts to ROSAT X-ray sources (Anderson et al. 2003), and a variety of stellar and calibrating objects (Stoughton et al. 2002; Adelman-McCarthy et al. 2006). These targets are observed by a pair of double spectrographs fed by 640 optical fibers, each $3''$ in diameter, plugged into aluminum plates 2.98° in diameter. The resulting spectra cover the wavelength range $3800 - 9200$ Å with a resolution of $\lambda/\Delta\lambda \approx 2000$. The finite size of the fiber cladding means that only one of two objects closer than $55''$ can be targeted on a given plate; this restriction results in a roughly 10% incompleteness in galaxy spectroscopy, but this incompleteness is well characterized and is generally straightforward to correct in statistical calculations (e.g., Zehavi et al. 2002).

This paper presents the Fifth Data Release (DR5) of the SDSS, which follows the Early Data Release of commissioning data (EDR; Stoughton et al. 2002) and the regular data releases DR1-DR4 (Abazajian et al. 2003, 2004, 2005; Adelman-McCarthy et al. 2006). These data releases are cumulative, so all observations in the earlier releases are also included in DR5. There have been no substantive changes to the imaging or spectroscopic software since DR2, so DR5 includes data identical to DR2-DR4 in the overlapping regions. Finkbeiner et al. (2004) presented a separate ("Orion") release of imaging data outside the formal SDSS footprint, mostly at low Galactic latitudes.

The Fifth Data Release includes all survey quality data that were taken as part of "SDSS-I," the phase of the SDSS that ran through June 2005, including a variety of imaging scans and spectroscopic observations taken outside of the standard survey footprint or with non-standard spectroscopic target selection. The second "SDSS-II" phase, which includes a number of new participating institutions and will continue through mid-2008, consists of three distinct surveys: the Sloan Legacy Survey, the Sloan Supernova Survey, and the Sloan Extension for Galactic Understanding and Exploration (SEGUE). The Legacy survey is essentially a continuation of SDSS-I, with the goal of completing imaging and spectroscopy over about 8000 deg^2 of the north Galactic cap. The Supernova Survey (J. Frieman et al. 2007, in preparation) repeatedly scans a 300 square degree area in the south Galactic cap during the fall months to detect and measure time variable objects, especially Type Ia supernovae (out to $z \approx 0.4$) that can be used to measure the cosmic expansion history. SEGUE includes 3500 deg^2 of new imaging, mostly at Galactic latitudes below those of the original SDSS footprint,

and spectroscopy of about 240,000 selected stellar targets to study the structure, chemical evolution, and stellar content of the Milky Way. Future SDSS data releases will include data from all three surveys, and some early data from SEGUE are included in DR5. An initial release of imaging data and uncalibrated object catalogs from the Autumn 2005 season of the Supernova Survey is available at `http://www.sdss.org/drsn1/DRSN1_data_release.html`, but it is not part of DR5.

Section 2 of this paper describes the contents of DR5, and §3 summarizes information about data quality, including new tests of spectrophotometric accuracy. Section 4 describes several new features of DR5: photometric redshifts for galaxies, "sector/region" tables for precisely defining the survey geometry, and tools for matching repeat observations of the same objects. We conclude in § 5.

2. What is included in DR5

As described by Stoughton et al. (2002), public SDSS data are available both as flat files (from the Data Archive Server, or DAS) and via a flexible web interface to the SDSS database (the Catalog Archive Server, or CAS). Information about and entry points to both interfaces can be found at `http://www.sdss.org/dr5`. The CAS is a convenient and powerful tool for selecting objects found in the SDSS based on their location, photometric parameters, and (if they were observed spectroscopically) spectroscopic parameters. FITS images and spectra for individual objects and fields are available from the CAS; the DAS should be used for bulk downloads of large quantities of data. Links to extensive documentation and examples are available on the above web site.

The principal SDSS imaging data are taken along a series of great-circle stripes that aim to fill a contiguous area in the north Galactic cap, and along three non-contiguous stripes in the south Galactic cap. Each filled stripe consists of two interleaved strips because of the gaps between columns of CCDs in the imaging camera (see Gunn et al. 1998; York et al. 2000). Figure 1 shows the region of sky included in DR5, in imaging (top) and spectroscopy (bottom). In contrast to DR4, the imaging available in DR5 covers an essentially contiguous region of the north Galactic cap, with **a few small patches totaling** $\sim$ 200 **square degrees remaining (nearly all of this area will be included in DR6). The area covered by the DR5 primary imaging survey (including the southern stripes but not counting these patches) is 8000 deg^2.** The great circle stripes in the north overlap at the poles of the survey; 21% of this region of sky is covered more than once. In any region where imaging runs overlap, one run is declared primary and used for spectroscopic target selection, and other runs are declared secondary. DR5 includes both the primary and

secondary (repeat) observations of each area and source (see §4.3).

As spectroscopic observations necessarily lag the imaging, the DR5 spectroscopic area still has the gap at intermediate declinations that was present in the DR4 imaging coverage. The area covered by the spectroscopic survey is 5713 deg^2. The spectroscopic data include 1,048,960 spectra, arrayed on 1639 plates of 640 fibers each. Thirty-two fibers per plate are devoted to measurements of sky. Automated spectral classification yields approximately 675,000 galaxies, 90,000 quasars, and 216,000 stars. Nearly 99% of all spectra are of high enough quality to yield an unambiguous classification and redshift; most of the unidentified targets are either faint ($r > 20$) or have featureless spectra (hot stars or blazar-like AGN; see Collinge et al. 2005). However, in rare cases the assigned redshift is far from the true redshift, so for an object with unusual properties it is important to examine the spectra and to check for flags that can indicate data quality or classification problems. As described in the DR4 paper (Adelman-McCarthy et al. 2006), a number of plates have duplicate observations, usually just one but in some cases several. DR5 includes 62 duplicates of 53 unique main survey plates, and ten duplicates of special plates which take spectra outside the standard survey target selection. Some main-survey objects are also reobserved on adjacent plates to check the end-to-end reproducibility of spectroscopy. In total, about 2% of main-survey objects have one or more repeat spectra.

In the Fall months, when the southern Galactic cap is visible in the northern hemisphere, the SDSS imaging has been confined to a stripe along the Celestial Equator, plus two "outrigger" stripes, centered roughly at $\delta = +15°$ and $\delta = -10°$, respectively (these are visible on the right-hand-side of the panels of Figure 1). We have performed multiple imaging passes of the southern equatorial stripe (a.k.a. Stripe 82, spanning $22^{\rm h}\,20^{\rm m} < \alpha < 3^{\rm h}\,20^{\rm m}$, $-1.25° < \delta < +1.25$, in J2000 coordinates), which can be used for variability studies and for co-addition to create deeper summed images. Previous data releases have included only a single epoch of these observations. In DR5, we make available 36 runs on the northern strip of this stripe and 29 runs on the southern strip; these are all the observations of Stripe 82 carried out before July 2005 that are of survey quality. Each individual run covers only part of the full right ascension range of the stripe; Figure 2 shows the number of passes available along the northern and southern strips, as a function of right ascension. The central regions of the stripe have typically been covered 10-20 times. The extra runs are available in DR5 only through the DR supplemental DAS, described at `http://www.sdss.org/dr5/start/aboutdr5sup.html`. In future data releases, they will be made available through the CAS as well. Note that DR5 does not include those runs on Stripe 82 at larger right ascension, in the region of Orion, as described by Finkbeiner et al. (2004). Those runs continue to be made available through the websites indicated in that paper.

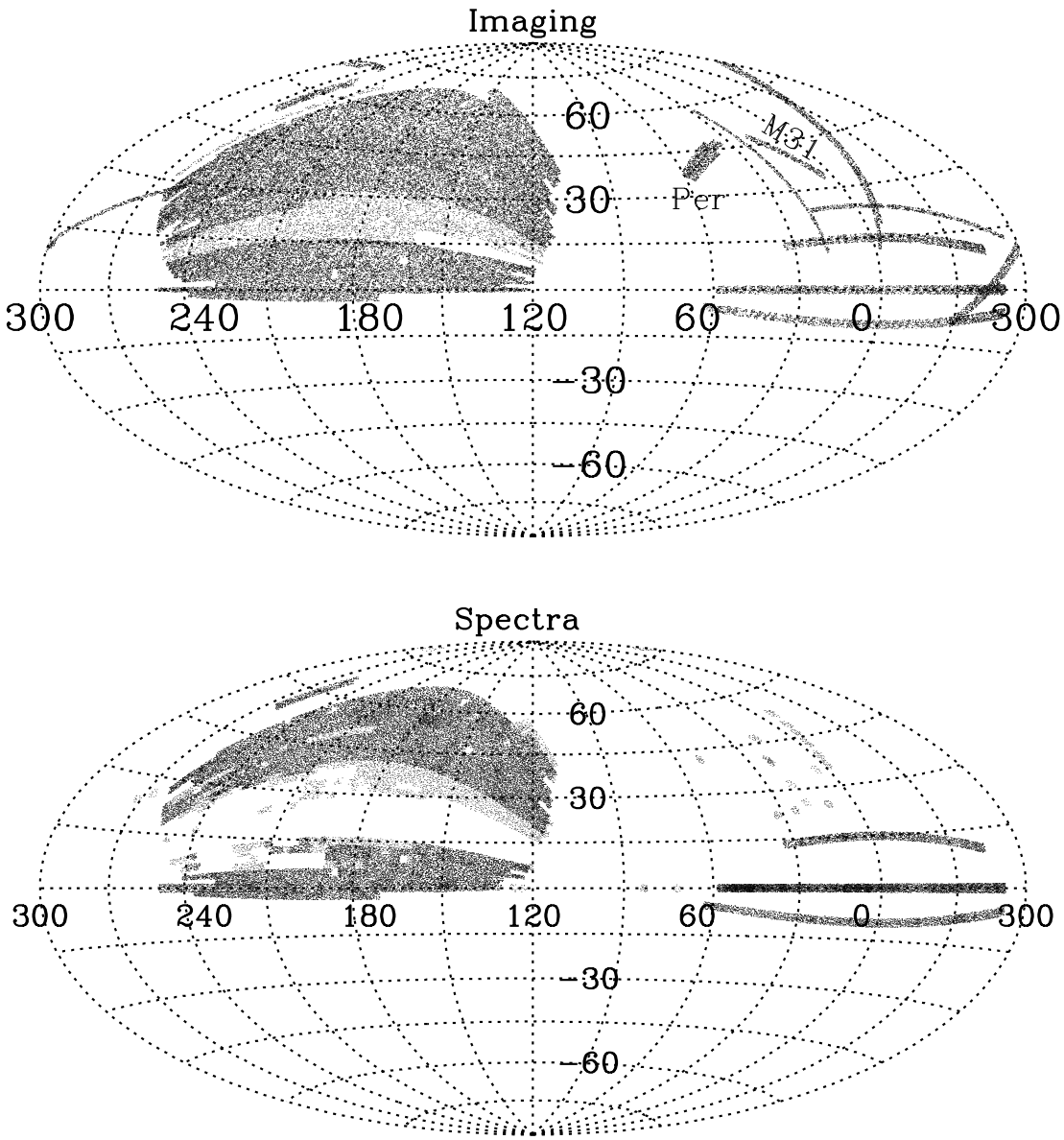

Fig. 1.— The distribution on the sky of SDSS imaging (upper panel) and spectroscopy (lower panel) included in DR5, shown in J2000 equatorial coordinates. The regions of sky that are new to DR5 are shaded more lightly. The upper panel includes both those regions included in the CAS (totaling 8000 deg^2) and the supplementary imaging runs available only through the DAS, which consist of SEGUE scans at low Galactic latitude and scans through M31 and the Perseus cluster.

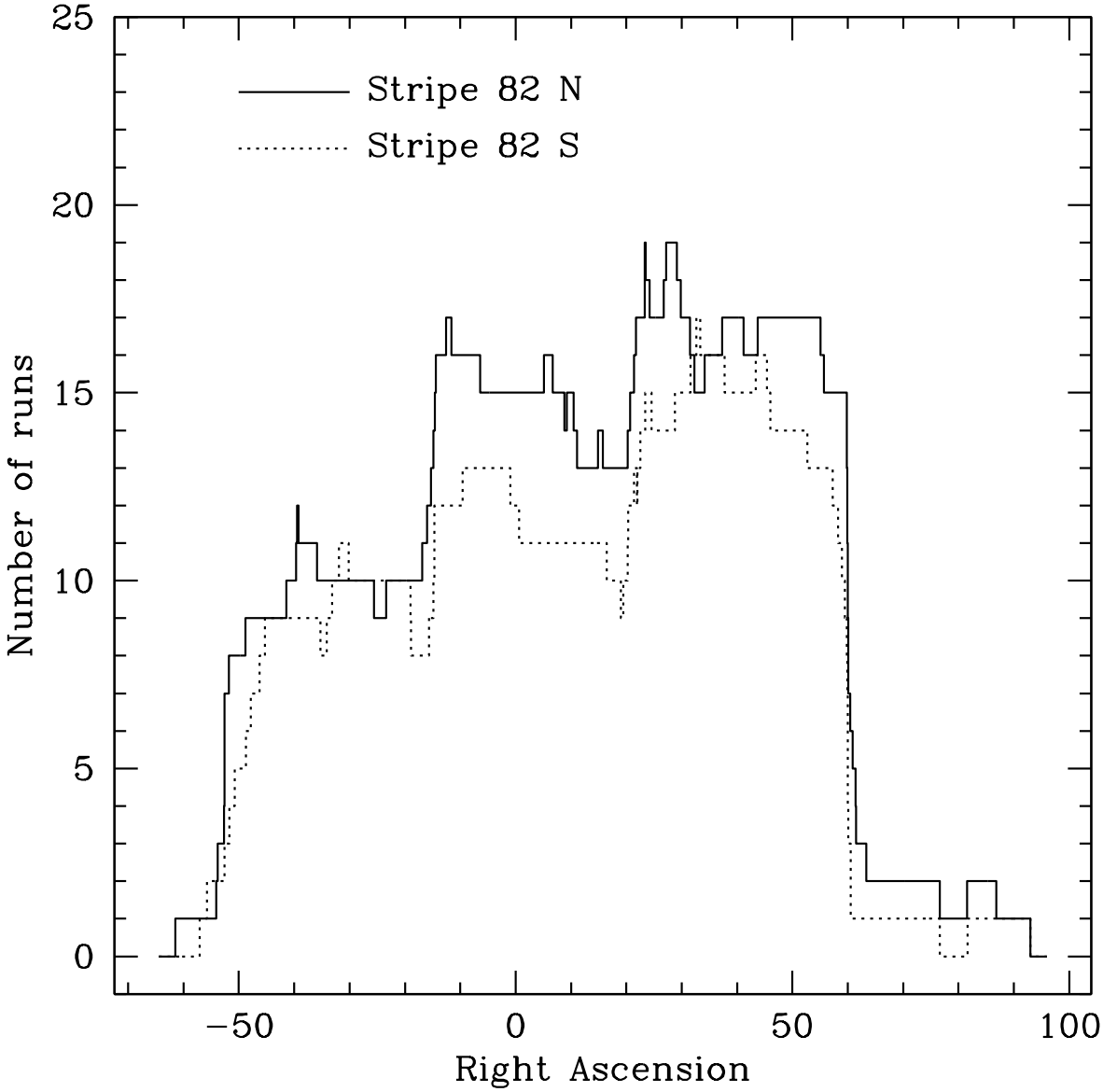

Fig. 2.— Coverage of the southern equatorial stripe in DR5. Solid and dotted lines show the number of photometric runs covering regions of different right ascension, for the northern and southern strips, respectively.

A combined, deep image of the full equatorial stripe is being prepared and will be made available in a future data release. However, for objects that can be detected in a single pass, the benefits of co-addition can mostly be realized simply by averaging the photometric measurements from the multiple passes, using the multiple entries in the photometric catalog rather than analyzing a summed image. Figure 3, based on the Stripe 82 stellar catalog of Ivezić et al. (in preparation), demonstrates this improvement, showing the $g - r$ vs. $u - g$ color-color diagram for blue, non-variable point sources (mostly white dwarfs) in Stripe 82. Data co-added at the catalog level have been used to search for faint quasars (Jiang et al. 2006), to measure the dispersion in galaxy colors on the red sequence (Cool et al. 2006), and to improve the signal-to-noise ratio of galaxy u-band Petrosian magnitudes (Baldry et al. 2005). The Stripe 82 data have also been used to search for variable and high proper motion objects (e.g., Ivezić et al. 2003) and to test the covariance of photometric errors among bands and among multiple objects in the same fields (Scranton et al. 2005). Because the catalogs from the multiple Stripe 82 scans are not yet available in the CAS, averaging or variability searches must be done by downloading object tables from the DAS and identifying repeat observations of the same object by positional matching.

In addition to the repeat scans on Stripe 82, several imaging runs outside of the standard footprint are included:

- Two runs that together make a 2.5° stripe crossing M31, the Andromeda Galaxy. These imaging data have been used to search for substructure in M31's halo (e.g., Zucker et al. 2004ab).
- Five runs that together cover 78 deg^2 centered roughly on the low-redshift Perseus cluster of galaxies.
- Ten runs of imaging data taken as part of the SEGUE survey, including stripes at $l = 50^\circ$ $(-46^\circ < b < -8^\circ)$, $l = 110^\circ$ $(-36^\circ < b < 29.5^\circ)$, and $l = 130^\circ$ $(-49 < b < -18.6)$, and a stripe that runs for 20 degrees along $\delta \approx 25^\circ$.

As with the repeat scans of Stripe 82, objects detected in these runs are recorded in the DR-supplemental DAS: `http://www.sdss.org/dr5/start/aboutdrsup.html`, but they are not, as yet, available in the CAS. All these runs are in quite crowded fields, as they tend to go to low Galactic latitude, or pass through the center of M31. The completeness and accuracy of the photometry produced by the automated SDSS pipeline becomes suspect in crowded fields, so these data should be used with care. Plots and tables of the field-by-field data quality for these runs may be accessed at `http://das.sdss.org/DRsup/data/imaging/QA/summaryQA_analyzePC.ht`

Because of the relatively small footprint of the imaging in the southern Galactic cap,

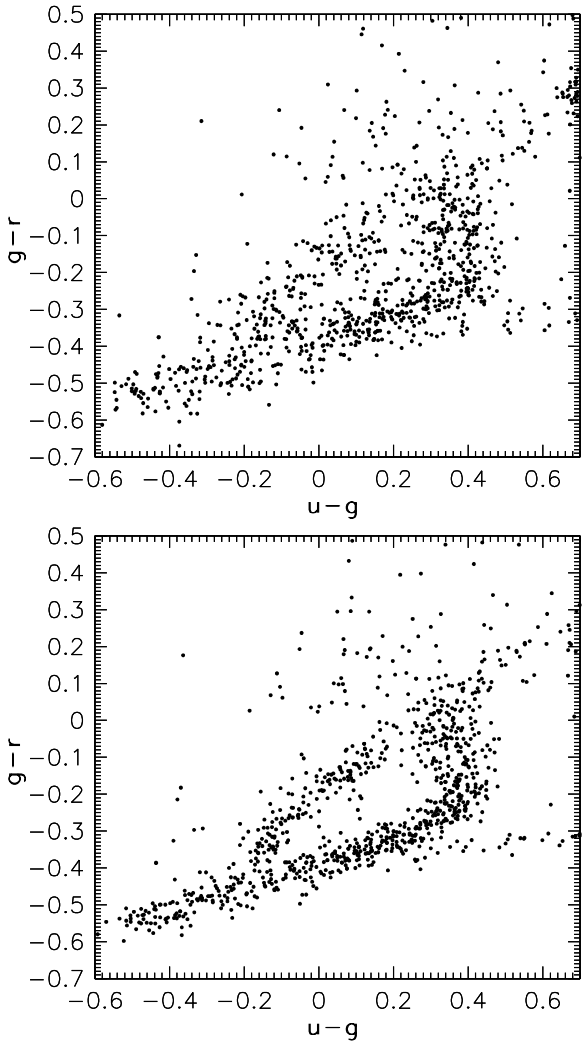

Fig. 3.— The $g - r$ vs. $u - g$ color-color diagram for the blue, non-variable point sources with $u < 20$ in the equatorial stripe (from Ivezić et al., in preparation). The top panel shows results using single-epoch DR5 photometry, while the bottom panel shows the striking improvement obtained by averaging the photometric measurements from all of the imaging passes, allowing clear separation between the sequences of helium white dwarfs (the top side of the "triangle") and hydrogen white dwarfs (which lie along the other two sides). **This region of color space also includes white dwarf–M dwarf pairs, hot subdwarfs, and quasars (see, e.g., the discussion of Eisenstein et al. [2006]).** Main sequence and red giant stars (far more numerous, of course), are mostly off the diagram to the upper right.

the spectroscopy of targets selected by our normal algorithms was completed quite early in the survey; most of these data were included already in DR1. We generally restrict imaging observations to pristine conditions, when the moon is below the horizon, the sky is cloudless, and the seeing is good. To make optimal use of the remaining time, we undertook a series of spectroscopic observing programs, based mostly on the imaging data of the equatorial stripe in the southern Galactic cap, designed to go beyond the science goals of the main survey. DR5 includes 299 plates from these programs, carried out in the Fall months of 2001–2004, with a total of 204,160 spectra. The great majority of these plates were already included in DR4; the target selection for them is described in the DR4 paper (Adelman-McCarthy et al. 2006), and we will not repeat it here. The science objectives include studies of galactic kinematics, calibration of photometric redshifts, evaluation of the completeness of the quasar survey (Vanden Berk et al. 2005), and surveys of galaxies that fall outside of the standard survey selection criteria (Baldry et al. 2005).

DR5 includes a total of 84 special plates that were not included in DR4. All of these were obtained as early data of the SEGUE program. Each SEGUE pointing includes two 640-fiber plates of different exposure times, with 592 brighter ($13 < g < 18$) and 560 fainter ($18 < g < 20$) stars targeted. The remaining targets are calibration standards and sky fibers. Target selection algorithms, which are outlined in Adelman-McCarthy et al. (2006) and will be described more fully in a future paper, identify candidate stars in the following categories: white dwarfs (25 per pointing), cool white dwarfs (10), A/BHB stars (150), F turnoff and sub-dwarf stars (150), G stars (375), K giants (100), low metallicity candidates (150), K dwarfs (125), M dwarfs (50), and AGB candidates (10). These plates are listed and described at `http://www.sdss.org/dr5/products/spectra/special.html`.

Tables 1 and 2 summarize the characteristics of the DR5 imaging and spectroscopic surveys, respectively. Note that the "star" and "galaxy" divisions in Table 1 refer to the photometric pipeline classifications; "stars" include quasars and any other unresolved sources, and "galaxies" are all resolved objects, including airplane and satellite trails, etc. Classifications in Table 2 are those returned by the spectroscopic pipeline; note, in particular, that the "quasar" classification (based on the presence of a securely detected, high excitation emission line with FWHM broader than 1000 km sec^{-1}) does not include any explicit luminosity cut.

DR5 contains several QSO-related tables and views. The `QuasarCatalog` table lists the individually inspected, luminosity and line-width restricted, bonafide quasars from the DR3 sample as published by Schneider et al. (2005). A similar catalog is now being created for DR5 (Schneider et al. 2007, in prep.) The `QSOBunch` table contains a record for each "object" flagged as a potential QSO in any of three catalog tables: `Target.PhotoObjAll`,

`Best.PhotoObjAll` or `SpecObj`. In such cases a bunch record describing the primary photo, target, and spectro objects within 1.5 arcseconds of that object is created. Identifiers of nearby objects from each catalog are combined into `QSOConcordanceAll` records that point to the `QSOBunch` record. Those identifiers in turn point to the `QSObest`, `QSOtarget`, and `QSOspec` tables that carry more detailed information about each object. Thus, the `QuasarCatalog` table provides straightforward access to a set of carefully vetted quasars with well defined selection criteria, while the `QSOConcordanceAll` table can be used to identify all objects that were flagged as potential quasars based on photometry and/or spectroscopy.

3. Data Quality

SDSS imaging data are obtained under photometric conditions, as determined by observations from the 0.5-m photometric monitoring telescope and a 10μm "cloud camera" (Hogg et al. 2001; Tucker et al. 2006). The median seeing of the imaging data is 1.4″ in the r band, and essentially all imaging data accepted as survey quality have seeing better than 2″ (see Figure 4). The 95% completeness limit for detection of point sources in the r band is 22.2 mag, **estimated from comparison to deeper surveys.** Constancy of stellar population colors shows that photometric calibration over the survey area is accurate to roughly 0.02 mag in the g, r and i bands, and 0.03 mag in u and z (Ivezić et al. 2004). Analysis of multiple observations of the southern Equatorial stripe shows that photometry of bright stars is repeatable at better than 0.01 mag in all bands and that the photometric pipeline correctly estimates random photometric errors (Z. Ivezić et al., in preparation). All magnitudes are roughly on an AB system (Oke & Gunn 1983) and use the "asinh" scale described by Lupton, Gunn, & Szalay (1999). The astrometric calibration precision is better than 0.1″ rms per coordinate (Pier et al. 2003).

The wavelength calibration uncertainty for SDSS spectra is roughly 0.05 Å. Note that spectra in DR5 (and DR2-DR4) are *not* corrected for Galactic extinction; this is a change relative to DR1. The spectra are flux-calibrated using observations of F subdwarfs, which are targeted for this purpose on each spectroscopic plate; **the calibration procedure is described in §4.1 of Abazajian et al. (2004). Wilhite et al. (2005) discuss the repeatability of stellar spectra taken more than 50 days apart. Their Figure 4 shows that the integrated flux has a 68% full-width of $\sim 10\%$ for non-variable stars, and their Figure 5 shows that the mean difference between spectra in two epochs of a single plate is $1-3\%$ over the full observed wavelength range, with no strong features at any wavelength.**

A useful way to test the quality of spectrophotometry on small scales (< 500Å) is to

Table 1. Characteristics of the DR5 Imaging Survey

Footprint area	8000 deg^2 (20% increment over DR4)
Imaging catalog	217 million unique objects
AB Magnitude limits:[a]	
u	22.0 mag
g	22.2 mag
r	22.2 mag
i	21.3 mag
z	20.5 mag
Median PSF width	$1.4''$ in r
RMS photometric calibration errors:	
r	2%
$u-g$	3%
$g-r$	2%
$r-i$	2%
$i-z$	3%
Astrometry errors	$< 0.1''$ rms absolute per coordinate
Object Counts:[b]	
Stars, primary	85,383,971
Stars, secondary	28,201,858
Galaxies, primary	131,721,365
Galaxies, secondary	33,044,047

[a]95% completeness for point sources in typical seeing; 50% completeness numbers are generally 0.4 mag fainter. **The difference between "asinh" magnitudes and conventional magnitudes is** $0.004-0.015$ **at the 95% limits and** $0.008-0.03$ **at the 50% limits, smaller than the uncertainty in conversion of magnitudes between surveys.**

[b]Primary imaging objects are those in the primary imaging area; secondary objects are in repeat imaging, so they are typically repeats of primary objects.

Table 2. Characteristics of the DR5 Spectroscopic Survey

Main Survey	
Footprint area	5713 deg^2 (19% increment over DR4)
Wavelength coverage	3800–9200Å
Resolution $\lambda/\Delta\lambda$	1800–2100
Signal-to-noise ratio[a]	> 4 per pixel at $g = 20.2$
Wavelength calibration errors	< 5 km sec^{-1}
Redshift accuracy	30 km sec^{-1} rms for Main galaxies
	$\sim 99\%$ of classifications and redshifts are reliable
Number of plates	1639
Number of spectra[b]	1,048,960
Galaxies	674,741
Science primary galaxies	561,530
Quasars	90,596
Science primary quasars	75,005
Stars	215,781
Sky	55,555
Unclassifiable	12,287
Additional Spectroscopy	
Repeat of main survey plates	62 plates
SEGUE and SEGUE test plates	80 plates (2 repeated)
Other southern programs	219 plates (8 repeated)

[a]Pixel size is 69 km s^{-1}, varying from 0.9Å (blue end) to 2.1Å (red end).

[b]Science primary objects define the set of unique science spectra of objects from main-survey plates (i.e., they exclude repeat observations, sky fibers, spectrophotometric standards, and objects from special plates).

observe a population of identical objects at a range of redshifts. Spectrophotometric residuals may then be computed by dividing the restframe spectra of objects in different redshift bins. While no ideal population of identical objects exists, elliptical galaxies have spectra that are similar, on average, over the redshift range $z = 0.04 - 0.20$, since they are no longer forming stars.

We select ellipticals for this test using their position in the color-magnitude diagram, with an additional cut on the Hα equivalent width of 2Å to exclude any objects with on-going star formation. We average 300 to 1000 spectra in the restframe in 160 bins of 0.001 in redshift from $z = 0.04 - 0.20$. To determine the spectrophotometry residuals we must fit out any evolution with redshift, which can arise from a combination of true passive evolution, slight changes in sample selection, and aperture effects. This is done by fitting a fourth-order polynomial to flux as function of redshift for each *rest-frame* wavelength. We divide the rest-frame spectra by these fits and interpolate back to the observed frame. The median of the residual spectra in the observed frame provides a measure of the spectrophotometry error, i.e., the mean factor by which the flux-calibrated spectrum provided by the spectroscopic pipeline is high or low compared to a perfectly calibrated spectrum. Since the evolutionary fits are themselves affected by the spectrophotometry errors, we apply the estimated correction to the averaged spectra and iterate the process, which converges rapidly.

Figure 5 shows the spectrophotometry residuals inferred from each of the 160 composite spectra, and the median of these residuals. There are sharp features associated with calcium and sodium absorption, probably originating in the Galactic interstellar medium, and with night sky emission lines. The most worrisome features are the wiggles below 4500 Å, with amplitude of $\sim 3\%$, centered on Ca H and K, Hδ, and Hγ. **The coincidence of these wiggles with known spectral features suggests that these residuals are caused by a systematic mismatch between the spectrophotometric standard stars and the model F-stars used in the calibration pipeline.**

One obvious question is the scale at which we can measure spectrophotometry errors with this technique. This scale is set by our ability to discriminate evolution effects from the spectrophotometry residuals, which in turn is related to the wavelength shift between our high- and low-redshift bins. We have tested the technique empirically by adding sine and cosine modulations with different periods to the observed frame and the seeing how well we recover them. Residuals seem to be well measured on scales less than 500Å, i.e., Figure 5 should reveal any systematic errors in SDSS photometry with periods shorter than this. On larger scales, we must rely on the F star spectral models, on tests against white dwarf model spectra (see figure 4 of Abazajian et al. 2004), and on checks of synthesized magnitudes against the photometry. Collectively, these tests imply that the flux-calibrated

SDSS spectra can be used for spectrophotometry at the few percent level.

4. New Features of DR5

4.1. Photometric Redshifts for Galaxies

DR5 includes two estimates of photometric redshifts for galaxies, calculated with two independent techniques.[1] The first uses the template fitting algorithm described by Csabai et al. (2003), which compares the expected colors of a galaxy (derived from template spectral energy distributions) with those observed for an individual galaxy. A common approach for template fitting is to take a small number of spectral templates T (e.g., E, Sbc, Scd, and Irr galaxies) and choose the best fit by optimizing the likelihood of the fit as a function of redshift, type, and luminosity, $p(z, T, L)$. We use a variant of this method that incorporates a continuous distribution of spectral templates, enabling the error function in redshift and type to be well defined. Since a representative set of photometrically calibrated spectra in the full wavelength range of the filters is not easy to obtain, we have started from the empirical templates of Coleman, Wu, & Weedman (1980), extended them with spectral synthesis models, and adjusted them to fit the colors of galaxies in the training set (Budavari et al. 2000). The results are listed in the CAS table `Photoz`, which includes the estimate of the redshift, spectral type, rest-frame colors, rest-frame absolute magnitudes, errors on all of these quantities, and a quality flag. All photometric objects have an entry in the `PhotoZ` table, regardless of whether they are photometrically classified as galaxies or stars, so it is essential to consult the quality flag and error characterizations when using the photometric redshifts.

The second photometric redshift estimate uses a neural network method that is very similar in implementation to that of Collister & Lahav (2004). The training set consists of 140,000 single pass SDSS photometry measurements with spectroscopic redshifts from various sources: the SDSS (110,000 redshifts), CNOC2 (Yee et al. 2000; 9000 redshifts) CFRS (Lilly et al. 1995; 1000 redshifts), DEEP and DEEP2 (Weiner et al. 2005; 1700 redshifts), TKRS/GOODS (Wirth et al. 2004; 300 redshifts), and the 2SLAQ LRG survey (Cannon et al. 2006; 27,000 redshifts). The SDSS portion of the training set consists of a representative sampling of the SDSS Main, LRG, and southern survey spectroscopic data; the other surveys are used to augment the training set at magnitudes fainter than probed by the SDSS spectroscopic samples. Note that the training set multiply counts independent,

[1]See `http://skyserver.elte.hu/PhotoZ/` and `http://yummy.uchicago.edu/SDSS/` for details.

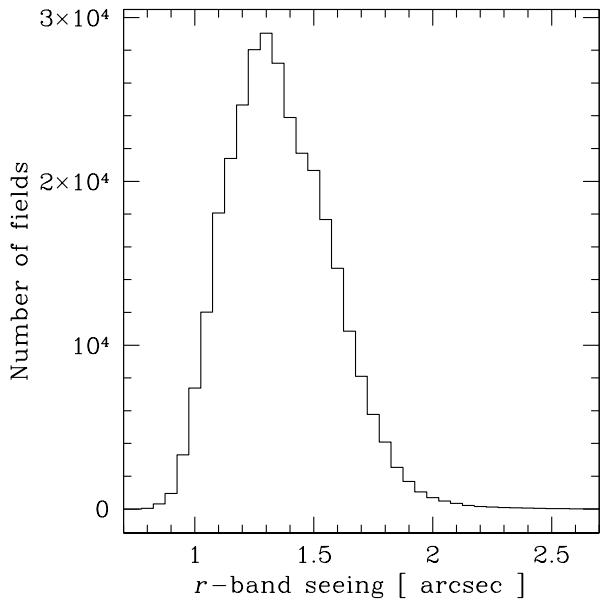

Fig. 4.— The distribution of image quality (FWHM of point sources) in the imaging survey, measured in r band.

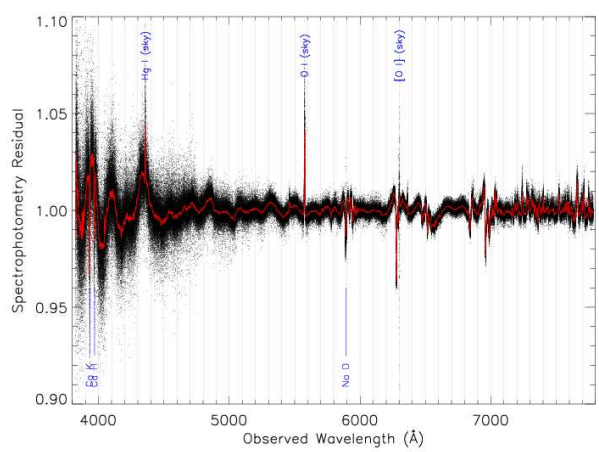

Fig. 5.— Test of spectrophotometric accuracy, performed by dividing the rest-frame spectra of elliptical galaxies observed over the redshift range $0.04 \leq z \leq 0.2$ (see text). Points show the residual inferred from 160 redshift-bin spectra (each an average of 300-1000 individual galaxies) spaced by $\Delta z = 0.01$, and the central line shows the median residual.

repeat SDSS photometric measurements of the same objects, in particular on SDSS Stripe 82. Photometric redshift errors are computed using the Nearest Neighbor Error method (NNE), which assigns to each object an error based on the photometric redshift error distribution of objects with similar magnitude and color in the training set (for which the true redshifts are known), and this approach is found to accurately predict the errors (H. Oyaizu et al., in preparation). The trained network is tested on a larger validation set consisting of 1,700,000 objects with SDSS photometry (counting independent repeat measurements) and for which spectroscopic redshifts are available. The input catalogs for these photometric redshift measurements were derived from the SDSS photo pipeline outputs, but with a few additional cuts employed to improve the star-galaxy separation, using the PSF probability and the lensing smear polarizability (Sheldon et al. 2004). The photometric sample was cut at a galaxy probability greater than 0.8, which is very stringent, and a smear polarizability less than 0.8, and further cuts on magnitude were also made; hence not all DR5 objects are included. The `Photoz2` table lists a photometric redshift, an error, and a quality flag. For objects with all five SDSS magnitudes measured, the flag is set to 0 if $r \leq 20$ or 2 if $r > 20$; photometric redshifts for flag = 2 objects are subject to larger uncertainties. Objects not satisfying the above conditions have flag set to 1 or 3 and their photometric redshifts should not be used. There are 12.6 million objects in the DR5 data set with a `Photoz2` flag of 0 and another 59.0 million with a flag of 2. In the validation set, 68% of flag = 0 galaxies have photometric redshift within 0.026 of the measured spectroscopic redshift (in the range $0.001 \leq z \leq 1.5$). The rms dispersion between photometric and spectroscopic redshifts is higher, $\sigma = 0.039$, a consequence of the non-Gaussian tails of the error distribution.

Figure 6 plots the two photometric redshift estimates against spectroscopic redshifts, and against each other, for 20,000 objects selected from the DR5 database. These are objects with SDSS spectroscopic redshifts, spectroscopically classified as galaxies, `PhotoZ` quality flag of 4 or 5, and `PhotoZ2` flag of 0 or 2. Both estimates show a tight correlation with spectroscopic redshift for the great majority of sources, while `PhotoZ` shows a somewhat larger fraction of outliers with overestimated photometric redshifts.

4.2. Regions and Sectors

Each survey observation, imaging or spectroscopic, covers a certain region of the sky. Doing statistical calculations with the SDSS data usually requires performing computations over these regions and the intersections among them, e.g., to normalize luminosity functions or calculate completeness corrections. Typical questions are: how much area do these regions

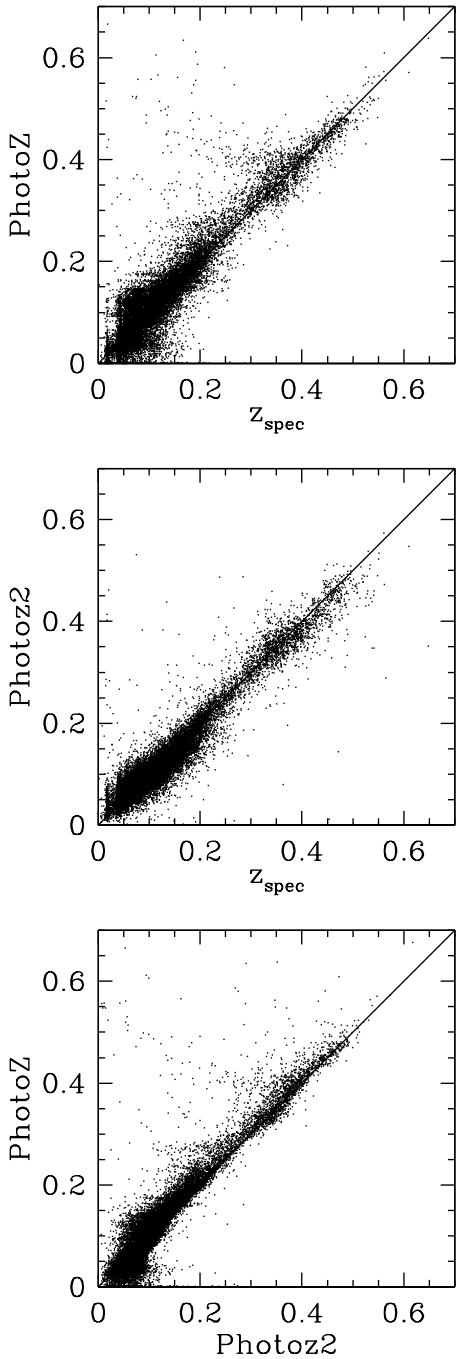

Fig. 6.— **Comparison of photometric redshift estimates** `PhotoZ` **and** `PhotoZ2` **to SDSS spectroscopic redshifts, and to each other.**

cover, how much do they overlap, and which regions contain a certain point or area of the sky? The DR5 CAS includes tables that precisely describe each region and built-in tools for finding the connections and overlaps between one kind of region and another. Each `Region` in the CAS is represented as a union of spherical polygons, and its area is analytically calculated and stored.

The SDSS has many different types of regions; they include the stripes, camera columns, segments, chunks, and spectroscopic tiles that are the basis of the SDSS observing and target selection strategy. The survey stripes overlap at the edges, with the overlap increasing towards the survey poles, so they are clipped into disjoint "staves" centered on each stripe that uniquely cover the survey area (like the staves of a barrel). The union of the staves within the survey boundaries defines the survey's "primary" photometric area. There are "holes" inside the stripes and staves, consisting of fields that were declared to be of inferior quality (e.g., because of degraded seeing or contamination by the saturated pixels of a bright star and its wings). The portions of these holes that lie within the primary survey area are called `TiHoles` to emphasize their role in the tiling process, as explained below.

As a simple example of the region tables, let us calculate the photometric survey area. Imaging data are imported to the database in "chunks," and the total area of these chunks can be obtained from the SQL (Structured Query Language) query[2]

```
select sum(area) from Region where type='CHUNK',
```

yielding 9560 deg^2. However, this counts overlapping areas more than once. To obtain the unique survey imaging footprint, we select only the "primary" region, the intersection of the chunks with the staves,

```
select sum(area) from Region where type='PRIMARY',
```

yielding 7897 deg^2. The total area and unique footprint area should be adjusted downwards by the area of the holes, obtained from the queries

```
select sum(area) from Region where type='HOLE'
```

for the chunks and

```
select sum(area) from Region where type='TIHOLE'
```

[2]See `http://cas.sdss.org/dr5/en/help/docs/sql_help.asp`. The text follows our standard capitalization conventions; for example, the various types of entries in the `Region` table (`CHUNK, TILE`, etc.) are listed in all capital letters. However, queries are not case-sensitive.

for the primary area. These queries yield 26 and 23 square degrees, respectively, making the final precise numbers for the photometric survey area 9534 deg^2 in total and 7875 unique deg^2 within the main survey boundaries. (The 8000 deg^2 figure quoted elsewhere includes a small amount of imaging outside of the ellipse that defines the main survey boundary.)

For analyses of spectroscopic samples, the issues are more complex. The SDSS spectroscopic survey aims to sample quasars and galaxies uniformly over the sky, with additional spectra for other samples (not necessarily uniform) of science targets, calibration objects, and sky. In practice, after an area has been observed by the photometric survey, a series of targeting pipelines creates lists of targets that satisfy the selection criteria. A "tiling" program (Blanton et al. 2003) runs over a subset of the observed area and assigns targets to circular "tiles" of diameter 2.98°; it also determines which targets are assigned fiber holes on which spectroscopic plugplate, imposing physical constraints such as the 55″ minimum spacing between fibers. A given run of the tiling program operates on the union of a set of "rectangular" (in spherical coordinates) `TilingGeometry` areas.

For calculations of galaxy or quasar clustering, one needs to compute the completeness of the spectroscopic sample as a function of sky position. The natural scale on which to do this is that of a `SECTOR`, a region that is covered by a unique set of `Tile` overlaps (e.g., by a particular spectroscopic plate, or by two or more plates that overlap). These are regions over which the completeness should be nearly uniform (see, e.g., Figure 1 of Percival et al. [2006] and earlier discussions by Tegmark et al. [2004] and Blanton et al. [2005]). The `Target` table lists (in the column `target.regionID`) the `SECTOR` for every object selected by the spectroscopic target selection algorithms, regardless of whether or not that object has been spectroscopically observed. To find the `SECTOR` for an object in the main table of spectroscopically observed objects, `SpecObj`, one must first identify the corresponding entry in the `Target` table. For example, the following query

```
select top 10 s.specObjID, t.regionID
from SpecObj s join Target t
  on s.targetID = t.targetID
```

returns the spectroscopic ID numbers and the `SECTOR` numbers of the first ten objects encountered in the `SpecObj` table. The database function `fRegionsContainingPointEQ` can be used to find the `SECTOR` that covers a specified point on the sky.

The following practical example illustrates several other features of these tables. The SDSS quasar target selection algorithm underwent significant changes in the early phases of the survey, reaching its final form (Richards et al. 2002) with `targetVersion` 3.1.0, following

DR1. A calculation of the quasar luminosity function should therefore be restricted to regions targeted with this or subsequent versions of the target selection code, and it should be normalized using the corresponding area, which the following query shows to be 4013 deg^2:

```
select sum(area)
from Region
where regionID in (
      select b.boxID
      from Region2Box b join TilingGeometry g
            on b.id = g.tilingGeometryID
      where b.boxType = 'SECTOR'
            and b.regionType = 'TIPRIMARY'
      group by b.boxID
      having min(g.targetVersion) >= 'v3_1_0'
)
```

This query uses the `Region2Box` table, which maps between various types of `Regions` and the `TilingGeometries` in which information about the target selection is stored. The `where` clause selects, from the table of all `Regions`, those which are `SECTOR`s in the primary tiled area and were targeted with a final version of the quasar target selection algorithm.[3]

In principle, these tables provide all the information needed for complex clustering calculations — e.g., determining local completeness corrections, generating appropriate catalogs of randomly distributed points, and identifying targeted objects that were not observed because of the minimum fiber spacing constraint. The queries required for such calculations are rather lengthy, and will be presented and documented elsewhere.

4.3. Match Tables

About 50 million photometric objects in the CAS lie in regions that have been observed more than once, because of stripe overlap or repeat scans. These repeat observations can be used to detect variable and moving objects. The `MatchHead` and `Match` tables of the

[3]This query is included as one of the sample queries in the DR5 documentation, under "Uniform Quasar Sample," together with a longer query that shows how to extract all quasars and quasar candidates from the corresponding sky area.

DR5 CAS provide convenient tools to examine the multiple observations of a single object, identified by positional matches with a 1″ tolerance, and collectively referred to as a *bundle*. The `MatchHead` table has the unique ID of the first object in the bundle (defined by observation date), the mean and variance of the coordinates of all matched detections, the number of matched detections, and the number of times the object was "missed" in other observations of the same sky area. Misses can occur because the object is variable, because it is moving, because inferior seeing moves it below the detection threshold, or because the original detection was spurious. The `Match` table lists all objects in each bundle.

As an example, the following query lists information about the multiple detections of an object at (ra,dec)=(194,0):

```
select MH.*
   from MatchHead MH
   join fGetNearbyObjEq(194,0,0.3) N on MH.objID = N.objID
```

The `fGetNearbyObjEq` function returns a table (assigned the name `N`) of all objects found within 0.3 arc-minutes of the desired coordinates. The `select` command returns all entries in the `matchHead` table (assigned the name `MH`) which, as a result of the `join` command, have an object ID that matches one returned by the neighborhood search. In this case, there is just one such match, hence a single bundle. One can get information on all the objects in the bundle with the query

```
select M.*
   from Match M
   join MatchHead MH on M.matchHead = MH.objID
   join fGetNearbyObjEq(194,0,0.3) N on MH.objID = N.objID
```

where the new `join` command selects out those `Match` tables whose `matchHead` agrees with that returned by the earlier query.

The DR5 CAS has 50,627,023 bundles described by `MatchHead` and 109,441,410 objects in the `Match` table. When an object is undetected in a repeat observation of the same area of sky, a surrogate object is placed in the `Match` table but marked as a "miss," with an additional flag to indicate if the miss could be caused by masking of the region in the second observation (e.g., because of a satellite trail or cosmic ray hit) or because it lies near the edge of the overlap region. A bundle may therefore consist of a single detection and one or more surrogates (and the object in the `MatchHead` may be a surrogate). There are 9.8 million

surrogates in the `Match` table. The presence of surrogate objects may simplify algorithmic searches for moving or variable objects.

Because the multiple imaging scans of the southern equatorial stripe are not yet in the CAS, the `Match` tables cannot be used to search for moving or variable objects in these data. However, this capability will be present in future data releases.

5. Conclusions

The Fifth Data Release of the Sloan Digital Sky Survey provides access to 8000 deg^2 of five-band imaging data and over one million spectra. These data represent a roughly 20% increment over the previous data release (DR4, Adelman-McCarthy et al. 2006). Both the catalog data and the source imaging data are available via the Internet. All the data products have been consistently processed by the same set of pipelines across several data releases. The previous data releases remain online and unchanged to support ongoing science studies. DR5 includes several qualitatively new features: multiple imaging scans of the southern equatorial stripe, special imaging scans of M31 and the Perseus cluster, database access to QSO catalogs and galaxy photometric redshifts, and database tools for precisely defining the survey geometry and for linking repeat imaging observations of matched objects. More than a thousand scientific publications have been based on the SDSS data to date, spanning an enormous range of subjects. Future data releases will increase the survey area, and they will provide qualitatively new kinds of data on the stellar kinematics and populations of the Milky Way and on Type Ia supernovae and other transient or variable phenomena, further extending this scientific impact.

We dedicate this paper to our colleague Jim Gray, who disappeared in January, 2007, while sailing near San Francisco. Jim dedicated an enormous amount of his time, his energy, and his remarkable talents to the SDSS over the course of many years. He played a critical role in the development of the SDSS database, including important contributions to the writing of this paper.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating

Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

Abazajian, K. et al. 2003, AJ, 126, 2018 (DR1)

Abazajian, K. et al. 2004, AJ, 128, 502 (DR2)

Abazajian, K. et al. 2005, AJ, 129, 1755 (DR3)

Adelman-McCarthy, J. K. et al. 2006, ApJS, 162, 38 (DR4)

Anderson, S. et al. 2003, AJ, 126, 2209

Baldry, I.K. et al. 2005, MNRAS, 358, 441

Blanton, M.R., Lin, H., Lupton, R.H., Maley, F.M., Young, N., Zehavi, I., & Loveday, J. 2003, AJ, 125, 2276

Blanton, M.R. et al. 2005, AJ, 129, 2562

Budavari, T. et al. 2000, AJ, 120, 1588

Cannon, R., et al. 2006, MNRAS, in press, astro-ph/0607631

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393

Collinge, M. et al. 2005, AJ, 129, 2542

Collister, A. A. & Lahav, O. 2004, PASP, 116, 345

Cool, R. J. et al. 2006, AJ, 131, 73Z

Csabai, I. et al. 2003, AJ, 125, 580

Eisenstein, D.J. et al. 2001, AJ, 122, 2267

Eisenstein, D. J., et al. 2006, ApJS, 167, 40

Finkbeiner, D.P. et al. 2004, AJ, 128, 2577

Fukugita, M., Ichikawa, T., Gunn, J.E., Doi, M., Shimasaku, K., & Schneider, D.P. 1996, AJ, 111, 1748

Gunn, J.E. et al. 1998, AJ, 116, 3040

Gunn, J.E. et al. 2006, AJ, 131, 2332

Hogg, D.W., Finkbeiner, D.P., Schlegel, D.J., & Gunn, J.E. 2001, AJ, 122, 2129

Ivezić, Ž. et al. 2003, MmSAI, 74, 978

Ivezić, Ž. et al. 2004, Astronomische Nachrichten, 325, 583

Jiang, L. et al. 2006, AJ, 131, 2788

Lilly, S.J., et al. 1995, ApJ, 455, 50

Lupton, R.H. 2005, AJ, submitted

Lupton, R.H., Gunn, J.E., & Szalay, A.S. 1999, AJ, 118, 1406

Lupton, R.H., Gunn, J.E., Ivezić, Ž., Knapp, G.R., Kent, S., & Yasuda, N. 2001, in *Astronomical Data Analysis Software and Systems X*, edited by F. R. Harnden Jr., F. A. Primini, and H. E. Payne, ASP Conference Proceedings, 238, 269

Mandelbaum, R. et al. 2005, MNRAS, 361, 1287

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

Percival, W. J. et al. 2006, ApJ, in press, astro-ph/0608636

Petrosian, V. 1976, ApJ, 209, L1

Pier, J.R., Munn, J.A., Hindsley, R.B., Hennessy, G.S., Kent, S.M., Lupton, R.H., & Ivezić, Ž. 2003, AJ, 125, 1559

Richards, G.T. et al. 2002, AJ, 123, 2945

Schneider, D.P. et al. 2005, AJ, 130, 367

Scranton, R. et al. 2005, ApJ, submitted, astro-ph/0508564

Sheldon, E.S., et al. 2004, AJ, 127, 2544

Smith, J.A. et al. 2002, AJ, 123, 2121

Stoughton, C. et al. 2002, AJ, 123, 485

Strauss, M.A. et al. 2002, AJ, 124, 1810

Tegmark, M. et al. 2004, ApJ, 606, 702

Tucker, D. et al. 2006, Astronomische Nachrichten, 327, 821

Vanden Berk, D.E. et al. 2005, AJ, 129, 2047

Weiner, B.J., et al. 2005, ApJ, 620, 595

Wirth, G.D., et al. 2004, AJ, 127, 3121

Yee, H.K.C., et al. 2000, ApJS, 129, 475

York, D.G. et al. 2000, AJ, 120, 1579

Zehavi, I. et al. 2002, ApJ, 571, 172

Zucker, D. et al. 2004a, ApJ, 612, L117

Zucker, D. et al. 2004b, ApJ, 612, L121

This preprint was prepared with the AAS LaTeX macros v5.2.