Exploring the Variable Sky with Sloan Digital Sky Survey

Branimir Sesar[1], Željko Ivezić[1], Robert H. Lupton[2], Mario Jurić[3], James E. Gunn[2], Gillian R. Knapp[2], Nathan De Lee[4], J. Allyn Smith[5], Gajus Miknaitis[6], Huan Lin[6], Douglas Tucker[6], Mamoru Doi[7], Masayuki Tanaka[8], Masataka Fukugita[9], Jon Holtzman[10], Steve Kent[6], Brian Yanny[6], David Schlegel[10], Douglas Finkbeiner[11], Nikhil Padmanabhan[10], Constance M. Rockosi[12], Nicholas Bond[2], Brian Lee[10], Chris Stoughton[6], Sebastian Jester[13], Hugh Harris[14], Paul Harding[15], Jon Brinkmann[16], Donald P. Schneider[17], Donald York[18]

ABSTRACT

We quantify the variability of faint unresolved optical sources using a catalog based on multiple SDSS imaging observations. The catalog covers the so-called

[1]University of Washington, Dept. of Astronomy, Box 351580, Seattle, WA 98195-1580

[2]Princeton University Observatory, Princeton, NJ 08544-1001

[3]Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540

[4]Dept. of Physics & Astronomy, Michigan State University, East Lansing, MI 48824-2320

[5]Dept. of Physics & Astronomy, Austin Peay State University, Box 4608, Clarksville, TN 37044

[6]Fermi National Accelerator Laboratory, Box 500, Batavia, IL 60510

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

[8]Dept. of Astronomy, Graduate School of Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-0033, Japan

[9]Institute for Cosmic Ray Research, University of Tokyo, Kashiwa, Chiba, Japan

[10]New Mexico State University, Box 30001, 1320 Frenger St., Las Cruces, NM 88003

[10]Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 50R5032, Berkeley, CA, 94720

[11]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

[12]University of California–Santa Cruz, 1156 High St., Santa Cruz, CA 95060

[13]School of Physics and Astronomy, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

[14]U.S. Naval Observatory, Flagstaff Station, Box 1149, Flagstaff, AZ 86002

[15]Department of Astronomy, Case Western Reserve University, Cleveland, Ohio 44106

[16]Apache Point Observatory, 2001 Apache Point Road, Box 59, Sunspot, NM 88349-0059

[17]Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA 16802

[18]University of Chicago, Astronomy & Astrophysics Center, 5640 S. Ellis Ave., Chicago, IL 60637

equatorial Stripe 82 (22h 24m $< \alpha_{J2000} <$ 04h 08m, $-1.27° < \delta_{J2000} < +1.27°$, ~ 290 deg^2), and contains 58 million photometric observations in the SDSS $ugriz$ system for 1.4 million sources that were observed at least 4 times in each of the gri bands (with a median of 10 observations obtained over $\sim$5 years). In each photometric bandpass we compute various low-order lightcurve statistics such as root-mean-square scatter (rms), χ^2 per degree of freedom, skewness, minimum and maximum magnitude, and use them to select and study variable sources. We find that 2% of unresolved optical sources brighter than $g = 20.5$ appear variable at the 0.05 mag level (rms) simultaneously in the g and r bands. The majority (2/3) of these variable sources are low-redshift (< 2) quasars, although they represent only 2% of all sources in the adopted flux-limited sample. We find that about 1/4 of variable stars are RR Lyrae stars, and that only 0.5% of stars from the main stellar locus are variable at the 0.05 mag level. We analyze the distribution of lightcurve skewness in the $g-r$ vs. $u-g$ color-color diagram and show that its distribution on the main stellar locus is bimodal (with one mode consistent with Algol-like behavior), while for quasars the skewness distribution is centered on zero. Using over six hundred RR Lyrae stars, we demonstrate rich halo substructure out to distances of 100 kpc. We find that at least 90% of quasars are variable at the 0.03 mag level (rms) and confirm that variability is as good a method for finding low-redshift quasars as is the UV excess color selection. We extrapolate these results to expected performance by the Large Synoptic Survey Telescope and estimate that it will obtain well-sampled 2% accurate multi-color lightcurves for $\sim$ 2 million low-redshift quasars, and will discover at least 50 million variable stars.

Subject headings: Galaxy: halo — Galaxy: stellar content — variables: RR Lyrae variable

1. Introduction

Variability is an important phenomenon in astrophysical studies of structure and evolution, both stellar and galactic. Some variable stars, such as RR Lyrae, are an excellent tool for studying the Galaxy. Being nearly standard candles (thus making distance determination relatively straightforward) and being intrinsically bright, they are a particularly suitable tracer of Galactic structure. In extragalactic astronomy, the optical continuum variability of quasars is utilized as an efficient method for their discovery (van den Bergh, Herbst & Pritchet 1973; Hawkins 1983; Hawkins & Veron 1995), and is also frequently used to con-

strain the origin of their emission (Kawaguchi et al. 1998; Trevese et al. 2001; Martini & Schneider 2003).

Despite the importance of variability, variable optical sky remains largely unexplored and poorly quantified, especially at the faint end. To what degree different variable populations contribute to the overall variability, how they distribute in the magnitude and color space, what the characteristic time-scales and the dominant mechanisms of variability are, are just some of the questions that still remain to be answered. To address these questions, several contemporary projects aimed at regular monitoring of the optical sky were started. Some of the more prominent surveys in terms of the sky coverage, depth, and cadence are:

- The Faint Sky Variability Survey (Groot et al. 2003) is a very deep ($17 < V < 24$) BVI survey of 23 deg^2 of sky, containing about 80,000 sources sampled at timescales ranging from minutes to years.
- The QUEST Survey (Vivas et al. 2001) monitors 700 deg^2 of sky from $V = 13.5$ to a limit of $V = 21$.
- ROTSE-I (Akerlof et al. 2000) monitors the entire observable sky twice a night from $V = 10$ to a limit of $V = 15.5$. Northern Sky Variability Survey (Woźniak et al. 2004) is based on ROTSE-I data.
- OGLE (most recently OGLE III; Udalski et al. 2002) monitors ~ 100 deg^2 towards the Galactic bulge from $I = 11.5$ to a limit of $I = 20$. Due to a very high stellar density towards the bulge, OGLE II has detected over 200,000 variable stars (Woźniak et al. 2002)

A more comprehensive review of past and ongoing variability surveys can be found in Becker et al. (2004).

Recognizing the outstanding importance of variable objects, the last Decadal Survey Report highly recommended a major new initiative for studying the variable sky, the Large Synoptic Survey Telescope (LSST; Tyson et al. 2002, Walker 2003). The LSST[1] will offer an unprecedented view of the faint variable sky: according to the current designs it will scan the entire accessible sky every three nights to a limit of $V \sim 25$ with two observations per night in two different bands. One of the LSST science goals[2] will be the exploration of the

[1]See http://www.lsst.org

[2]For more details see http://www.lsst.org/Science/science_goals.shtml

transient optical sky: discovery and analysis of rare and exotic objects (e.g. neutron star and black hole binaries), gamma-ray bursts, X-ray flashes, and of new classes of transients, such as binary mergers and stellar disruptions by black holes. The observed volume of space, and the requirement to recognize and monitor these events — in real time — on a "normally" variable sky, will present a challenge to the project.

Since LSST will utilize[3] the Sloan Digital Sky Survey (SDSS; York et al. 2000) photometric system (*ugriz*, Fukugita et al. 1996; Gunn et al. 1998), multiple photometric observations obtained by SDSS represent an excellent dataset for a pre-LSST study that characterizes the faint variable sky and quantifies the variable population and its distribution in the magnitude-color-variability space. Here we present such a study.

In Section 2 we give a brief overview of SDSS imaging survey and repeated scans of a $\sim$ 290 deg^2 large region called Stripe 82. In Section 3, we describe methods used to select candidate variable sources from the SDSS-I Stripe 82 data assembled, averaged and recalibrated by Ivezić et al. (2007), and present tests that show the robustness of the adopted selection criteria. In the same section, we discuss the distribution of selected variable sources in the magnitude-color-variability space. The Milky Way halo structure traced by selected candidate RR Lyrae stars is discussed in Section 4, and in Section 5 we estimate the fraction of variable quasars. Implications for surveys such as LSST are discussed in Section 6, and our main results are summarized in Section 7.

2. Overview of SDSS Imaging and Stripe 82 Data

The quality of photometry and astrometry, as well as the large area covered by the survey, make SDSS stand out among available optical sky surveys (Sesar et al. 2006). SDSS is providing homogeneous and deep ($r < 22.5$) photometry in five bandpasses (u, g, r, i, and z, Gunn et al. 1998; Hogg et al. 2002; Smith et al. 2002; Gunn et al. 2006; Tucker et al. 2006) accurate to 0.02 mag (root-mean-square scatter, hereafter rms, for sources not limited by photon statistics, Ivezić et al. 2003a; Scranton et al. 2002; and with a zeropoint uncertainty of 0.02 mag, Ivezić et al. 2004a). The survey sky coverage of 10,000 deg^2 in the northern Galactic cap, and 300 deg^2 in the southern Galactic cap will result in photometric measurements for well over 100 million stars and a similar number of galaxies. The recent Data Release 5 (Adelman-McCarthy et al. 2007)[4] lists photometric data for 215 million

[3]LSST will also use the Y band at ~ 1 μm. For more details see the LSST Science Requirement Document at http://www.lsst.org/Science/lsst_baseline.shtml

[4]Please see http://www.sdss.org/dr5

unique objects observed in 8000 deg^2 of sky as part of "SDSS-I" phase that ran through June 2005. Astrometric positions are accurate to better than $0.1''$ per coordinate (rms) for sources with $r < 20.5$ (Pier et al. 2003), and the morphological information from the images allows reliable star-galaxy separation to $r \sim 21.5$ (Lupton et al. 2002). In addition, the 5-band SDSS photometry can be used for very detailed source classification; e.g. separation of quasars and stars (Richards et al. 2002), spectral classification of stars to within 1-2 spectral subtypes (Lenz et al. 1998; Finlator 2000; Hawley et al. 2002), and even remarkably efficient color selection of the Horizontal Branch and RR Lyrae stars (Yanny et al. 2000; Sirko et al. 2004; Ivezić et al. 2005) and low-metallicity G and K giants (Helmi et al. 2003).

The equatorial Stripe 82 region (22h 24m $< \alpha_{J2000} <$ 04h 08m, $-1.27° < \delta_{J2000} < +1.27°$, ~ 290 deg^2), observed in the southern Galactic cap, presents a valuable data source for variability studies. The region was repeatedly observed (65 imaging runs by July 2005, but not all cover the entire region), and it is the largest source of multi-epoch data in the SDSS. By averaging the repeated observations of Stripe 82 sources, a more accurate photometry than the nominal 0.02 mag single-scan accuracy can be achieved. This motivated Ivezić et al. (2007) to produce a catalog of recalibrated Stripe 82 observations. The catalog lists 58 million photometric observations for 1.4 million sources that were observed at least 4 times in each of the *gri* bands (with a median of 10 observations obtained over ~ 5 years). The random photometric errors are below 0.01 mag for stars brighter than (19.5, 20.5, 20.5, 20, 18.5) in *ugriz*, respectively (about twice as good as for individual SDSS runs), and the spatial variation of photometric zeropoints is not larger than $\sim$0.01 mag (rms). In addition, various low-order statistics such as root-mean-square scatter (Σ), χ^2 per degree of freedom (χ^2), skewness, minimum and maximum magnitude, were computed for each *ugriz* band and each source.

Separation of quasars and stars, as well as efficient color selection of Horizontal Branch and RR Lyrae stars, depend on accurate u band photometry. To ensure this, we select 748,084 sources from the Ivezić et al. (2007) catalog with at least 4 observations in the u band. A catalog of variable sources selected from this sample is analyzed in Section 3.

3. Analysis of Stripe 82 Catalog of Variable Sources

In this section we describe methods for selecting candidate variable sources, and present tests that show robustness of the adopted selection criteria. The distribution of selected variable sources in the magnitude-color-variability space is also presented and discussed.

3.1. Methods and Selection Criteria

Due to a relatively small number of observations per source and random sampling, we do not perform lightcurve fitting, but instead use low order statistics to select candidate variables and study their properties. There are four parameters (median magnitude, root-mean-square scatter Σ, χ^2, and lightcurve skewness) measured in five photometric bands (u, g, r, i, and z), for a total of 20 parameters. In the analysis presented here, we utilize eight of them:

- median magnitudes in the ugr bands (corrected for interstellar extinction using the map from Schlegel, Finkbeiner & Davis 1998) because the $g-r$ vs. $u-g$ color-color diagram has the most classification power (e.g. Smolčić et al. 2004 and references therein)
- Σ and χ^2 in the g and r bands, and
- lightcurve skewness in the g band, which has the highest signal-to-noise.

The dependence of Σ on magnitude in the $ugriz$ bands, is shown in Figure 1. For sources brighter than 18, 19.5, 19.5, 19, and 17.5 mag in the $ugriz$, respectively, SDSS delivers 2% photometry with little or no dependence on magnitude. The observed rms scatter Σ includes both the intrinsic variability and the measurement error. We determine the latter by fitting a fourth degree polynomial to median Σ values in 0.5 mag wide bins. We define the magnitude-independent rms scatter σ as

$$\sigma = (\Sigma^2 - \Sigma_{fit}^2)^{1/2} \quad (1)$$

for $\Sigma > \Sigma_{fit}$, and $\sigma = 0$ otherwise.

As the first variability selection criterion, we adopt $\sigma(g) \geqslant 0.05$ mag and $\sigma(r) \geqslant 0.05$ mag (hereafter written as $\sigma(g,r) \geqslant 0.05$ mag). At the bright end, this criterion is equivalent to selecting sources with rms scatter greater than $2.5\sigma_0$, where $\sigma_0 = 0.02$ mag is the measurement noise. Selection cuts are applied simultaneously in the g and r bands to reduce the number of "false positives" (intrinsically non-variable sources selected as candidate variable sources due to measurement noise). About 6% of sources pass the σ cut in each band separately, and $\sim$ 3% of sources pass the cut in both bands simultaneously. By selecting sources with $\sigma(g,r) \geqslant 0.05$ mag, we also select faint sources that have large σ due to large photometric errors at the faint end. To only select faint sources with statistically significant rms scatter, we apply the χ^2 test as the second selection cut.

In the χ^2 test, the value of χ^2 per degree of freedom, calculated with respect to a weighted mean magnitude and using errors computed by the photometric pipeline, determines whether the observed lightcurve is consistent with the Gaussian distribution of errors.

Large χ^2 values show that the rms scatter is inconsistent with random fluctuations. Ivezić et al. (2003a) and Ivezić et al. (2007) used multi-epoch SDSS observations to show that the photometric error distribution in SDSS roughly follows a Gaussian distribution. A comparison of χ^2 distributions in the g and r bands with a reference Gaussian χ^2 distribution, is shown in Figure 2. As evident, χ^2 distributions in both bands roughly follow the reference Gaussian χ^2 distribution for $\chi^2 < 1$, demonstrating that median photometric errors are correctly determined. The discrepancy for larger χ^2 is due to variable sources rather than non-Gaussian error distribution, as we demonstrate below.

The second selection cut, $\chi^2(g) \geqslant 3$ and $\chi^2(r) \geqslant 3$ (hereafter written as $\chi^2(g,r) \geqslant 3$), selects $\sim 90\%$ of $\sigma(g,r) \geqslant 0.05$ mag sources, as shown in Figure 2 (middle panels). The effectiveness of the χ^2 test is demonstrated in the bottom panel of Figure 2. For magnitudes fainter than $g = 20.5$, the fraction of candidate variables decreases as photometric errors increase. The selection is relatively uniform for sources brighter than $g = 20.5$, and we adopt this value as the flux limit for the selected variable sample.

There are 662,195 sources brighter than $g = 20.5$ in the full sample. Using $\sigma(g,r) \geqslant 0.05$ mag and $\chi^2(g,r) \geqslant 3$ as the selection criteria, we select 13,051 candidate variable sources. Therefore, *at least* 2% *of unresolved optical sources brighter than* $g = 20.5$ *appear variable at the* $\geqslant 0.05$ *mag level (rms) simultaneously in the* g *and* r *bands.*

3.2. The Counts of Variable Sources

In this section we estimate the completeness and efficiency of the candidate variable sample, and discuss the dependence of counts, rms scatter, $\sigma(g)/\sigma(r)$ ratio, and the g band lightcurve skewness on the position in the $g-r$ vs. $u-g$ color-color diagram.

3.2.1. Completeness

The selection completeness, defined as the fraction of true variable sources recovered by the algorithm, depends on the lightcurve shape and amplitudes. Due to a fairly large number of observations (median of 10), and small $\sigma(g,r)$ cutoff compared to typical amplitudes of variable sources (e.g. most RR Lyrae stars and quasars have peak-to-peak amplitudes $\lesssim 1$ mag), we expect the completeness to be fairly high for RR Lyrae stars ($\gtrsim 99\%$, see Section 4) and quasars ($\sim 90\%$, see Section 5). The completeness for other types of variable sources, such as flares and eclipsing binaries, is hard to estimate, but is probably low due to sparse sampling.

3.2.2. Efficiency

The selection efficiency, defined as the fraction of true variable sources in the candidate variable sample, determines the robustness of the selection algorithm. The main diagnostic for the robustness of the adopted selection criteria is the distribution of selected candidates in SDSS color-magnitude and color-color diagrams. The position of a source in these diagrams is a good proxy for its classification (Lenz et al. 1998; Fan 1999; Finlator 2000; Smolčić et al. 2004).

Figure 3 compares the distribution of candidate variable sources to that of all sources in the $g - r$ vs. $u - g$ color-color diagram. Were the selection a random process, the selected candidates would have the same distribution as the full sample. The distributions of candidate variables and of the full sample are remarkably different, demonstrating that the candidate variables are *not* randomly selected from the parent sample.

The three dominant classes of variable object are quasars, RR Lyrae stars, and stars from the main stellar locus. The most obvious difference between the two distributions is a much higher fraction of low-redshift quasars (< 2.2, recognized by their UV excess, $u - g < 0.7$, see Richards et al. 2002) and RR Lyrae stars ($u - g \sim 1.15$, $g - r < 0.3$, see Ivezić et al. 2005) in the variable sample, than in the full sample.

In order to quantify these differences, as well as those in other parts of the color-color diagram, we follow Sesar et al. (2006) and divide the $g - r$ vs. $u - g$ color-color diagram into six characteristic regions, each dominated by a particular type of source, as shown in Figure 4. The fractions and counts of variable and all sources in each region are listed in Table 1 for $g < 19$, $g < 20.5$, and $g < 22$ flux-limited samples. Notably, in the adopted $g < 20.5$ flux limit, the fraction of Region II sources (dominated by low-redshift quasars) in the variable sample is 63%, or ~ 30 times greater than the fraction of Region II sources in the full sample ($\sim 2\%$). The fraction of Region IV sources (which include RR Lyrae stars) in the variable sample is also high when compared to the full sample (~ 6 times higher).

As shown in Table 1, in the $g = 20.5$ flux-limited sample, we find that low-redshift quasars and RR Lyrae stars (i.e. Regions II and IV) make 70% of the variable population, while representing only 3% of all sources. Quasars alone account for 63% of the variable population. Stars from the main stellar locus represent 95% of all sources and 25% of the variable sample: about 0.5% of the stars from the locus are variable at the $\geqslant 0.05$ mag level.

3.3. The Properties of Variable Sources

Various lightcurve properties, such as shape and amplitude, are expected to be correlated with stellar types. In this section we study the distribution of the rms scatter in the u and g bands, and $\sigma(g)/\sigma(r)$ ratio as a function of the $u-g$ and $g-r$ colors. To emphasize trends, we bin sources and present median values for each bin.

The distribution of the median $\sigma(u)$ and $\sigma(g)$ values in the $g-r$ vs. $u-g$ color-color diagram is shown in the top two panels of Figure 5. RR Lyrae stars show larger rms scatter ($\gtrsim 0.3$ mag) in the u and g bands, than low-redshift quasars or stars from the main stellar locus. Quasars also show slightly larger rms scatter in the u band (~ 0.1 mag) than in the g band (~ 0.07 mag). If we define the degree of variability as the root-mean-square scatter in the g band, then on average RR Lyrae stars show the greatest variability, followed by quasars and the main stellar locus stars.

Another distinctive characteristic of variable sources is the ratio of flux changes in different bandpasses. This property can be used to select different types of variable sources. For example, RR Lyrae stars are bluer when brighter, a behavior used by Ivezić et al. (2000) to select RR Lyrae using 2-epoch SDSS data. Here we derive a new parameter, $\sigma(g)/\sigma(r)$, to express the ratio of flux changes in the g and r bands, and study its distribution in the $g-r$ vs. $u-g$ color-color diagram. In particular, we examine this distribution and its median values for three dominant classes of variable sources: quasars, RR Lyrae stars, and stars from the main stellar locus.

The bottom left panel in Figure 5 shows the distribution of median $\sigma(g)/\sigma(r)$ values as a function of $u-g$ and $g-r$ colors. Using Fig. 5 we note that on average:

- RR Lyrae stars have $\sigma(g)/\sigma(r) \sim 1.4$
- Main stellar locus stars have $\sigma(g)/\sigma(r) \sim 1$, and
- Quasars show a $\sigma(g)/\sigma(r)$ gradient in the $g-r$ vs. $u-g$ color-color diagram.

The average value of $\sigma(g)/\sigma(r) \sim 1.4$ in Region IV indicates that RR Lyrae stars dominate the variable source count in this region. The ratio of 1.4 for RR Lyrae stars was also previously found by Ivezić et al. (2000). While Figure 5 only presents median values of the rms scatter, Figure 6 shows how the rms scatter in the g and r bands correlates with the $u-g$ color for individual sources. Variable sources that follow the $\sigma(g) = 1.4\sigma(r)$ relation also correlate with the $u-g$ color, and have $u-g \sim 1$, as expected for RR Lyrae stars.

The average ratio of $\sigma(g)/\sigma(r) \sim 1$ (i.e. gray flux variations) for stars in the main stellar locus suggests that the variability could be caused by eclipsing systems. Distribution

of the lightcurve skewness in the g band for main stellar locus stars further strengthens this possibility, as discussed in Section 3.4 below.

Gradient in the $\sigma(g)/\sigma(r)$ observed for low-redshift quasars in the $g-r$ vs. $u-g$ color-color diagram suggests that the variability correlation between the g and r bands is more complex than in the case of RR Lyrae or main stellar locus stars. Wilhite et al. (2006) show that the photometric color changes for quasars depend on the combined effects of continuum changes, emission-line changes, redshift, and the selection of photometric bandpasses. They note that due to the lack of variability of the lines, measured photometric color is not always bluer in brighter phases, but depends on redshift and the filters used. To verify the dependence of broad-band photometric variability on redshift, we plot $\sigma(g)/\sigma(r)$ vs. redshift for all spectroscopically confirmed quasars from Schneider et al. (2005) which are in Stripe 82, as shown in Figure 7. We confirm that the broad-band photometric variability depends on the redshift, and that the $\sigma(g)/\sigma(r)$ gradient in the $g-r$ vs. $u-g$ color-color diagram can be explained by the increase in $\sigma(g)/\sigma(r)$ from ~ 1 to ~ 1.6 in the 1.0 to 1.6 redshift range. The implied correlation of the $u-g$ and $g-r$ colors with redshift is consistent with the discussion in Richards et al. (2002). The lack of noticeable correlation of $\sigma(g)$ with redshift is due to combined effects of the dependence of $\sigma(g)$ on the rest-frame wavelength and time which cancel out (for a detailed model see Ivezić et al. 2004b).

3.4. Skewness as a Proxy for Dominant Variability Mechanism

Lightcurve skewness, as a measure of the lightcurve asymmetry, provides additional information on the type of variability[5]. Negatively skewed, asymmetric lightcurves indicate variable sources that spend more time fainter than $(m_{min}+m_{max})/2$, where m_{min} and m_{max} are magnitudes at the minimum and maximum. Type ab RR Lyrae stars, for example, have negatively skewed lightcurves (skewness ~ -0.5, Wils, Lloyd & Bernhard 2006). Positively skewed, asymmetric lightcurves indicate variable sources that spend more time brighter than $(m_{min}+m_{max})/2$ (e.g. eclipsing systems). Sources with symmetric lightcurves will have skewness ~ 0.

Bottom right panel in Figure 5 shows the distribution of the median lightcurve skewness in the g band as a function of the position in the $g-r$ vs. $u-g$ color-color diagram. On average, quasars and c type RR Lyrae stars ($u-g \sim 1.15$, $g-r < 0.15$) have skewness $g \sim 0$, ab type RR Lyrae ($u-g \sim 1.15$, $g-r > 0.15$) have negative skewness (skewness $g \sim -0.5$), and stars in the main stellar locus have positive skewness.

[5]Equations used in calculating skewness are listed in the Appendix A

Figure 8 shows the distribution of the lightcurve skewness in the ugi bands for spectroscopically confirmed quasars from Schneider et al. (2005) which are in Stripe 82, candidate RR Lyrae stars (selection details are discussed in Section 4 below), and main stellar locus stars from our variable sample. Stars in the main stellar locus show bimodal skewness distribution in the g band. This distribution suggests at least two, and perhaps more, different populations of variables. Similar bimodality is also discernible in the r band, while it is less pronounced in the i band and not detected in the u and z bands.

A comparison of $u-g$ and $g-r$ color distributions for variable main stellar locus stars brighter than $g = 19$ and a subset with highly asymmetric lightcurves (skewness $g \sim 2.5$) is shown in Figure 9. The subset with asymmetric lightcurves has an increased fraction of stars with colors $u-g \sim 2.5$ and $g-r \sim 1.4$, that correspond to M stars. This may indicate that M stars have a higher probability of being associated with an eclipsing companion than stars with earlier spectral types. However, the selection effects due to low luminosity of M dwarfs are probably important. Quasars have symmetric lightcurves (skewness ~ 0) and their distribution of skewness does not change between bands.

4. The Milky Way Halo Structure Traced by Candidate RR Lyrae Stars

Studies of substructures in the Galactic halo, such as clumps and streams, can help constrain the formation history of the Milky Way. One of the best tracers to study the outer halo are RR Lyrae stars because they are nearly standard candles, are sufficiently bright to be detected at large distances ($5-100$ kpc for $14 < r < 20.7$), and are sufficiently numerous to trace the halo substructure with a high spatial resolution.

In this section we fine tune criteria for selecting candidate RR Lyrae stars, and estimate the selection completeness and efficiency. Using selected candidate RR Lyrae stars, we recover a known halo clump associated with the Sgr dwarf tidal stream, and find several new halo substructures.

4.1. Criteria for Selecting RR Lyrae Stars

Figures 3, 4, and 5 show that RR Lyrae stars occupy a well-defined region (Region IV) in the $g-r$ vs. $u-g$ color-color diagram, and Figure 6 shows how RR Lyrae stars follow the $\sigma(g) = 1.4\sigma(r)$ relation. Motivated by these results, we introduce color and $\sigma(g)/\sigma(r)$ cuts to specifically select candidate RR Lyrae stars from the variable sample, and study their distribution in the rms scatter-color-lightcurve skewness parameter space.

RR Lyrae stars have distinctive colors and can be selected with the following criteria (Ivezić et al. 2005):

$$0.98 < u - g < 1.30 \tag{2}$$

$$-0.05 < D_{ug} < 0.35 \tag{3}$$

$$0.06 < D_{gr} < 0.55 \tag{4}$$

$$-0.15 < r - i < 0.22 \tag{5}$$

$$-0.21 < i - z < 0.25 \tag{6}$$

where

$$D_{ug} = (u - g) + 0.67(g - r) - 1.07 \tag{7}$$

and

$$D_{gr} = 0.45(u - g) - (g - r) - 0.12. \tag{8}$$

We apply these cuts to our sample of candidate variables and select 846 sources. Their distribution in the $g - r$ vs. $u - g$ color-color diagram and rms scatter in the g band are shown in Figure 10 (top left panel). The distribution of sources in the RR Lyrae region is inhomogeneous. Sources with large rms scatter in the g band ($\gtrsim 0.2$ mag) are centered around $u - g \sim 1.15$, and are separated by $g - r \sim 0.12$ into two groups. A comparison with Figure 3 from Ivezić et al. (2005) suggests that these large rms scatter sources might be RR Lyrae type *ab* (RRab, $g - r > 0.12$) and type *c* stars (RRc, $g - r < 0.12$). Small rms scatter sources ($\lesssim 0.1$ mag) have a fairly uniform distribution, and are slightly bluer with $u - g \lesssim 1.1$.

The distribution of sources from the RR Lyrae region in the $\sigma(r)$ vs. $\sigma(g)$ diagram is presented in the top right panel of Figure 10. The majority of large rms scatter sources follow the $\sigma(g) = 1.4\sigma(r)$ relation, as expected for RR Lyrae stars. Since RR Lyrae stars are bluer when brighter, or equivalently, have greater rms scatter in the g band than in the r band, we require $1 < \sigma(g)/\sigma(r) \leqslant 2.5$ and select 683 candidate RR Lyrae stars.

A comparison of $u - g$ color distributions for candidate RR Lyrae stars and of sources with RR Lyrae colors, but not tagged as RR Lyrae stars, presented in the bottom left panel of Figure 10, demonstrates the robustness of the RR Lyrae selection. The two distributions are very different, with the candidate RR Lyrae distribution peaking at $u - g \sim 1.15$, as expected for RR Lyrae stars.

One property that distinguishes RRab from RRc stars is the shape (or skewness) of their lightcurves (in addition to lightcurve amplitude and period). RRab stars have asymmetric lightcurves, while RRc lightcurves are symmetric. In the top left panel of Figure 10, we

noted that $g - r \sim 0.12$ seemingly separates high rms scatter sources into two groups. If $g - r \sim 0.12$ is the boundary between the RRab and RRc stars, then the same boundary should show up in the distribution of lightcurve skewness as a function of the $g - r$ color. As shown in Figure 10 (bottom left panel), this is indeed the case. On average, sources with $g - r < 0.12$ have skewness $g \sim 0$ (symmetric lightcurves), as RRc stars, while $g - r > 0.12$ sources have skewness $g \sim -0.5$ (asymmetric lightcurves) typical of RRab stars.

We show in Section 4.2 that candidate RR Lyrae stars with skewness $g > 1$ are contaminated by eclipsing variables. Therefore, to reduce the contamination by eclipsing variables, we also require skewness $g \leqslant 1$, and select 634 sources as our final sample of candidate RR Lyrae stars.

4.2. Completeness and efficiency

The selection completeness, defined as the fraction of recovered RR Lyrae stars, will depend on the color cuts, $\sigma(g, r)$ cutoff, and the number of observations. The color cuts (Eqs. 2 to 8) applied in Section 4.1 were chosen to select an almost 100% complete sample of RR Lyrae stars (Ivezić et al. 2005). With the $\sigma(g, r)$ cutoff at 0.05 mag (small compared to the $\lesssim 1$ mag typical peak-to-peak amplitudes of RR Lyrae stars), and a fairly large number of observations per source (median of 10), we estimate the RR Lyrae selection completeness to be $\gtrsim 95\%$ (see Appendix in Ivezić et al. 2000)

To determine the selection efficiency, defined as the fraction of true RR Lyrae stars in the RR Lyrae candidate sample, we positionally match 683 candidate RR Lyrae stars selected by $1 < \sigma(g)/\sigma(r) \leqslant 2.5$, to a sample of sources selected from the Lightcurve-Motion Catalog (Bramich et al. 2007) as RR Lyrae stars and classified using lightcurves by De Lee et al. (2007), and find 613 matches. 70 candidate RR Lyrae stars from our sample, for some reason, do not have a match in the Lightcurve-Motion Catalog (De Lee, private communication). We find that 71% of sources in our candidate RR Lyrae sample are classified as RRab and RRc by De Lee et al., 28% are classified as variable non-RR Lyrae stars, and only 1% are spurious, non-variable sources. The most significant contamination comes from a population of variable sources bluer than $u - g \sim 1.1$ (dotted line, bottom left panel Figure 11), possibly Population II δ Scuti stars, also known as SX Phoenicis (Hoffmeister, Richter & Wenzel 1985). The top left and the bottom right panels in Figure 11, show that RRab and RRc-dominated regions are separated by $g - r \sim 0.12$, as already hinted in Figure 10. Also, variable non-RR Lyrae sources with skewness $g > 1$ are classified by De Lee et al. (2007) as eclipsing variables, justifying our skewness $g \leqslant 1$ cut.

To summarize, using color criteria and criteria based on $\sigma(g)$, $\sigma(r)$, and lightcurve skewness in the g band RR Lyrae stars are selected with $\gtrsim 95\%$ completeness and $\sim 70\%$ efficiency.

4.3. The Spatial Distribution of Candidate RR Lyrae Stars

Using the selection criteria from Section 4.1 we isolate 634 RR Lyrae candidates. The magnitude-position diagram for these candidates within 2.5 deg from the Celestial Equator is shown in Figure 12.

As discussed by Ivezić et al. (2005), an advantage of the data representation utilized in Figure 12 (magnitude–right ascension diagram) is its simplicity – only "raw" data are shown, without any post-processing. However, the magnitude scale is logarithmic and thus the spatial extent of structures is heavily distorted. In order to avoid these shortcomings, we have applied a Bayesian method for estimating continuous spatial density distribution developed by Ivezić et al. (2005), see their Appendix B. The resulting density map is shown in the right panel in Figure 13. The advantage of that representation is that it better conveys the significance of various local overdensities. For comparison, we also show a map of the northern part of the equatorial strip constructed using 2-epoch data discussed by Ivezić et al. (2000).

We detect several new halo substructures at $\gtrsim 3\sigma$ significance (compared to expected Poissonian fluctuations) and present their approximate locations and properties in Table 2. The most distant clump is at 100 kpc from the Galactic center. The strongest clump in the left wedge belongs to the Sgr dwarf tidal stream as well as the clump marked by C in the right wedge (Ivezić et al. 2003a). We note that the apparent "clumpiness" of the candidate RR Lyrae distribution increases with increasing radius, similar to CDM predictions by Bullock, Kravtsov & Weinberg (2001). A detailed comparison of their models with the data presented here will be discussed elsewhere (Sesar et al., in prep).

5. Are All Quasars Variable?

The optical continuum variability of quasars has been recognized since their first optical identification (Matthews & Sandage 1963), and it has been proposed and utilized as an efficient method for their discovery (van den Bergh, Herbst & Pritchet 1973; Hawkins 1983; Hawkins & Veron 1995). The observed characteristics of the variability of quasars are frequently used to constrain the origin of their emission (e.g. Kawaguchi et al. 1998

and references therein; Martini & Schneider 2003). Recently, significant progress in the description of quasar variability has been made by employing SDSS data (de Vries, Becker & White 2003; Vanden Berk et al. 2004; Ivezić et al. 2004b; de Vries et al. 2005; Sesar et al. 2006). Here we expand these studies by quantifying the efficiency of quasar discovery using variability.

A preliminary comparison of color and variability based methods for selecting quasars using SDSS data was presented by Ivezić et al. (2003b). They found that 47% of spectroscopically confirmed quasars with UV excess have the g band magnitude difference between two observations obtained two years apart larger than 0.15 mag. We can improve on their analysis because now there are significantly more observations obtained over a longer time period. Since quasars vary erratically and the rms scatter of their variability (the so-called structure function) increases with time (e.g. Vanden Berk et al. 2004 and references therein), the variability selection completeness is expected to be higher than $\sim 50\%$ obtained by Ivezić et al. (2003b).

First, although the adopted variability selection criteria discussed above are fairly conservative, we find that at least 63% of low-redshift quasars are variable at the $\geqslant 0.05$ mag level (simultaneously in the g and r bands over observer's time scales of several years) in the $g < 20.5$ flux-limited sample. Second, even this estimate is only a lower limit: given the spectroscopic confirmation for a large flux-limited sample of quasars, it is possible to relax the adopted variability selection cutoff without a prohibitive contamination by non-variable sources.

There are 2,492 quasars in the catalog of spectroscopically confirmed SDSS quasars (Schneider et al. 2005) from Stripe 82. The fraction of these objects that vary more than σ in the g and r bands, as a function of σ, is shown in Figure 14. We also show the analogous fraction for stars from the stellar locus. About 93% of quasars vary with $\sigma > 0.03$ mag. For a small fraction of these objects the measured rms scatter is due to photometric noise, and the stellar data limit this fraction to be at most 3%. Conservatively assuming that none of these 3% of stars is intrinsically variable, we estimate that *at least* 90% *of quasars are variable at the 0.03 mag level*.

6. Implications for Surveys such as LSST

The Large Synoptic Survey Telescope (LSST) is a proposed imaging survey that aims to obtain repeated multi-band imaging to faint limiting magnitudes over a large fraction of

the sky. The LSST Science Requirement Document[6] calls for ~ 1000 repeated observations of $\sim 20,000$ deg^2 large sky area, distributed over the six $ugrizY$ photometric bandpasses and over 10 years long survey. The results presented here can be extrapolated to estimate the lower limit on the number of variable sources that LSST would discover.

The single-epoch LSST images will have a 5σ detection limit[7] at $r \sim 24.7$. Hence, 2% accurate photometry, comparable to the subsample with $g < 20.5$ discussed here, will be available for stars with $r \lesssim 22$. The USNO-B catalog (Monet et al. 2003) shows that there are about 10^9 stars with $r < 21$ over the whole sky. About half of these stars are in the parts of the sky to be surveyed by LSST. The simulations based on contemporary Milky Way models, such as those developed by Robin et al. (2003) and Jurić et al. (2007), predict that there are about twice as many stars with $r < 22$ than with $r < 21$ across the whole sky. Hence, it is expected that LSST will detect about a billion stars with $r < 22$. These estimate is uncertain to within a factor of two or so due to unknown details in the spatial distribution of dust in the Galactic plane and towards the Galactic center.

We found that at least 0.5% of stars from the main stellar locus can be detected as variable with photometry accurate to $\sim 2\%$. This is only a lower limit because a much larger number of LSST observations obtained over a longer timespan than SDSS data discussed here would increase this fraction. Hence, our results imply that *LSST will discover at least 50 million variable stars*. Unlike the SDSS sample, where RR Lyrae stars account for $\sim 25\%$ of all variable *stars*, the number of RR Lyrae stars in LSST sample will be negligible compared to other types of variable star.

As estimated by Jurić et al. (2007) using deeper coadded SDSS photometry, there are about 100 deg^{-2} low-redshift quasars with $r < 22$. Therefore, with a sky coverage of $\sim 20,000$ deg^2, LSST will obtain well-sampled accurate multi-color lightcurves for ~ 2 million low-redshift quasars. Even at the redshift limit of ~ 2, this sample will be complete to $M_r \sim -24$, that is, almost to the formal quasars luminosity cutoff, and will represent an unprecedented sample for studying quasar physics.

7. Conclusions and Discussion

We have designed and tested algorithms for selecting candidate variable sources from a catalog based on multiple SDSS imaging observations. Using a sample of 13,051 selected

[6]Available at http://www.lsst.org/Science/lsst_baseline.shtml

[7]An LSST Exposure Time Calculator is available at www.lsst.org

candidate variable sources in the adopted $g = 20.5$ flux-limited sample, we find that at least 2% of unresolved optical sources appear variable at the $\geqslant 0.05$ mag level, simultaneously in the g and r bands. A similar fraction of variable sources ($\sim 1\%$) was also found by (Sesar et al. 2006) using recalibrated photometric POSS and SDSS measurements, and by Morales-Rueda et al. (2006) using the Faint Sky Variability Survey data ($\sim 1\%$).

Thanks to the multi-color nature of the SDSS photometry, and especially to the u band data, we can obtain robust classification of selected variable sources. The majority (2/3) of variable sources are low-redshift (< 2) quasars, although they represent only 2% of all sources in the adopted $g < 20.5$ flux-limited sample. We find that about 1/4 of variable stars are RR Lyrae stars, and that only 0.5% of stars from the main stellar locus are variable at the 0.05 mag level.

The distribution of lightcurve skewness in the g band for main stellar locus stars is bimodal, suggesting at least two, and possibly even four, different populations of variables. About a third of variable stars from the stellar locus show gray flux variations in the g and r bands ($\sigma(g)/\sigma(r) \sim 1$), and positive lightcurve skewness, suggesting variability caused by eclipsing systems. This population has an increased fraction of M type stars.

RR Lyrae stars show the largest rms scatter in the u and g bands, followed by low-redshift quasars. The ratio of rms scatter in the g and r bands for RR Lyrae is ~ 1.4, in agreement with Ivezić et al. (2000) results based on 2-epoch photometry. The mean lightcurve skewness for RR Lyrae stars is ~ -0.5, in agreement with Wils, Lloyd & Bernhard (2006). We selected a sample of 634 candidate RR Lyrae stars, with an estimated $\gtrsim 95\%$ completeness and $\sim 70\%$ efficiency. Using these stars, we detected rich halo substructure out to distances of 100 kpc. The apparent "clumpiness" of the candidate RR Lyrae distribution increases with increasing radius, similarly to CDM predictions by Bullock, Kravtsov & Weinberg (2001).

Low-redshift quasars show a dependence of $\sigma(g)/\sigma(r)$ on redshift, consistent with discussions in Richards et al. (2002) and Wilhite et al. (2006). The lightcurve skewness distribution for quasars is centered on zero in all photometric bands. We find that at least 90% of quasars are variable at the 0.03 mag level (rms) on time scales of order several years. This confirms that variability is as efficient a method for finding low-redshift quasars as is the UV excess color selection. The fraction of variable quasars at the $\geqslant 0.1$ mag level obtained here (30%, see Figure 14) is comparable to 36% found by Rengstorf, Brunner & Wilhite (2006).

The multiple photometric observations obtained by SDSS represent an excellent dataset for estimating the impact of surveys such as LSST on studies of the variable sky. Our results indicate that LSST will will obtain well-sampled 2% accurate multi-color lightcurves for ~ 2

million low-redshift quasars, and will discover at least 50 million variable stars. The number of variable stars discovered by LSST will be of the same order as the number of *all* stars detected by SDSS. With about 1000 data points in six photometric bands, it will be possible to recognize and classify variable objects using lightcurve moments of higher order than skewness discussed here, including lightcurve folding for periodic variables.

We acknowledge support by NSF grant AST-0551161 to LSST for design and development activity.

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.

A. Appendix

Skewness for small samples was calculated using equations from

http://www.xycoon.com/skewness_small_sample_test_1.htm.

$$skewness = \frac{n^2 M_3}{(n-1)(n-2)s^3} \tag{A1}$$

$$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \langle x \rangle)^3 \tag{A2}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \langle x \rangle)^2} \tag{A3}$$

where n is the number of observations, x_i is the magnitude, and $\langle x \rangle$ is the mean magnitude.

REFERENCES

Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S. et al., 2007, submitted

Akerlof, C., Amrose, S., Balsano, R. et al., 2000, AJ, 119, 1901

Becker, A. C., Wittman, D. M., Boeshaar, P. C. et al., 2004, ApJ, 611, 418

Bramich, D. et al., 2007, in prep

Bullock, J. S., Kravtsov, A. V., & Weinberg D. H., 2001, ApJ, 548, 33

De Lee, N. et al., 2007, in prep

de Vries, W. H., Becker, R. H., & White, R. L., 2003, AJ, 126, 1217

de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C., 2005, AJ, 129, 615

Fan, X. 1999, AJ, 117, 2528

Finlator, K., Ivezić, Ž., Fan, X. et al., 2000, AJ, 120, 2615

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

Groot, P. J., Vreeswijk, P. M., Huber, M. E. et al., 2003, MNRAS, 339, 427

Gunn, J. E., Carr, M., Rockosi, C. et al., 1998, AJ, 116, 3040

Gunn, J.E., Siegmund, W.A., Mannery, E.J., et al., 2006, AJ, 131, 2332

Hawkins, M. R. S., 1983, MNRAS, 202, 571

Hawkins, M. R. S., & Veron, P., 1995, MNRAS, 275, 1102

Hawley, S. L., Covey, K. R., Knapp, G. R. et al., 2002, AJ, 123, 3409

Helmi, A., Ivezić, Ž., Prada, F. et al., 2003, ApJ, 586, 195

Hoffmeister, C., Richter, G., & Wenzel, W., 1985, Variable Stars (New York:Springer)

Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J. & Gunn, J.E., 2002, AJ, 122, 2129

Huber, M. E., Everett, M. E., & Howell, S. B., 2006, AJ, 132, 633

Ivezić, Ž., Goldston, J., Finlator, K. et al., 2000, AJ, 120, 963

Ivezić, Ž., Lupton, R. H., Anderson, S. et al., 2003a, Proceedings of the Workshop *Variability with Wide Field Imagers*, Mem. Soc. Ast. It., 74, 978 (also astro-ph/0301400)

Ivezić, Ž., Lupton, R. H., Johnston, D. E. et al., 2003b, Proceedings of "AGN Physics with the SDSS", Richards, G.T. & Hall, P.B., eds., in press (also astro-ph/0310566)

Ivezić, Ž., Lupton, R. H., Schlegel, D. et al., 2004a, Astronomische Nachrichten, 325, No. 6-8, 583-589 (also astro-ph/0410195)

Ivezić, Ž., Lupton, R. H., Jurić, M. et al., 2004b, Proceedings of "The Interplay among Black Holes, Stars, and ISM in Galactic Nuclei", IAU Symposium No. 222, Bergmann, Th. S., Ho, L. C., & Schmitt, H. R., eds., p. 525 (also astro-ph/0404487)

Ivezić, Ž., Vivas, A. K., Lupton, R. H. & Zinn, R., 2005, AJ, 129, 1096

Ivezić, Ž., Smith, J. A., Miknaitis, G. et al., 2007, in prep

Jurić, M., Ivezić, Ž., Brooks, A. et al., 2007, submitted to AJ (also astro-ph/0510520)

Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L., 1998, ApJ, 504, 671

Lenz, D. D., Newberg, J., Rosner, R., Richards, G. T., & Stoughton, C., 1998, ApJS, 119, 121

Lupton, R. H., Ivezić, Ž., Gunn, J. E., Knapp, G. R., Strauss, M. A. & Yasuda, N., 2002, in "Survey and Other Telescope Technologies and Discoveries", Tyson, J. A. & Wolff, S., eds. Proceedings of the SPIE, 4836, 350

Martini, P., & Schneider, D. P., 2003, ApJ, 597, L109

Matthews, T. A., & Sandage, A. R., 1963, ApJ, 138, 30

Monet, D. G., Levine, S. E., Canzian, B. et al., 2003, AJ, 125, 984

Morales-Rueda, L., Groot, P. J., Augusteijn, T. et al., 2006, MNRAS, 371, 1681

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

Richards, G. T., Fan, X., Newberg, H. J. et al., 2002, AJ, 123, 2945

Rengstorf, A. W., Mufson, S. L., Andrews, P. et al., 2004, AJ, 617, 184

Rengstorf, A. W., Brunner, R. J., & Wilhite, B. C., 2006, AJ, 131, 1923

Robin, A. C., Reylé, C., Derriére, S., & Picaud, S., 2003, A&A, 409, 523

Schlegel, D., Finkbeiner, D. P., & Davis, M., 1998, ApJ500, 525

Schneider, D. P., Hall, P. B., Richards, G. T. et al, 2005, AJ, 367

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

Sesar, B., Svilković, D., Ivezić, Ž. et al., 2006, AJ, 131, 2801

Sirko, E., Goodman, J., Knapp, G. et al., 2004, AJ, 127, 899S

Smith, J. A., Tucker, D. L., Kent, S. M. et al., 2002, AJ, 123, 2121

Smolčić, V., Ivezić, Ž., Knapp, G. R. et al., 2004, ApJ, 615L, 141

Tonry, J. L., Howell, S. B., Everett, M. E. et al., 2005, PASP, 117, 281

Trevese, D., Kron, R. G., & Bunone, A., 2001, ApJ, 551, 103

Tucker, D., Kent, S., Richmond, M.W., et al., 2006, AN, 327, 821

Tyson, J.A., 2002, in "Survey and Other Telescope Technologies and Discoveries", Tyson, J. A. & Wolff, S., eds. Proceedings of the SPIE, 4836, 10

Udalski, A., Zebrun, K., Szymanski, M. et al., 2002, Acta Astronomica, 52, 115

van den Bergh, S., Herbst, E., & Pritchet, C. 1973, AJ, 78, 375

Vanden Berk, D. E., Wilhite, B. C., Kron, R. G. et al., 2004, ApJ, 601, 692

Vivas, A. K., Zinn, R., Andrews, P. et al., 2001, ApJ, 554, L33

Walker, A.R. 2003, Proceedings of the Workshop *Variability with Wide Field Imagers*, Mem. Soc. Ast. It., 74, 999 (also astro-ph/0303012)

Wilhite, B. C., Vanden Berk, D. E., Kron, R. G. et al., 2006, ApJ, 633, 638

Wils, P., Lloyd, C., & Bernhard, K., 2006, MNRAS, 368, 1757

Woźniak, P. R., Udalski, A., Szymanski, M. et al., 2002, Acta Astronomica, 52, 129

Woźniak, P. R. et al., 2004, AJ127,2436

Yanny, B., Newberg, H. J., Kent, S. et al., 2000, ApJ, 540, 825

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

Table 1. The distribution of candidate variable sources in the $g-r$ vs $u-g$ diagram

		$g < 19$				$g < 20.5$				$g < 22$			
Region[a]	Name[b]	% all[c]	% var[d]	var/all[e]	N_{var}/N^f_{all}	% all[c]	% var[d]	var/all[e]	N_{var}/N^f_{all}	% all[c]	% var[d]	var/all[e]	N_{var}/N^f_{all}
I	white dwarfs	0.14	0.59	4.25	3.50	0.24	0.40	1.69	3.34	0.28	0.45	1.64	4.51
II	low-redshift QSOs	0.45	30.88	68.83	56.58	1.90	62.90	33.03	65.10	4.07	70.01	17.22	47.30
III	dM/WD pairs	0.08	0.53	6.54	5.37	0.83	2.08	2.50	4.92	1.21	3.79	3.13	8.61
IV	RR Lyrae stars	1.28	16.81	13.11	10.78	1.33	7.95	5.99	11.81	1.48	6.41	4.33	11.90
V	stellar locus stars	96.27	48.77	0.51	0.42	94.49	25.15	0.27	0.52	91.89	18.33	0.20	0.55
VI	high-redshift QSOs	1.78	2.42	1.36	1.12	1.21	1.52	1.26	2.48	1.07	1.01	0.95	2.60
	total count	411,667	3,384			662,195	13,051			748,067	20,553		

[a]These regions are defined in the $g-r$ vs. $u-g$ color-color diagram, with their boundaries shown in Fig. 4

[b]An approximate description of the dominant source type found in the region

[c]The fraction of all sources in a magnitude-limited sample found in this color region, with the magnitude limits listed on top.

[d]The number of candidate variables from the region, expressed as a fraction of all variable sources

[e]The ratio of values listed in columns d) and c)

[f]The number of candidate variables from the region, expressed as a fraction of all sources in that region

Table 2. Approximate locations and properties of detected overdensities

Label[a]	N[b]	$\langle R.A.\rangle^c$	$\langle d\rangle^d$	$\langle r\rangle^e$	$\langle u-g\rangle^f$	$\langle g-r^g\rangle$	$\langle skew\ g\rangle^h$	N_b/N_r^i
A	84	330.95	21	17.02	1.14	0.18	-0.50	0.62
B	144	309.47	22	16.76	1.12	0.16	-0.57	0.64
C	54	33.69	25	17.61	1.13	0.20	-0.68	0.29
D	8	347.91	29	18.02	1.14	0.23	-0.50	0.38
E	11	314.06	43	18.84	1.09	0.20	-0.41	0.75
F	11	330.26	48	19.16	1.07	0.20	-0.46	0.38
G	10	354.81	55	19.46	1.10	0.22	-0.69	0.38
H	7	43.57	57	19.32	1.05	0.04	0.06	1.34
I	4	311.34	72	19.98	1.08	0.11	-0.10	2.0
J	26	353.58	81	20.21	1.11	0.20	-0.27	0.58
K	8	28.39	84	20.35	1.10	0.20	0.14	0.44
L	3	339.01	92	20.45	1.06	0.16	0.08	0.67
M	5	39.45	102	20.73	1.07	0.11	0.36	1.67

[a]Overdensity's label from Fig. 13

[b]Number of candidate RR Lyrae in the overdensity

[c]Median Right Ascension

[d]Median distance (in kpc)

[e]Median r band magnitude

[f]Median $u-g$ color

[g]Median $g-r$ color

[h]Median lightcurve skewness in the g band

[i]The number ratio of candidate RR Lyrae with $g-r<0.12$ and $g-r>0.12$

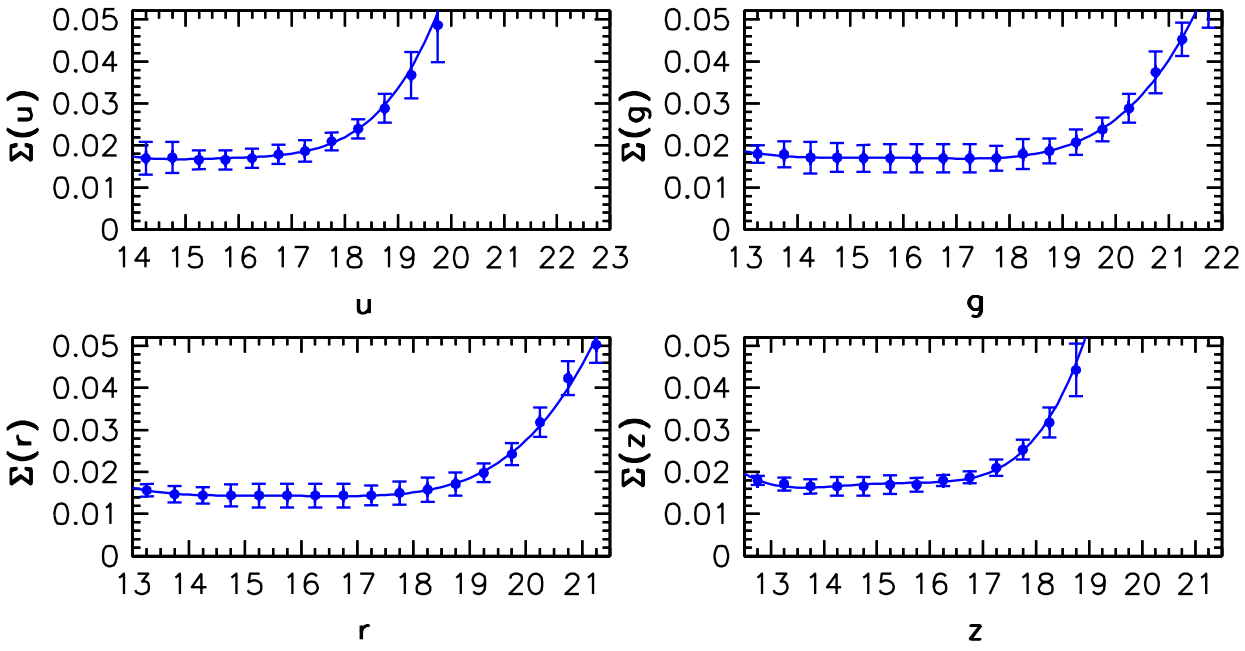

Fig. 1.— The dependence of the median root-mean-square (rms) scatter Σ in SDSS $ugrz$ bands on magnitude (symbols). The vertical bars show the rms scatter of Σ in each bin (not the error of the median). The dependence of Σ in the i band is similar to the r band dependence. In each band, a fourth-degree polynomial is fitted through medians and shown by the solid line.

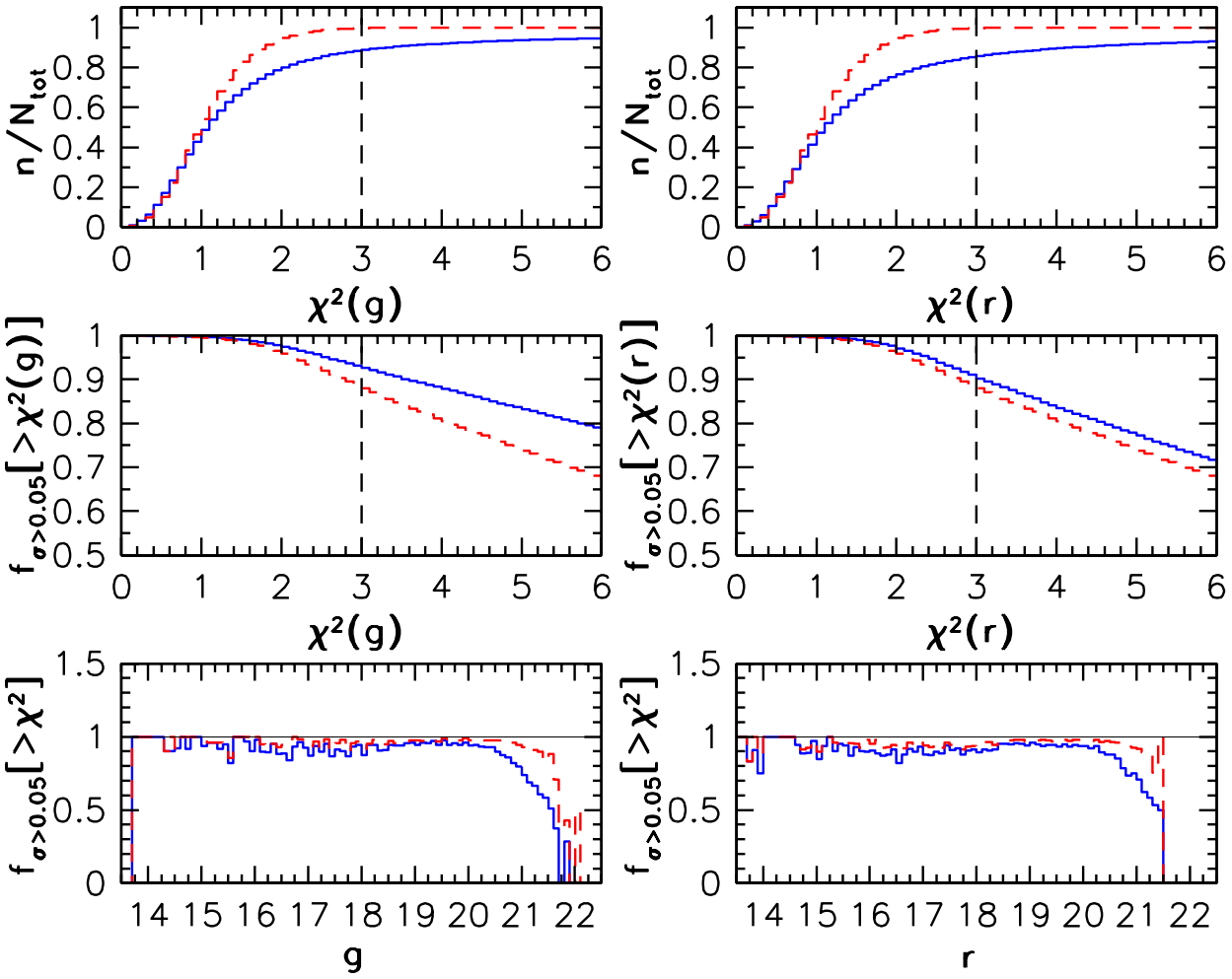

Fig. 2.— *Top*: The cumulative distribution of χ^2 g and r values for all sources (solid line) and a reference Gaussian χ^2 distribution with 9 degrees of freedom (dashed line). Vertical dashed lines show adopted selection cuts on $\chi^2(g)$ and $\chi^2(r)$ values. *Middle*: The fraction of $\sigma(g,r) \geqslant 0.05$ mag sources with χ^2 per degree of freedom greater than χ^2 (only in the g or r band: solid line, both in the g and r bands: dashed line). *Bottom*: The fraction of $\sigma(g,r) \geqslant 0.05$ mag sources with $\chi^2(m) \geqslant 2$ (dashed line) or $\chi^2(m) \geqslant 3$ (solid line) as a function of magnitude for $m = g, r$ bands, respectively.

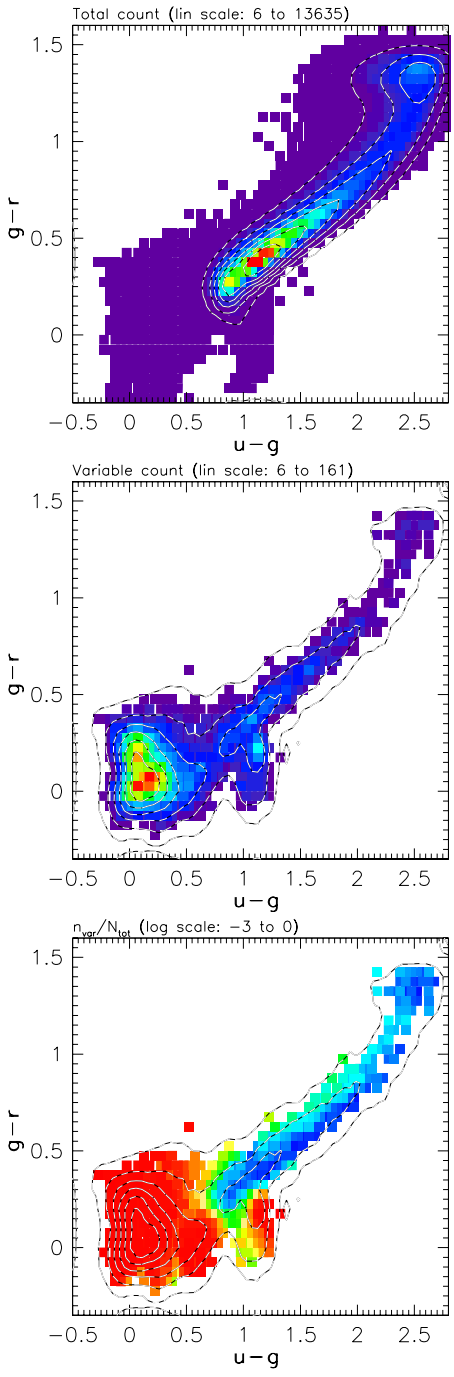

Fig. 3.— The distribution of counts for the full sample (top), candidate variable sample (middle), and the ratio of two counts (bottom) in the $g-r$ vs. $u-g$ color-color diagram for sources brighter than $g = 20.5$, binned in 0.05 mag bins. Contours outline distributions of unbinned counts. Note the remarkable difference between the distribution of all sources and that of the variable sample, which demonstrates that the latter are robustly selected.

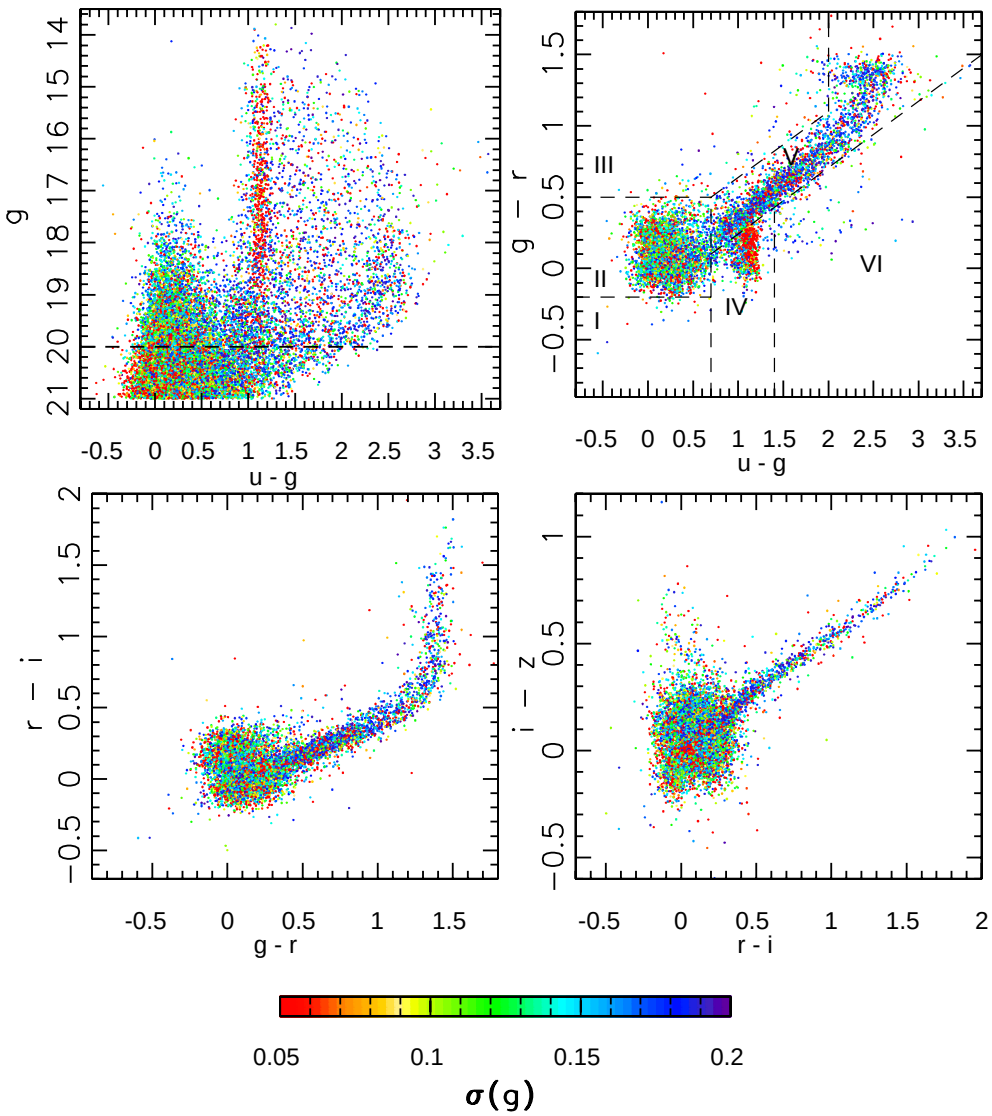

Fig. 4.— The distribution of 18,329 candidate variable sources brighter than $g = 21$, in representative SDSS color-magnitude and color-color diagrams. Candidate variables are color-coded by their rms scatter in the g band (0.05-0.2, see the legend, red if larger than 0.2). Only sources brighter than $g = 20$ are plotted in color-color diagrams. Note how RR Lyrae stars ($u - g \sim 1.15$, red dots, $\sigma(g) \gtrsim 0.2$ mag) and low-redshift quasars ($u - g \leqslant 0.7$, green dots, $\sigma(g) \gtrsim 0.1$ mag) stand out as highly variable sources. The regions marked in the top right panel are used for quantitative comparison of the overall and variable source distributions (see Table 1).

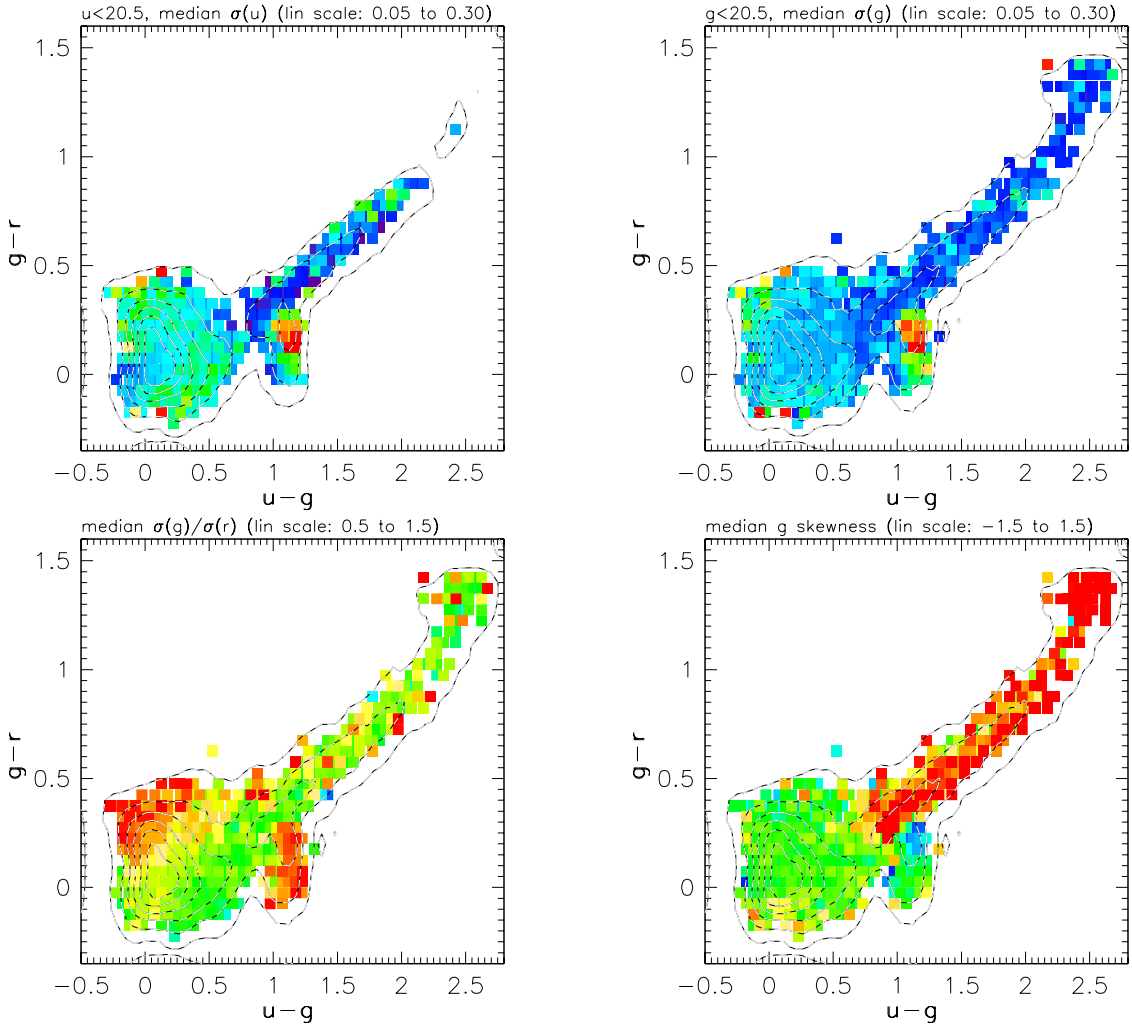

Fig. 5.— The distribution of the rms scatter $\sigma(u)$ (top left), rms scatter $\sigma(g)$ (top right), $\sigma(g)/\sigma(r)$ ratio (bottom left), and lightcurve skewness in the g band (bottom right) for the variable sample in the $g-r$ vs. $u-g$ color-color diagram. Sources are binned in 0.05 mag wide bins and the median values are color-coded. Color ranges are given at the top of each panel, going from blue to red, where the green color is in the mid-range. Values outside the range saturate in blue or red. Contours outline the count distributions on linear scale in steps of 15%. Magnitude limit is $g < 20.5$, with an additional $u < 20.5$ limit in the top left panel. *Bottom left:* On average, RR Lyrae stars have $\sigma(g)/\sigma(r) \sim 1.4$, main stellar locus stars have $\sigma(g)/\sigma(r) \sim 1$, while low-redshift quasars show a gradient of $\sigma(g)/\sigma(r)$ values. *Bottom right:* On average, quasars and c type RR Lyrae stars ($u-g \sim 1.15$, $g-r < 0.15$) have skewness ~ 0, ab type RR Lyrae ($u-g \sim 1.15$, $g-r > 0.15$) have negative skewness, and stars in the main stellar locus have positive skewness.

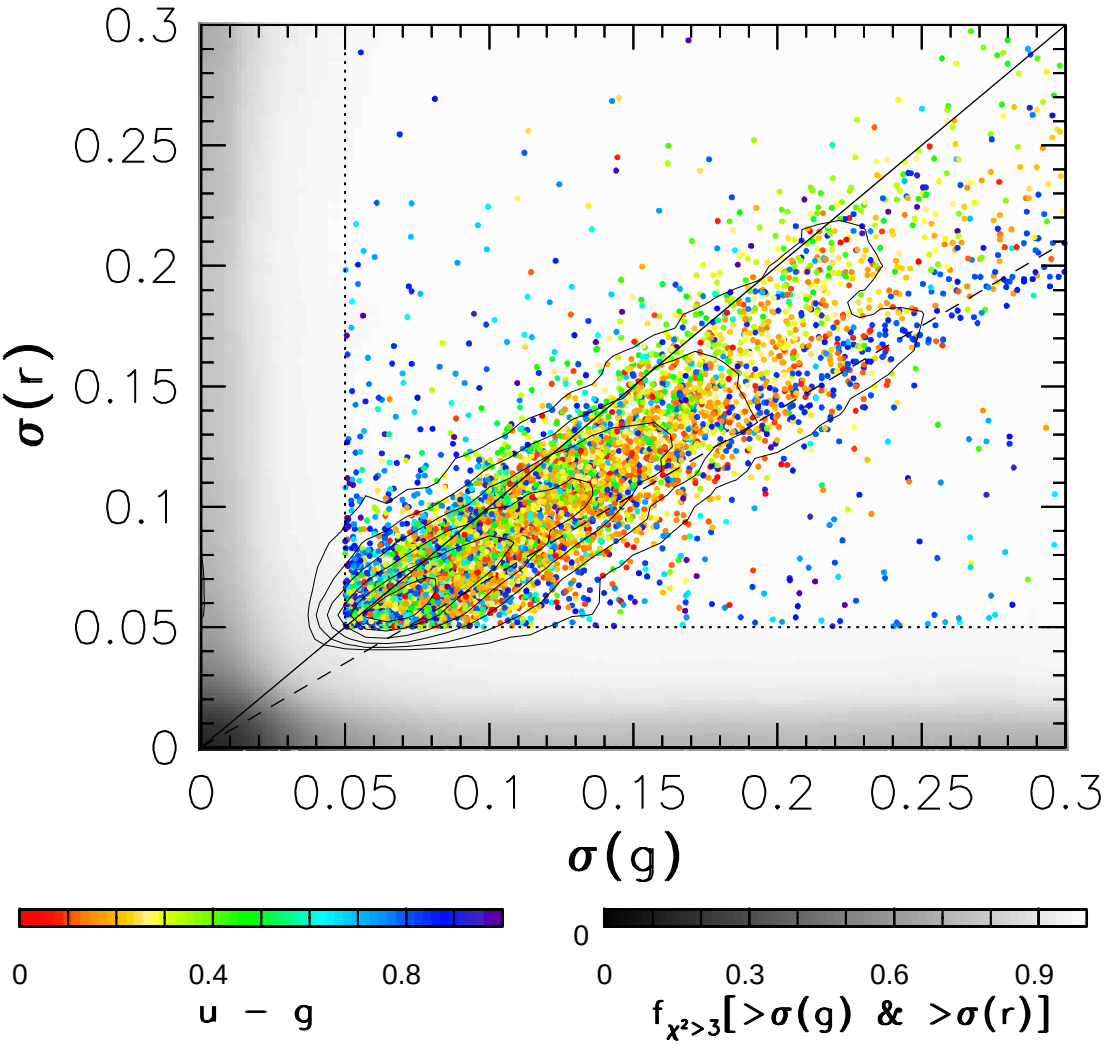

Fig. 6.— The distribution of candidate variable sources in the $g < 20.5$ flux-limited sample is shown by linearly-spaced contours, and by symbols color-coded by the $u - g$ color for sources with $\sigma(g) \geqslant 0.05$ mag and $\sigma(r) \geqslant 0.05$ mag. The thick solid line shows $\sigma(g) = \sigma(r)$, while the dashed line shows $\sigma(g) = 1.4\sigma(r)$ relation representative of RR Lyrae stars. Note that sources following the $\sigma(g) = 1.4\sigma(r)$ relation tend to have $u - g \sim 1$, as expected for RR Lyrae stars. The greyscale background shows the fraction of $\chi^2(g, r) \geqslant 3$ sources which also have $\sigma(g) \geqslant x$ and $\sigma(r) \geqslant y$ (large χ^2 sources also have large σ). The dotted lines show the adopted $\sigma(g, r)$ selection cut.

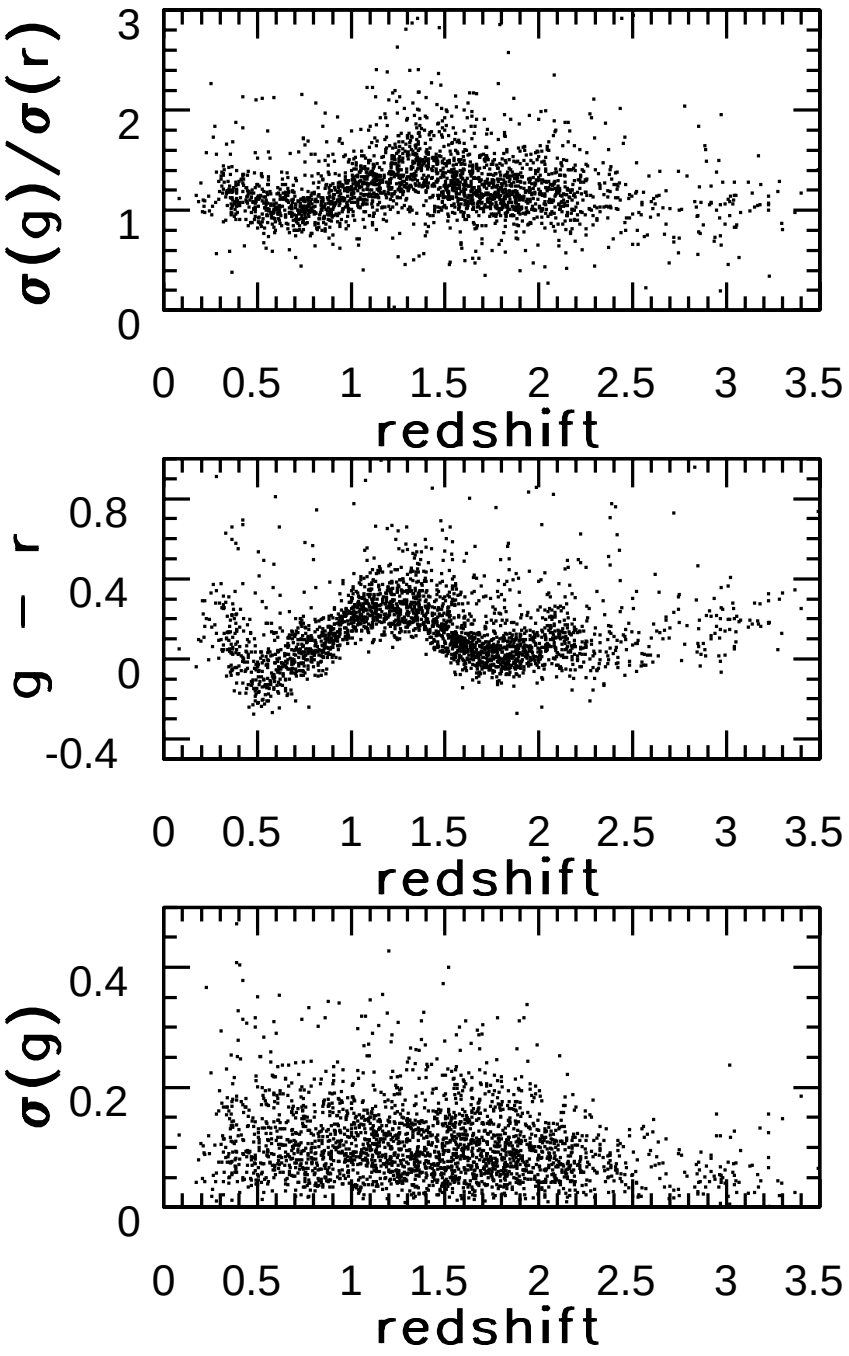

Fig. 7.— The dependence of $\sigma(g)/\sigma(r)$ (top), $g-r$ (middle), and $\sigma(g)$ on redshift for a sample of spectroscopically confirmed quasars from Schneider et al. (2005). The $\sigma(g)/\sigma(r)$ gradient shown in Fig. 5 (bottom left panel) can be explained by the local maximum of $\sigma(g)/\sigma(r)$ in the 1.0 to 1.6 redshift range.

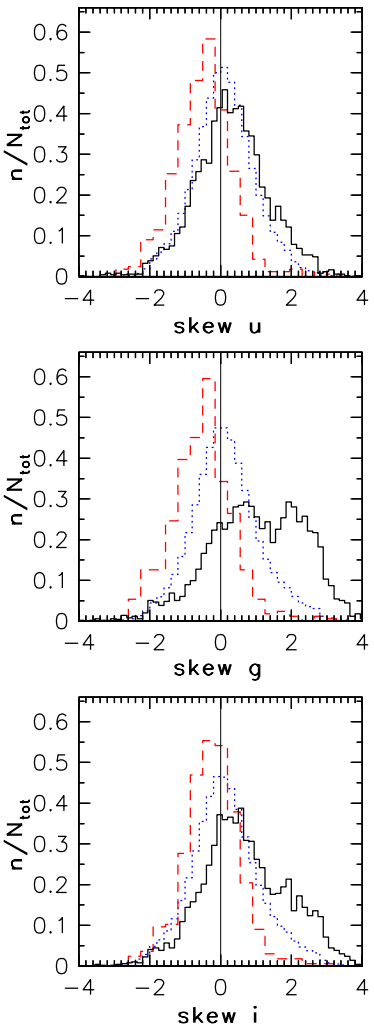

Fig. 8.— The lightcurve skewness distribution in the *ugi* bands for spectroscopically confirmed quasars (dotted line), candidate RR Lyrae stars (dashed line), and variable main stellar locus stars (solid line, Region V, see Fig. 4 for the definition). The distribution of the skewness in the r band is similar to the g band distribution, and the distribution of skewness in the z band is similar to the u band distribution. Stars in the main stellar locus, show bimodality in skewness distribution in the g band, suggesting at least two, and perhaps as many as four, different populations of variables. Similar bimodality is also discernible in the r band, while it is less pronounced in the i band and not detected in the u and z bands. Quasars have symmetric lightcurves (skewness ~ 0) and their distribution of skewness does not change between bands.

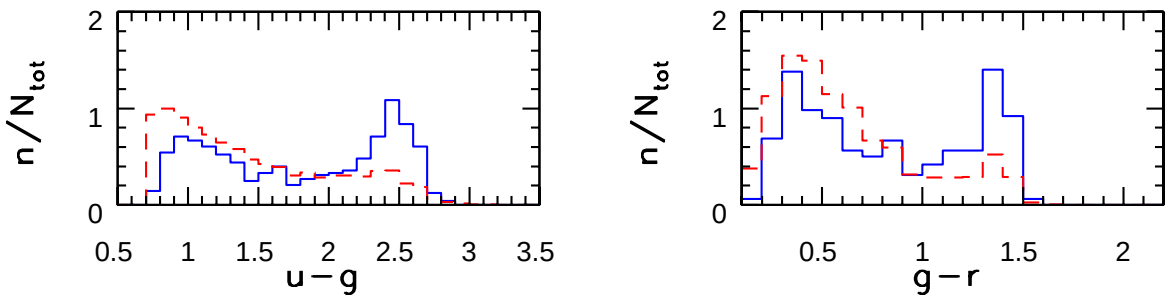

Fig. 9.— A comparison of the $u-g$ (left) and $g-r$ (right) color distributions for variable main stellar locus stars brighter than $g = 19$ (dashed lines), and a subset with highly asymmetric lightcurves (skew $g \sim 2.5$, solid lines). The subset with highly asymmetric lightcurves has an increased fraction of stars with colors $u - g \sim 2.5$ and $g - r \sim 1.5$, characteristic of M stars.

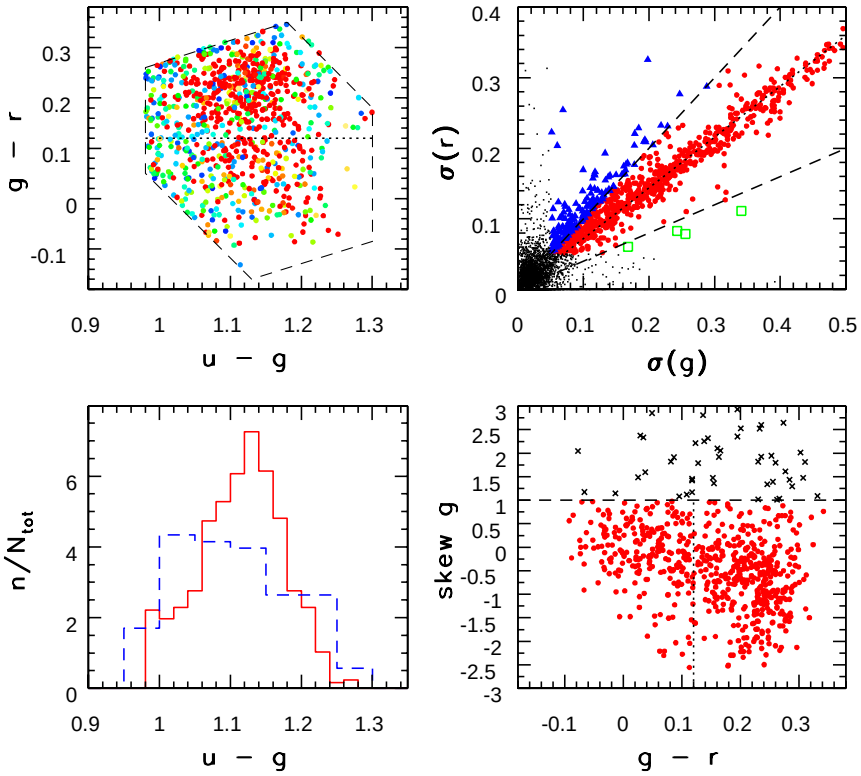

Fig. 10.— *Top left:* The distribution of 846 candidate variable sources from the RR Lyrae region (dashed lines, see Fig. 3 in Ivezić et al. 2005) in the $g-r$ vs. $u-g$ color-color diagram. The symbols mark the time-averaged values and are color-coded by $\sigma(g)$ (0.05 to 0.2, blue to red). The dotted horizontal line shows the boundary between the RRab and RRc-dominated regions. *Top right:* Sources from the top left panel divided into 3 groups according to their $\sigma(g)/\sigma(r)$ values: candidate RR Lyrae stars with $1 < \sigma(g)/\sigma(r) \leqslant 2.5$ (large dots), sources with $\sigma(g)/\sigma(r) \leqslant 1$ (triangles), and sources with $\sigma(g)/\sigma(r) > 2.5$ (squares). Small dots show sources with RR Lyrae colors that fail the variability criteria. The dashed lines show the $\sigma(g) = \sigma(r)$ and $\sigma(g) = 2.5\sigma(r)$ relations, while the dotted line shows the $\sigma(g) = 1.4\sigma(r)$ relation. *Bottom left:* A comparison of the $u-g$ color distributions for candidate RR Lyrae stars (solid line) and sources with RR Lyrae colors but not tagged as RR Lyrae stars (dashed line). *Bottom right:* The dependence of the lightcurve skewness in the g band on the $g-r$ color for candidate RR Lyrae stars. The boundary $g-r = 0.12$ (vertical dotted line) separates candidate RR Lyrae stars into those with asymmetric (skewness $g \sim -0.5$) and symmetric (skewness $g \sim 0$) lightcurves, corresponding to RRab and RRc stars, respectively. The condition skewness $g \leqslant 1$ (horizontal dashed line) is used to reduce the contamination of the RR Lyrae sample by eclipsing variables.

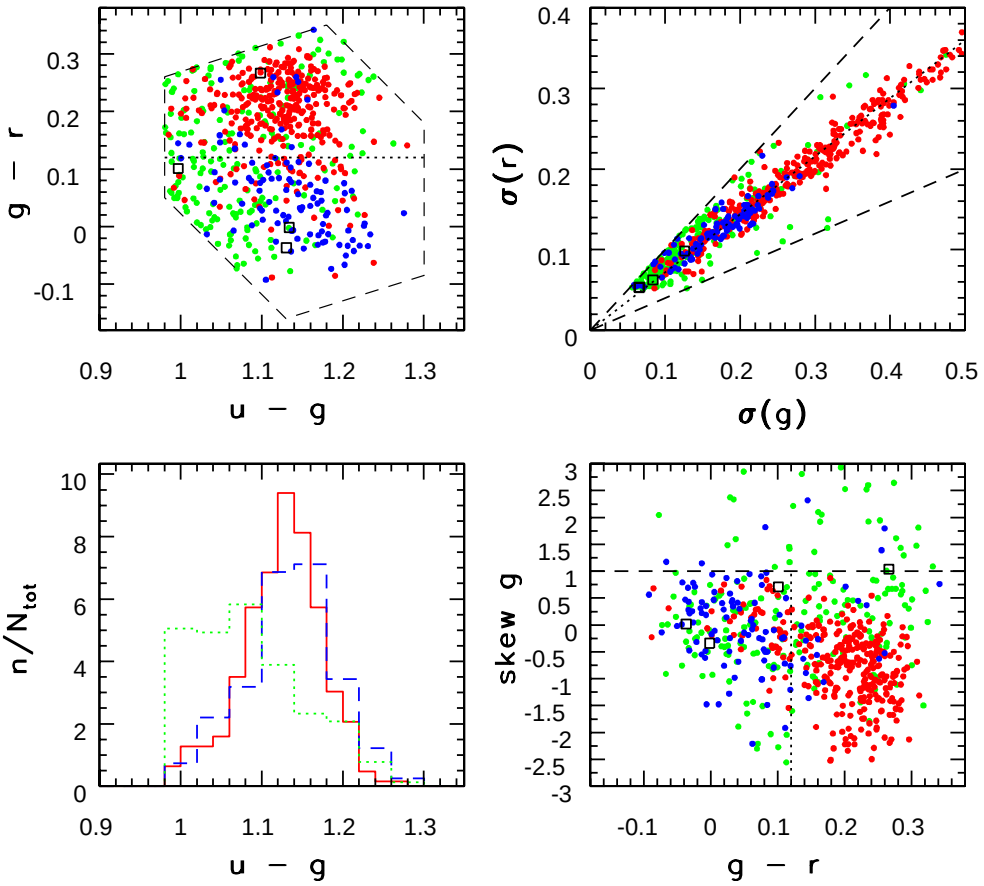

Fig. 11.— The distribution of candidate RR Lyrae stars selected with $1 < \sigma(g)/\sigma(r) \leqslant 2.5$ and classified by De Lee et al. (2007), shown in diagrams similar to Fig. 10. Symbols show RRab stars (red dots), RRc stars (blue dots), variable non-RR Lyrae stars (green dots), and non-variable sources (open boxes, only four sources). A comparison of the $u - g$ color distribution for RRab (solid line), RRc (dashed line), and variable non-RR Lyrae stars (doted line) is shown in the bottom left panel.

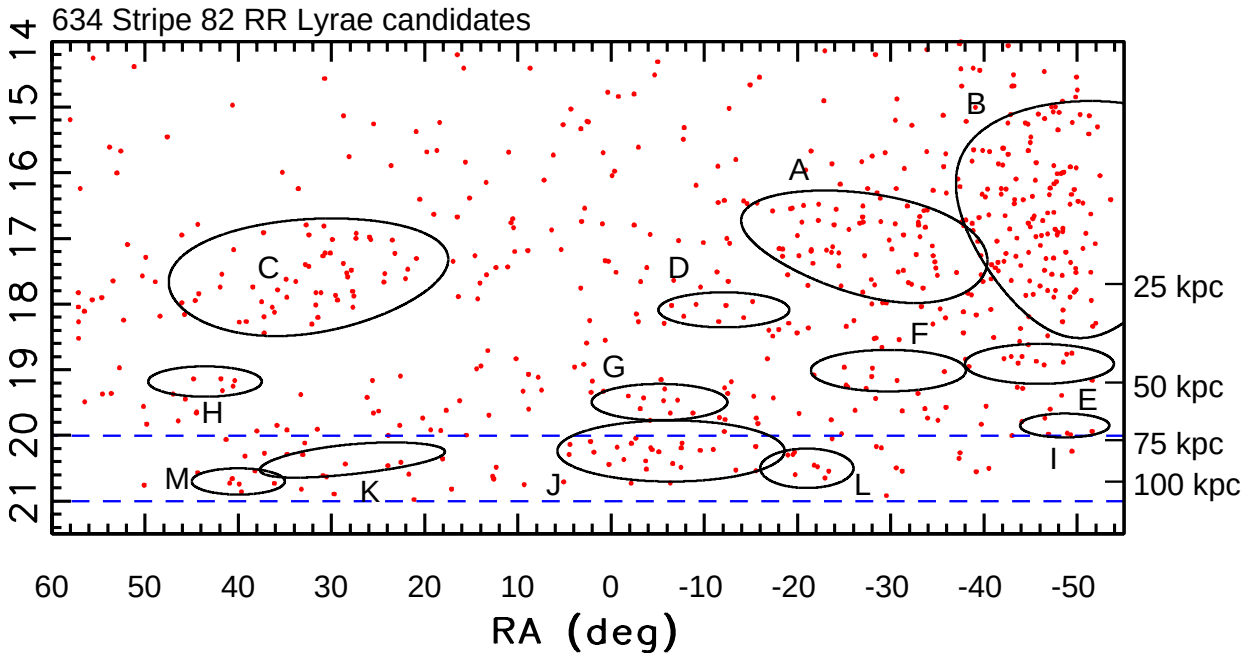

Fig. 12.— The magnitude-position distribution of 634 Stripe 82 RR Lyrae candidates within $-55° < R.A. < 60°$ and $|Dec| \leqslant 1.27°$. Approximate distance (shown on the right y-axis) is calculated assuming $M_r = 0.7$ mag for RR Lyrae. Dashed lines show where sample completeness decreases from approximately 99% to 60% due to the χ^2 cut (see the bottom right panel in Fig. 2). Closed curves are remapped ellipses and circles from Fig. 13 that mark the halo substructure.

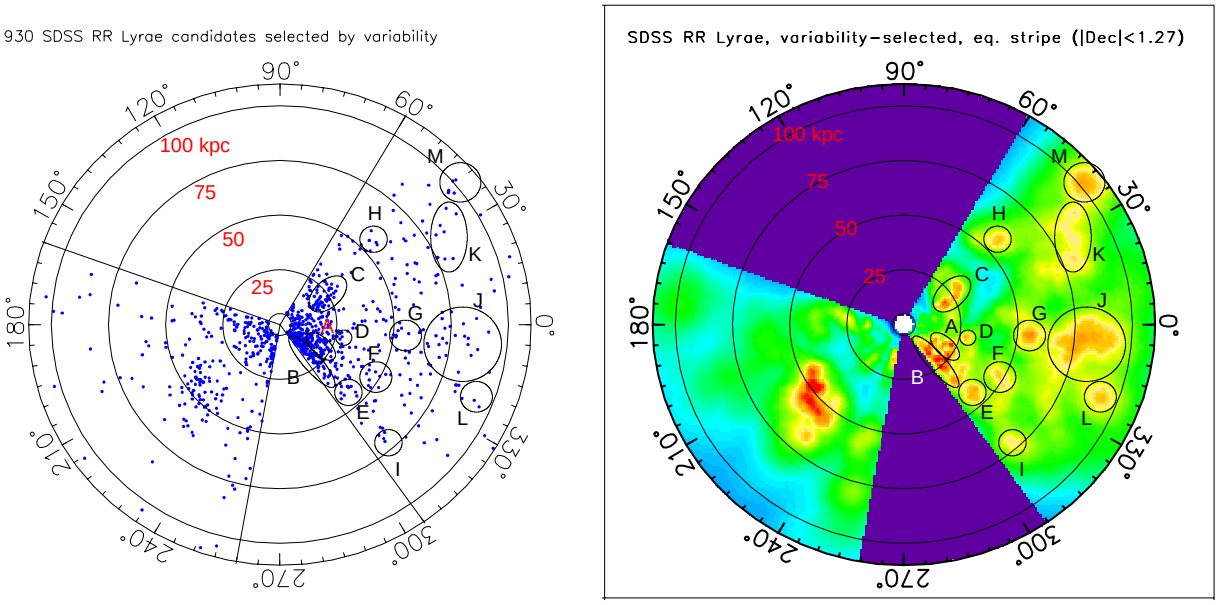

Fig. 13.— *Left:* The spatial distribution of candidate RR Lyrae stars discovered by SDSS along the Celestial Equator. Distance is calculated assuming Eq. 3 from Ivezić et al. (2005) and $M_V = 0.7$ mag as the absolute magnitude of RR Lyrae in the V band. The right wedge corresponds to candidate RR Lyrae selected in this work (634 candidates, shown in Fig. 12) and the left wedge is based on the sample from Ivezić et al. (2000) (296 candidates). *Right:* The number density distribution of candidate RR Lyrae stars shown in the top panel, computed using an adaptive Bayesian density estimator developed by Ivezić et al. (2005). The color scheme represents the number density multiplied by the cube of the galactocentric radius, and displayed on a logarithmic scale with a dynamic range of 300 (from light blue to red). The green color corresponds to the mean density – both wedges with the data would have this color if the halo number density distribution followed a perfectly smooth r^{-3} power-law. The purple color marks the regions without the data. The yellow regions are formally $\sim 3\sigma$ significant overdensities, and orange/red regions have an even higher significance (using only the counts variance). The strongest clump in the left wedge belongs to the Sgr dwarf tidal stream as well as the clump marked by C in the right wedge (Ivezić et al. 2003a). An approximate location and properties of labeled overdensities are listed in Table 2. The Ivezić et al. (2000) sample is based on only 2 epochs and thus has a much lower completeness ($\sim 56\%$) resulting in a lower density contrast.

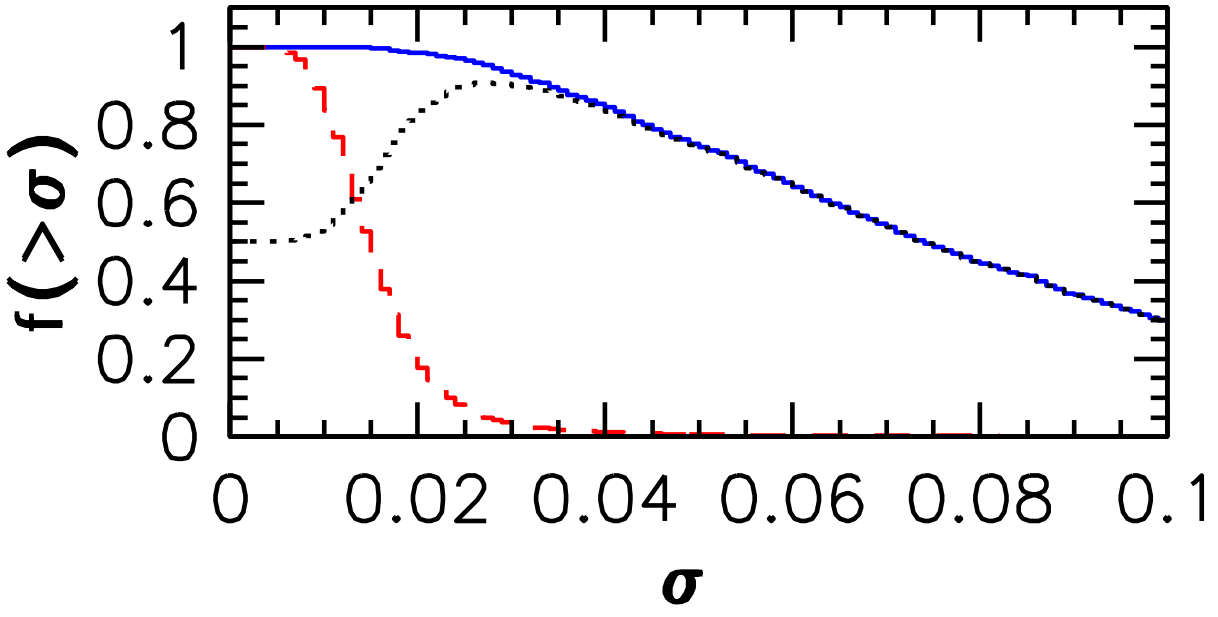

Fig. 14.— The fraction of spectroscopically confirmed QSOs (solid line) and the fraction of sources from the stellar locus (dashed line) brighter than $g = 19.5$ and $r = 19.5$ that have rms scatter larger than σ in the g and r bands. The ratio $f_{QSO}/(1+f_{loc})$ (dotted line), which corresponds to the implied fraction of variable QSOs, peaks at a level of 90% for $\sigma = 0.03$ mag.