

## THE SELECTION OF RR LYRAE STARS USING SINGLE-EPOCH DATA

ŽELJKO IVEZIĆ,<sup>1,2</sup> A. KATHERINA VIVAS,<sup>3</sup> ROBERT H. LUPTON,<sup>2</sup> AND ROBERT ZINN<sup>4</sup>

*Received 2003 October 20; accepted 2004 November 12*

### ABSTRACT

We use a complete sample of RR Lyrae stars discovered by the Quasar Equatorial Survey Team survey using light curves to design selection criteria based on Sloan Digital Sky Survey (SDSS) colors. Thanks to the sensitivity of the  $u - g$  color to surface gravity and of the  $g - r$  color to effective temperature and to the small photometric errors ( $\sim 0.02$  mag) delivered by SDSS, RR Lyrae stars can be efficiently and robustly recognized even with single-epoch data. In a 100% complete color-selected sample, the selection efficiency (the fraction of RR Lyrae stars in the candidate sample) is 6%, and, by adjusting color cuts, it can be increased to 10% with a completeness (the fraction of selected RR Lyrae stars) of 80% and to 60% with 28% completeness. Such color selection produces samples that are sufficiently clean for statistical studies of the Milky Way's halo substructure, and we use it to select 3,643 candidate RR Lyrae stars from SDSS Data Release 1. We demonstrate that this sample recovers known clumps of RR Lyrae stars associated with the Sgr dwarf tidal tail and the Palomar 5 globular cluster and use it to constrain the halo substructure away from the Sgr dwarf tidal tail. These results suggest that it will be possible to study the halo substructure out to  $\sim 70$  kpc from the Galactic center in the entire area imaged by the SDSS, and not only in the multiply observed regions.

*Key words:* Galaxy: halo — Galaxy: stellar content — Galaxy: structure — stars: variables: other

*Online material:* color figures

### 1. INTRODUCTION

Studies of substructures, such as clumps and streams, in the Galactic halo can help constrain the formation history of the Milky Way. Hierarchical models of galaxy formation predict that these substructures should be ubiquitous in the outer halo, where the dynamical timescales are sufficiently long for them to remain spatially coherent (Johnston et al. 1996; Mayer et al. 2002; Helmi 2002). One of the best tracers to study the outer halo are RR Lyrae stars because

They are nearly standard candles (dispersion of  $\sim 0.13$  mag; Vivas et al. 2001), and thus it is straightforward to determine their distance, and

They are sufficiently bright ( $\langle M_V \rangle = 0.7 - 0.8$ ; Layden et al. 1996; Gould & Popowski 1998) to be detected at large distances (5–100 kpc for  $14 < r < 20.7$ ).

One of the disadvantages of using RR Lyrae stars as tracers of the halo is that they appear only in very old populations, and their number (if any) depends on the morphology of the horizontal branch of the stellar population, which is probably a function of metallicity and age. However, it seems reasonable to assume that the progenitors of any stream in the halo resemble the present-day satellite galaxies of the Milky Way. If so, we expect RR Lyrae stars to be indeed good tracers of the substructures in the halo, since all of the satellite galaxies contain large numbers of these variable stars. Recent surveys for RR Lyrae stars (Vivas et al. 2001, 2004; Ivezić et al. 2000, hereafter I00, 2004b, 2005) have already detected several substructures in the halo.

RR Lyrae stars are typically found by obtaining well-sampled light curves. The Quasar Equatorial Survey Team (QUEST) survey is the largest such survey that is capable of discovering RR Lyrae stars in the outer halo. Using a 1 m Schmidt telescope, the QUEST survey has so far discovered about 500 RR Lyrae stars in  $400 \text{ deg}^2$  of sky (Vivas et al. 2004). Nevertheless, I00 demonstrated that RR Lyrae stars can be efficiently and robustly found even with two-epoch data, using accurate multiband photometry obtained by the Sloan Digital Sky Survey (SDSS). The QUEST survey later demonstrated (Vivas et al. 2001) that most ( $> 90\%$ ) of the SDSS candidates are indeed RR Lyrae stars and also confirmed the estimate of the sample completeness ( $\sim 35\% \pm 5\%$ ).

To extend the above surveys for RR Lyrae stars to a significant fraction of the sky (say, one quarter) is difficult. The QUEST survey will cover up to  $700 \text{ deg}^2$ , whereas the SDSS survey, which should observe close to one quarter of the sky, will obtain only single-epoch data for most of the scanned area. However, here we demonstrate that the distinctive SDSS colors of RR Lyrae stars allow their selection using only a single epoch of data. We use a complete sample of RR Lyrae stars discovered by the QUEST survey and design optimal selection criteria based on SDSS colors. The data and the selection method are described in § 2, we select and analyze candidate RR Lyrae stars from SDSS Data Release 1 in § 3, and we summarize and discuss the results in § 4.

### 2. THE SELECTION OF RR LYRAE STARS USING SDSS COLORS

#### 2.1. The SDSS and QUEST Data

The SDSS (York et al. 2000) is revolutionizing studies of the Galactic halo because it is providing homogeneous and deep ( $r < 22.5$ ) photometry in five passbands ( $u$ ,  $g$ ,  $r$ ,  $i$ , and  $z$ ; Fukugita et al. 1996; Gunn et al. 1998; Smith et al. 2002; Hogg et al. 2001) accurate to 0.02 mag (Ivezić et al. 2003). The

<sup>1</sup> Princeton University Observatory, Princeton, NJ 08544.

<sup>2</sup> H. N. Russell Fellow, on leave from the University of Washington.

<sup>3</sup> Centro de Investigaciones de Astronomía (CIDA), Apdo. Postal 264, Mérida 5101-A, Venezuela.

<sup>4</sup> Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06511.

survey sky coverage of up to 10,000 deg<sup>2</sup> in the Northern Galactic Cap will result in photometric measurements for over 100 million stars and a similar number of galaxies. Astrometric positions are accurate to better than 0".1 per coordinate (rms) for sources with  $r < 20.5$  mag (Pier et al. 2003), and the morphological information from the images allows reliable star-galaxy separation to  $r \sim 21.5$  mag (Lupton et al. 2002).

Here we use SDSS imaging data that are part of the SDSS Data Release 1 (Abazajian et al. 2003, hereafter DR1). DR1 includes 2099 deg<sup>2</sup> of five-band imaging data to a depth of  $r \sim 22.6$ . SDSS equatorial observing runs 752 and 756 overlap with the QUEST observations in a 89 deg<sup>2</sup> large region defined by  $-1^\circ < \text{decl.}_{2000} < 0^\circ$  and  $09^{\text{h}}44^{\text{m}} < \text{R.A.}_{2000} < 15^{\text{h}}40^{\text{m}}$ . This region contains about 210,000 unique, stationary unresolved sources with  $14 < r < 20$  and with mean Galactic coordinates  $l = 290^\circ$ ,  $b = 53^\circ$ . In the same region there are 162 RR Lyrae stars discovered by the QUEST survey and described by Vivas et al. (2001). The RR Lyrae stars span a range of magnitudes of  $V \sim 14\text{--}19.7$ . The discovery of these variables in this region of the sky was based on high-quality light curves, each containing 25–35 different epochs. This sample of RR Lyrae stars has a high completeness (>90%) for the *ab*-type variables (fundamental-mode pulsators). The completeness decreases for the low-amplitude type *c* RR Lyrae stars to 55%–75%, depending on the magnitude of the star.

When computing the efficiency of the selection algorithms described below, we exclude SDSS objects in the region  $-0.58 < \text{decl.} < -0.51$ , which was not observed by QUEST because it fell on a gap between the columns of CCDs in the QUEST camera. All magnitudes have been corrected by interstellar extinction using the dust maps and transformations given by Schlegel et al. (1998).

## 2.2. The SDSS Observations of the QUEST RR Lyrae Stars

We searched for the 162 QUEST RR Lyrae stars in the SDSS DR1 database<sup>5</sup> within a circle of radius 2" centered on the QUEST position and found all of them. The distribution of distances between the QUEST and SDSS positions has a median of 0".5 and an rms scatter of 0".16. (The distributions of  $\text{R.A.}_{2000}$  and  $\text{decl.}_{2000}$  differences show offsets of 0".3 for each coordinate.)

The SDSS processing flags (for details, see DR1 and Stoughton et al. 2002) indicated that nine stars may have substandard photometry (complex blends, cosmic rays, or bad pixels), with probable errors sometimes as large as 0.05 mag. Since the selection algorithm discussed here relies on accurate color measurements, hereafter we consider the sample of 153 stars with impeccable photometry. Note that the SDSS imaging in five bands is obtained within 5 minutes of time (54 s exposures); hence, the SDSS color measurements for RR Lyrae stars are not significantly affected by their variability.

The line marked by circles in the bottom panel in Figure 1 shows the distribution of differences between the mean  $V$  magnitude measured by the QUEST survey,  $V_{\text{QUEST}}^{\text{mean}}$ , and a single-epoch synthetic SDSS-based  $V_{\text{SDSS}}$  magnitude computed from (Fukugita et al. 1996)

$$V_{\text{SDSS}} - r = 0.44(g - r) - 0.02. \quad (1)$$

Reassuringly, the mean value of the shown distribution is consistent with zero to within 0.02 mag. The distribution is skewed

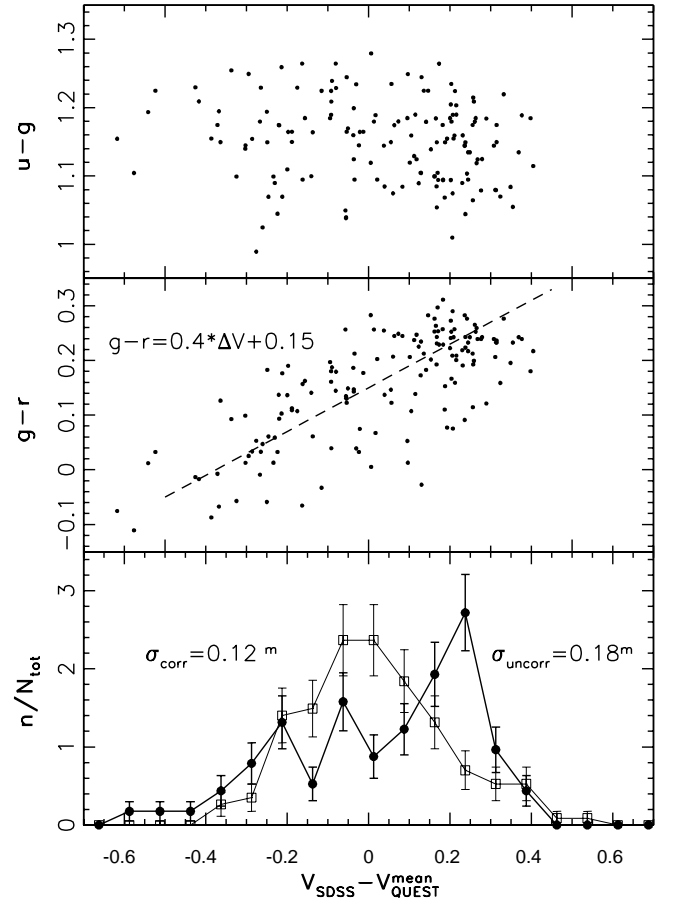


FIG. 1.—*Top and middle:* Correlations between the single-epoch  $u - g$  and  $g - r$  colors measured by SDSS and the difference between the mean  $V$  magnitude measured by the QUEST survey ( $V_{\text{QUEST}}^{\text{mean}}$ ) and a single-epoch synthetic  $V$  magnitude measured by the SDSS ( $V_{\text{SDSS}}$ ), for 153 RR Lyrae stars observed by both surveys. Note that RR Lyrae stars have bluer  $g - r$  colors when brighter, whereas there is no discernible correlation for the  $u - g$  color. The dashed line in the middle panel shows a best-fit relation between the  $g - r$  color and  $V_{\text{SDSS}} - V_{\text{QUEST}}^{\text{mean}}$ . *Bottom:* Comparison of the distribution of  $V_{\text{SDSS}} - V_{\text{QUEST}}^{\text{mean}}$  differences (circles) to the distribution of differences when  $V_{\text{SDSS}}$  is corrected for this correlation (squares; see eq. [3]).

because RR Lyrae stars have asymmetric light curves. (They spend more than 50% of their variability cycle fainter than their mean magnitude.)

The top and middle panels show the correlations between the  $u - g$  and  $g - r$  colors measured by SDSS and the  $V$  magnitude difference. As expected, RR Lyrae stars have bluer  $g - r$  colors when brighter, whereas there is no discernible correlation for  $u - g$  color, which may be due to shock wave–related activity (Smith 1995). Note that the  $g - r$  color spans twice as large a range as does the  $u - g$  color.

The  $g - r$  color is correlated with  $V_{\text{SDSS}} - V_{\text{QUEST}}^{\text{mean}}$ . The best-fit relation, shown in the middle panel by the dashed line, is

$$g - r = 0.4(V_{\text{SDSS}} - V_{\text{QUEST}}^{\text{mean}}) + 0.15. \quad (2)$$

This relation can be used to correct a bias in single-epoch SDSS measurements due to unknown phase, such that

$$V_{\text{SDSS}}^{\text{RRLyrae}} = r - 2.06(g - r) + 0.355, \quad (3)$$

where all measurements have been corrected for ISM reddening. The line marked by squares in the bottom panel in Figure 1

<sup>5</sup> Available from <http://www.sdss.org>.

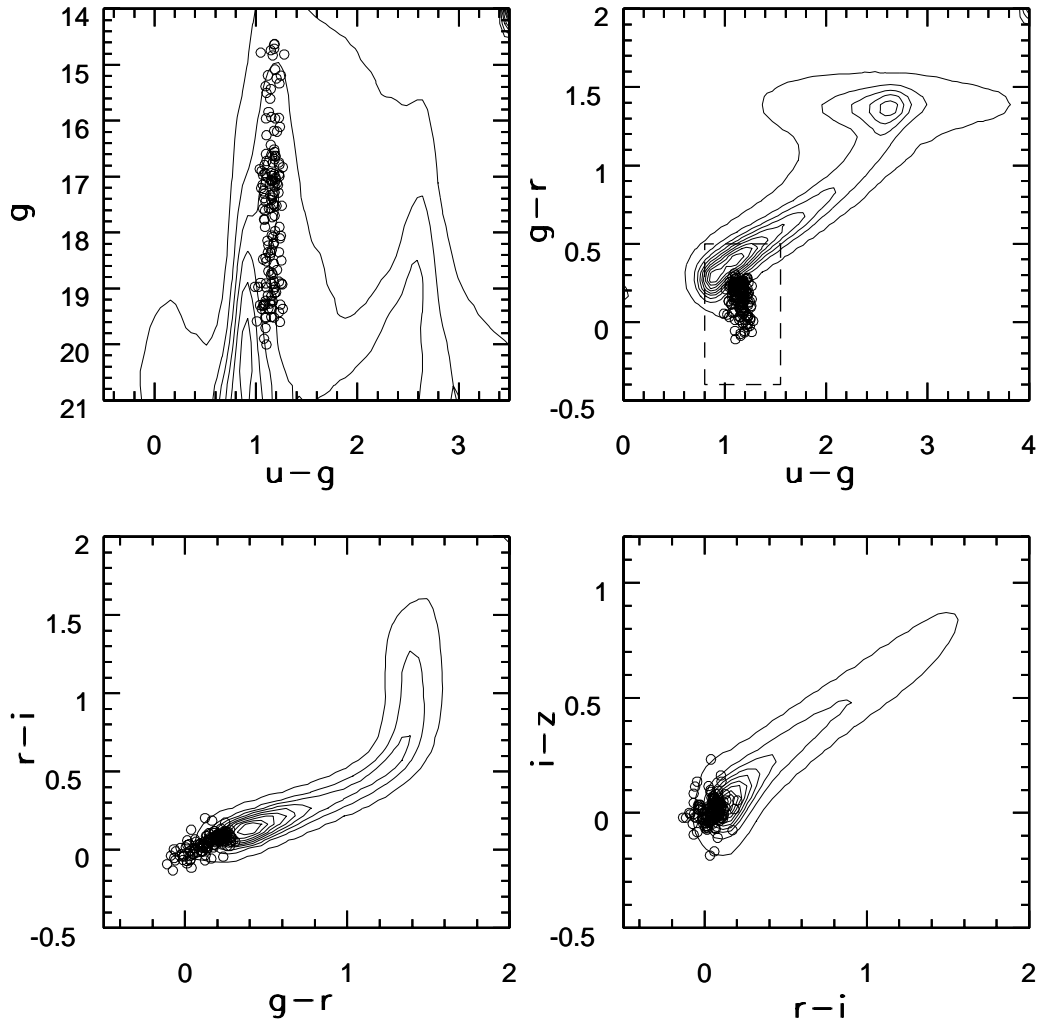


FIG. 2.—Comparison of the distribution of point sources in the SDSS color-magnitude and color-color diagrams (*linearly spaced contours*) and the distribution of RR Lyrae stars (symbols). The symbol size corresponds to 3–5 times the photometric errors, depending on the scale of individual panels. The rectangle shown by the dashed lines in the top right panel is the region that is shown magnified in Fig. 3. Note that RR Lyrae stars span a very narrow range of  $u - g$  color ( $u - g \sim 0.3 \pm 0.06$ ).

shows the distribution of  $V_{\text{SDSS}}^{\text{RRLyr}} - V_{\text{QUEST}}^{\text{mean}}$ . The rms scatter is significantly decreased compared to the scatter in  $V_{\text{SDSS}} - V_{\text{QUEST}}^{\text{mean}}$  (0.12 mag vs. 0.18 mag, as marked in the figure). This is a relation that produces unbiased RR Lyrae star distances with a minimal scatter (0.12 mag) from single-epoch SDSS measurements. It is remarkable that the scatter in mean magnitude estimated from single-epoch SDSS measurements is as small as the intrinsic uncertainty in RR Lyrae absolute magnitudes. We note that assuming a constant  $M_r$  (instead of  $M_V$ ) to determine distances results in practically no bias, but the scatter is increased to 0.20 mag. (For constant  $M_g$  the bias is 0.23 mag, with a comparable scatter.)

### 2.3. The Colors of RR Lyrae Stars in the SDSS Photometric System

Figure 2 shows the distribution of all point sources with  $r < 20$  in the SDSS color-magnitude and color-color diagrams as linearly spaced contours. In color-color diagrams, red is always toward the upper right. For a detailed description of stellar colors in the SDSS photometric system, see Finlator et al. (2000) and references therein. The 153 QUEST RR Lyrae stars are shown as symbols. The symbol size corresponds to 3–5 times the photometric errors, depending on the scale of individual panels. (For a detailed analysis of SDSS photometric errors,

see Ivezić et al. 2003.) That is, the scatter of points is due to intrinsic differences among RR Lyrae stars and the variation of colors with phase. Nevertheless, RR Lyrae stars span a very narrow range of SDSS colors. The color limits for the sample discussed here are

$$0.99 < u - g < 1.28, \quad (4)$$

$$-0.11 < g - r < 0.31, \quad (5)$$

$$-0.13 < r - i < 0.20, \quad (6)$$

$$-0.19 < i - z < 0.23. \quad (7)$$

In particular, both the range ( $\sim 0.30$  mag) and the rms scatter ( $\sim 0.06$  mag) are the smallest for the  $u - g$  color.

The RR Lyrae fraction is 1 in 1300 among all SDSS stars with  $r < 20$ . In a subsample selected using the above color cuts, the RR Lyrae star fraction is 6%. We show next how this fraction can be increased to over 60% by optimizing the color selection boundaries.

### 2.4. The Optimization of the Color Selection

RR Lyrae stars are found furthest from the locus of other stars in the  $g - r$  versus  $u - g$  color-color diagram. The relevant part of this diagram, outlined by the small rectangle in Figure 2, is

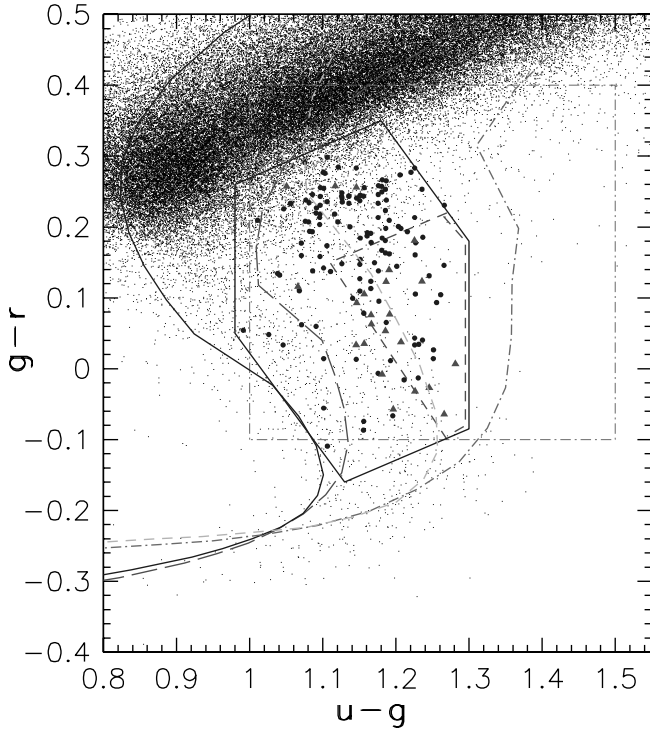


FIG. 3.—Color-selection criteria for RR Lyrae stars. The small dots show all SDSS point sources with  $r < 20$ , and the large symbols are confirmed RR Lyrae stars (filled circles are stars of type *ab* and triangles are *c* type). The photometric errors are comparable to the radii of the large dots. The solid polygon is a suggested boundary for the 100% completeness, with efficiency of 6%. The dashed lines are an example of a restricted selection boundary that results in a completeness of 28% and an efficiency of 61%. The dot-dashed rectangle is the selection boundary from a variability study by Ivezić et al. (2000), shown here for reference. The curved lines are model predictions based on Kurucz (1993) model atmospheres, for various combinations of metallicity,  $[\text{Fe}/\text{H}]$ , and gravity,  $\log(g)$ : solid,  $(-2, 4)$ ; long-dashed,  $(0, 4)$ ; short-dashed,  $(-2, 2)$ ; dot-dashed,  $(0, 2)$ . The short-dashed line corresponds to low-metallicity giants and provides the best agreement with the observed RR Lyrae distribution.

shown magnified in Figure 3. The small dots show all SDSS point sources with  $14 < r < 20$ , and the symbols are confirmed RR Lyrae stars (solid circles are *ab*-type RR Lyrae stars, and triangles correspond to the *c*-type stars). The photometric errors are comparable to the radius of the large circles, except for objects in the top left corner, which have faint *u*-band magnitudes ( $\lesssim 21.5$ ).

The lines in Figure 3 show the locus of colors synthesized from Kurucz model atmospheres (Kurucz 1993) for different metallicities and surface gravities. The position along the locus is parametrized by the effective temperature, which primarily controls the  $g - r$  color. The  $u - g$  color, in the range of colors populated by RR Lyrae stars, becomes redder as the metallicity increases and surface gravity decreases (see Lenz et al. 1998 for more details).

Most of the contamination in a sample of candidate RR Lyrae stars selected using the simple color cuts listed above comes from the main stellar locus visible in the top left corner. The second most significant source of contamination is the A star locus running from the main locus toward the bottom right. This motivates revised selection boundaries shown as the polygon outlined by solid lines. The edges with positive  $d(g - r)/d(u - g)$  slope have a constant distance from the main stellar locus

$$D_{ug} = (u - g) + 0.67(g - r) - 1.07, \quad (8)$$

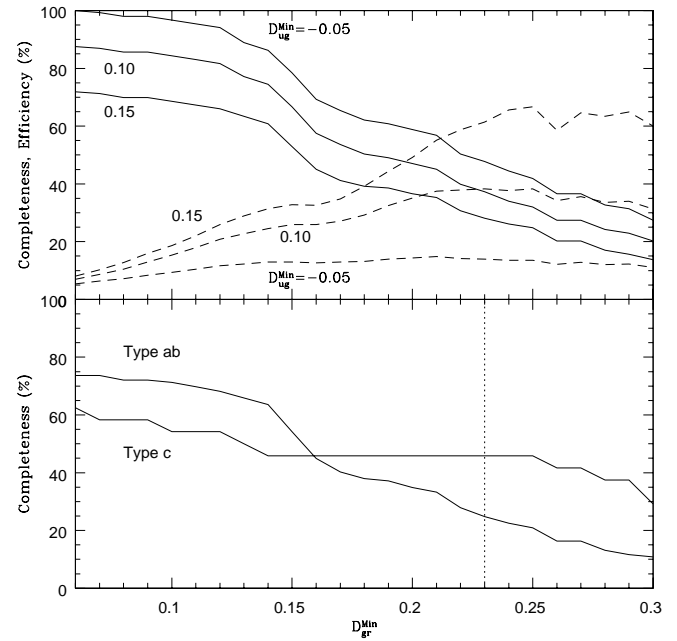


FIG. 4.—*Top*: Dependence of the selection completeness (solid lines) and efficiency (dashed lines) on  $D_{ug}^{\text{Min}}$  (different curves, as labeled) and  $D_{gr}^{\text{Min}}$ . *Bottom*: Comparison of the completeness estimates for the different types of RR Lyrae variables for a color cut with  $D_{ug}^{\text{Min}} = 0.15$ . The dotted line marks a cut with  $D_{gr}^{\text{Min}} = 0.23$ . Type *c* RR Lyrae stars have a higher selection efficiency for  $D_{gr}^{\text{Min}} > 0.16$  than type *ab* RR Lyrae stars.

and the edges with negative slope have a constant distance from the A star locus

$$D_{gr} = 0.45(u - g) - (g - r) - 0.12. \quad (9)$$

The QUEST RR Lyrae stars span the ranges  $-0.05 < D_{ug} < 0.35$  and  $0.06 < D_{gr} < 0.55$ , resulting in a selection efficiency (the fraction of RR Lyrae stars among the selected candidates) of 6%.

The selection efficiency can be further increased by increasing the lower limits on  $D_{ug}$  and  $D_{gr}$ , which are  $D_{ug}^{\text{Min}}$  and  $D_{gr}^{\text{Min}}$ , respectively (of course, the sample completeness then becomes less than 100%), and keeping the restrictions in colors ( $u - g$ ), ( $r - i$ ), and ( $i - z$ ). For example, the dashed lines show a restricted selection boundary obtained with  $D_{ug}^{\text{Min}} = 0.15$  and  $D_{gr}^{\text{Min}} = 0.23$ , which results in a completeness of 28% and 61% efficiency.<sup>6</sup>

The top panel in Figure 4 shows a detailed dependence of the completeness (solid lines) and efficiency (dashed lines) as functions of  $D_{gr}^{\text{Min}}$  for three different values of  $D_{ug}^{\text{Min}}$ , as indicated. The largest possible selection efficiency using single-epoch data is about 65%. The remaining contaminants are probably dominated by quasars (Richards et al. 2001) and nonvariable horizontal-branch stars.

The restricted color criteria preferentially select type *c* RR Lyrae stars because they have bluer  $g - r$  colors than *ab*-type stars (see Fig. 3). The bottom panel in Figure 4 compares the completeness estimates for the two different types of RR Lyrae variables for a color cut with  $D_{ug}^{\text{Min}} = 0.15$ . When  $D_{gr}^{\text{Min}} > 0.16$ , type *c* RR Lyrae stars have a higher selection efficiency. For

<sup>6</sup> The efficiency quoted here is obtained with  $r < 20$  and no limit on  $u$ . Since the faintest QUEST RR Lyrae star in the sample has  $u = 21.2$  and  $r = 19.7$ , the selection efficiency could be increased by a factor of 1.2 by requiring  $u < 21.1$  and  $r < 19.7$ .

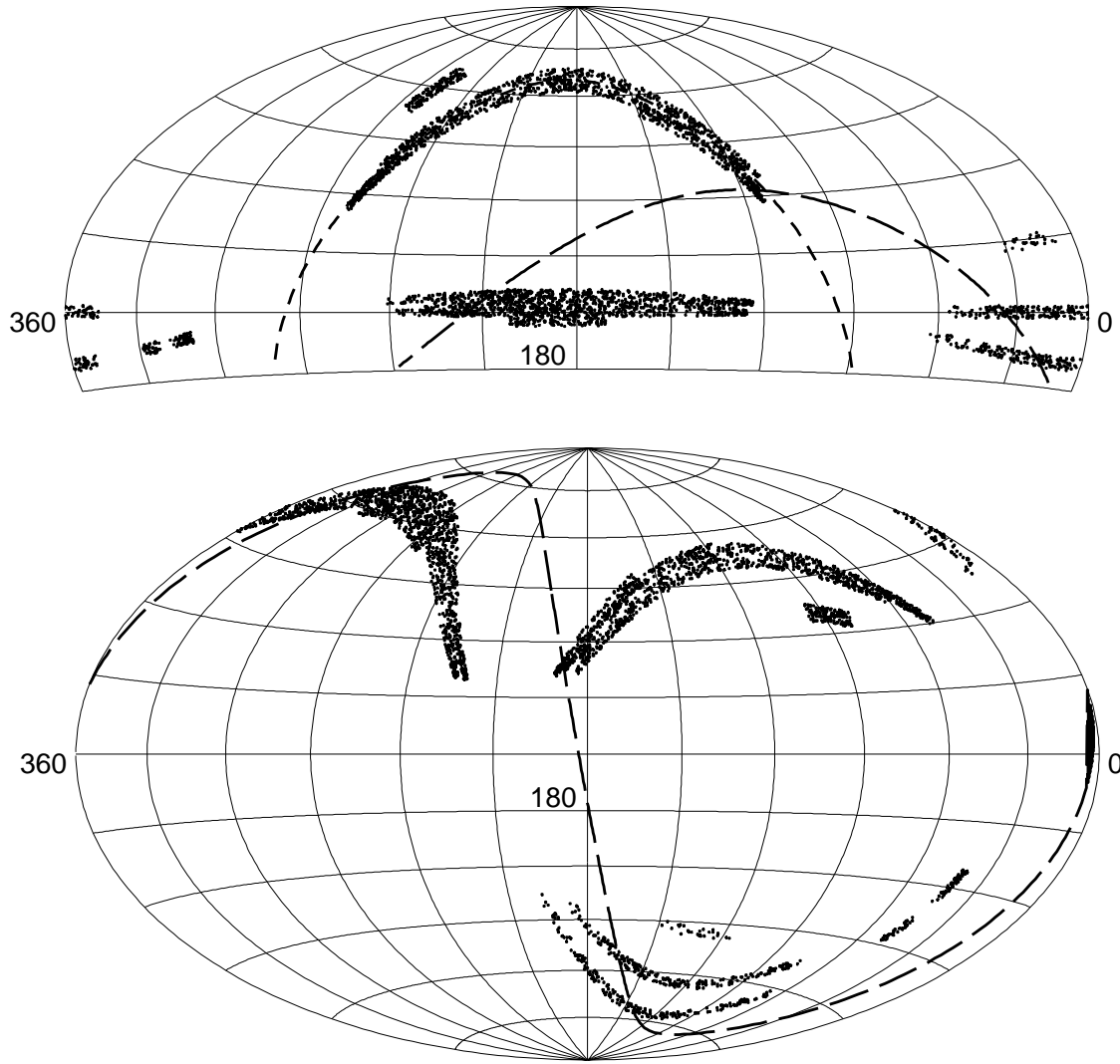


FIG. 5.—Distribution of color-selected RR Lyrae candidates from SDSS Data Release 1, shown in Aitoff projections using equatorial (*top*) and Galactic (*bottom*) coordinates. The long-dashed line indicates the position of the Sgr dwarf tidal stream, and the short-dashed line is a great circle that tracks a large fraction of the SDSS DR1 region (node =  $95^\circ$ , inclination =  $65^\circ$ ). The  $r$  vs. position angle distributions for these stars are shown in Figs. 6 (Sgr dwarf tidal stream) and 7 (the latter great circle).

example, in the region enclosed by the dashed lines in Figure 3, we recover 45% of all QUEST type  $c$  RR Lyrae stars but only 25% of the more common  $ab$  types.

To summarize, the proposed selection criteria are

$$14 < r < 20, \quad (10)$$

$$0.98 < u - g < 1.30, \quad (11)$$

$$D_{ug}^{\text{Min}} < D_{ug} < 0.35, \quad (12)$$

$$D_{gr}^{\text{Min}} < D_{gr} < 0.55, \quad (13)$$

$$-0.15 < r - i < 0.22, \quad (14)$$

$$-0.21 < i - z < 0.25, \quad (15)$$

where  $D_{ug}^{\text{Min}}$  and  $D_{gr}^{\text{Min}}$  can be chosen using Figure 4 to yield desired selection completeness and efficiency, depending on a specific purpose.

We emphasize two important disclaimers. First, the above selection method is fairly sensitive to color errors. In particular, it may be severely affected by errors in corrections for interstellar reddening at low Galactic latitudes. Analysis of the position of the stellar locus in the SDSS bands shows that the

Schlegel et al. (1998) extinction map is sufficiently accurate (errors in colors of less than 0.01 mag) to at least  $20^\circ$  from the plane (Ivezić et al. 2004a). However, we do not have any evidence yet that the proposed selection method will remain robust closer to the Galactic plane. Second, the values of completeness and efficiency shown in Figure 4 are normalized to those of the QUEST RR Lyrae survey. Although the contamination by other types of variable star in the QUEST survey is negligible (thus, the QUEST efficiency is considered to be fairly close to 100%), extensive simulations show that QUEST does miss some RR Lyrae stars (see Vivas et al. 2004 for details). The larger amplitude  $ab$ -type variables have a high completeness, between 88% and 95% depending on magnitude. The completeness decreases to 55%–75% for the  $c$ -type RR Lyrae stars. It is not possible to estimate an overall completeness because we do not know beforehand the number ratio of both types of RR Lyrae stars. However, previous surveys show that type  $ab$  stars are in general more common than the type  $c$  variables (Suntzeff et al. 1991; Vivas et al. 2004). Since  $ab$ -type RR Lyrae stars have a high completeness in the QUEST survey, we can roughly say that the completeness values shown in the top panel of Figure 4 are not far (say within about 20%) from

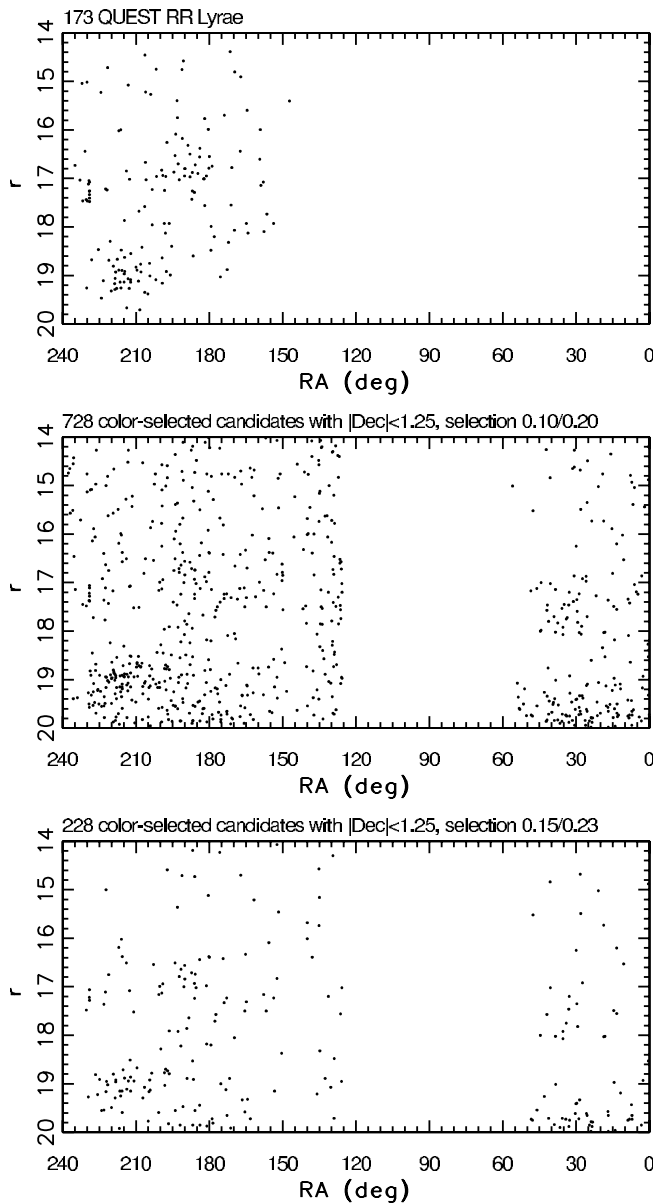


FIG. 6.—Comparison of the  $r$  vs. right ascension distribution of QUEST RR Lyrae stars (top) and color-selected candidates using SDSS single-epoch measurements (middle and bottom). The QUEST sample of confirmed RR Lyrae stars is practically complete in the region  $150^\circ < \text{R.A.} < 240^\circ$ , whereas the estimated completeness and efficiency for SDSS samples are 50%/35% and 28%/60%, for the middle and bottom panels, respectively (in the sampled R.A. range). Note that the clumps associated with the Sgr dwarf tidal tail (R.A.  $\sim 215^\circ$ ,  $r \sim 19$ ) and Pal 5 globular cluster (R.A.  $\sim 230^\circ$ ,  $r \sim 17.4$ ), as well as a clump at (R.A.  $\sim 190^\circ$ ,  $r \sim 17$ ), are recovered by color-selected SDSS samples. The clump at (R.A.  $\sim 35^\circ$ ,  $r \sim 17.5$ ) is also associated with the Sgr dwarf tidal tail.

their absolute values. Moreover, it is very likely that the color selection proposed in this work will naturally select some of the stars missed by QUEST, increasing the overall completeness of the method.

### 3. CANDIDATE RR LYRAE STARS IN SDSS DATA RELEASE 1

We apply the color selection method discussed in § 2 to SDSS Data Release 1. DR1 includes sky regions with known halo substructures traced by RR Lyrae stars and can thus be used to test the performance of the method for discovering such structures. DR1 also includes areas for which variability data

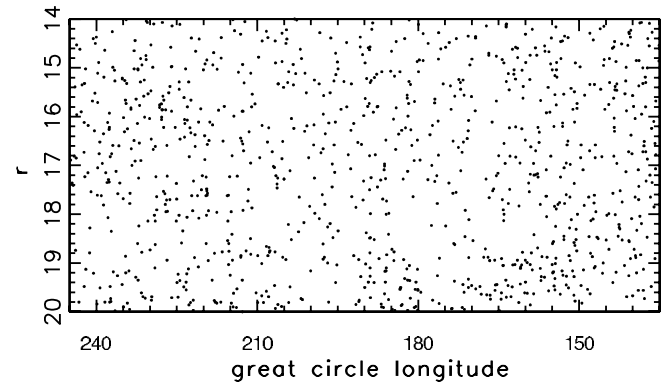


FIG. 7.—The  $r$  vs. great circle longitude distribution of SDSS color-selected candidates (selection 0.10/0.20) along a great circle marked by the short-dashed line in Fig. 5. The structure is not as pronounced along this great circle, as it is along the celestial equator (see the middle panel of Fig. 6; note different scale for x-axis).

do not exist (either from QUEST or from repeated SDSS scans) and that may exhibit previously uncharted substructure.

The conditions  $D_{ug}^{\text{Min}} = 0.10$  and  $D_{gr}^{\text{Min}} = 0.20$  result in a sample of 3643 candidate RR Lyrae stars selected from the SDSS DR1 database. Here we require that the processing flags SATURATED and BRIGHT are not set and use underreddened magnitudes for the  $14 < r < 20$  condition. (Of course, the color selection must be done with dereddened magnitudes.) A subsample satisfying  $D_{ug}^{\text{Min}} = 0.15$  and  $D_{gr}^{\text{Min}} = 0.23$  contains 896 stars. The completeness and efficiency, determined using Figure 4, for the first selection criteria are  $C = 50\%$ ,  $E = 35\%$ , and for the second selection criteria are  $C = 28\%$ ,  $E = 60\%$ . The distribution of selected candidates on the sky is shown in Figure 5 (which closely outlines the SDSS DR1 area).

#### 3.1. A Self-Consistency Test

The mean density of RR Lyrae stars can be estimated from

$$\rho_{\text{RRLyrae}} = \frac{N_s E_s}{A_{\text{DR1}} C_s}, \quad (16)$$

where the area included in DR1 is  $A_{\text{DR1}} = 2099 \text{ deg}^2$  and  $N_s$  is the number of selected stars using particular values of  $D_{ug}^{\text{Min}}$  and  $D_{gr}^{\text{Min}}$ . The estimate  $\rho_{\text{RRLyrae}}$  should be nearly the same for both samples and should agree with the value of  $\rho_{\text{RRLyrae}} \sim 1.3 \text{ deg}^{-2}$ , determined from the QUEST data for their first 400  $\text{deg}^2$  of sky (Vivas et al. 2004). The values obtained here, 1.21 and  $0.91 \text{ deg}^{-2}$ , are in good agreement with each other, indicating that the selection method is robust. (The training sample included data for a 24 times smaller area.)

We estimate that SDSS DR1 contains  $\sim 2200$  RR Lyrae stars and that 1170 are included in our sample of 3643 candidates. The smaller, more restrictive, sample of 896 stars contains 540 probable RR Lyrae stars.

#### 3.2. A Test of the Ability to Recover Halo Substructure

We analyze the spatial structure of selected candidates by examining their distribution in the  $r$  versus position diagrams for narrow strips on the sky. The equatorial strip (decl.  $\sim 0^\circ$ ) contains several known clumps of RR Lyrae stars (I00; Vivas et al. 2001).

The distribution of QUEST RR Lyrae stars in the  $r$  versus right ascension diagram along the celestial equator is shown in the top panel in Figure 6. The range of  $r$  from 14 to 20 corresponds to distances of 5–70 kpc. (Recall that the strip

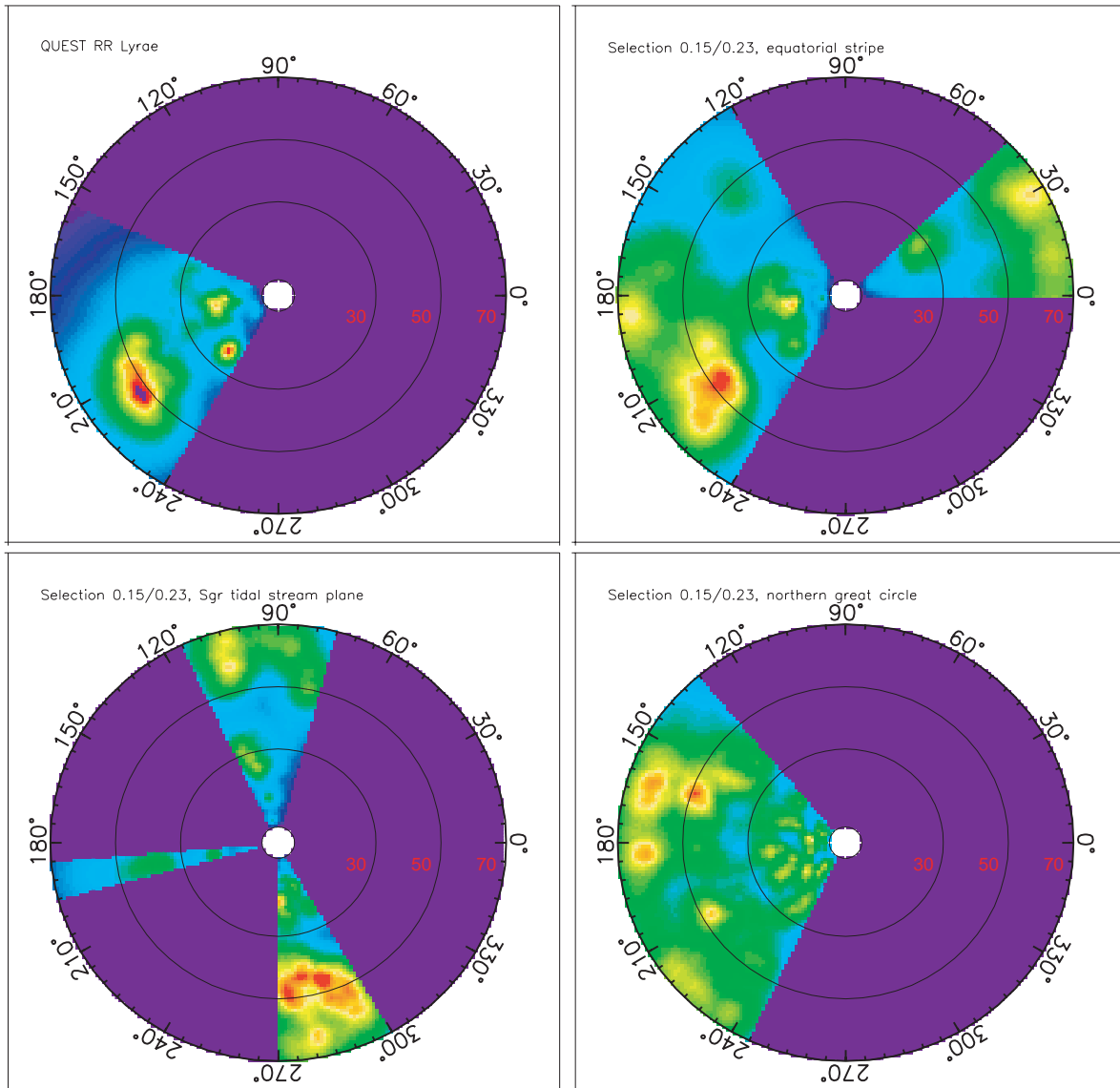


FIG. 8.—Application of the method described in Appendix B to the data shown in Figs. 6 and 7. Each panel corresponds to a different sample; the top left is the equatorial QUEST RR Lyrae sample, and the other three are selected from SDSS data using a restricted selection boundary obtained with  $D_{ug}^{\text{Min}} = 0.15$  and  $D_{gr}^{\text{Min}} = 0.23$  (which results in a completeness of 28% and an efficiency of 61%). The top right panel shows candidates with  $|\text{decl.}| < 1^{\circ}25$  (equatorial stripe; compare to the top left panel), and the two bottom panels show candidates within  $5^{\circ}$  from the great circles shown in Fig. 5 (left panel for the Sgr dwarf tidal stream and right panel for the northern great circle). 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 1000 (from light blue to red). The green color corresponds to the mean density—all 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 (using only the counts variance).

width in the declination direction for the QUEST subsample is  $1^{\circ}$ .) Three especially prominent features are the clump associated with the Sgr dwarf tidal tail (R.A.  $\sim 215^{\circ}$ ,  $r \sim 19$ ), the Palomar 5 globular cluster and associated tidal debris (R.A.  $\sim 230^{\circ}$ ,  $r \sim 17.4$ ), and a clump at (R.A.  $\sim 190^{\circ}$ ,  $r \sim 17$ ).

These three features are recovered by the color-selected SDSS DR1 candidate RR Lyrae stars, whose  $r$  versus right ascension distributions are shown in the middle and bottom panels in Figure 6. In particular, the feature at (R.A.  $\sim 190^{\circ}$ ,  $r \sim 17$ ), detected at a  $5\sigma$  level above the background by Vivas & Zinn (2003) using a complete sample of confirmed RR Lyrae stars, is clearly visible.

The color-selected samples also recover the so-called “southern” clump at R.A.  $\sim 30^{\circ}$  and  $r \sim 17$ –18 (associated with the Sgr dwarf tidal tail; see the great circle marked by the long-dashed line in Fig. 5) that was discovered using A-colored stars by Yanny et al. (2000). Furthermore, the faint clump at R.A.  $\sim 30^{\circ}$  and  $r >$

19 is present in a sample of candidate RR Lyrae stars selected from repeated SDSS observations (Ivezić et al. 2005). Given these successful recoveries of known structure, we conclude that the color-selected samples of candidate RR Lyrae stars are sufficiently clean and robust to study halo substructure out to distances of  $\sim 70$  kpc.

### 3.3. Is the Sgr Tidal Stream the Most Prominent Halo Feature?

To reliably answer the question posed in the title of this subsection, one would need an all-sky survey of several halo tracers. While such data do not exist yet (the upcoming large-scale synoptic surveys will eventually discover all halo RR Lyrae stars), a study of Two Micron All Sky Survey data by Majewski et al. (2003) provided the first all-sky view of halo structure, traced by M giants. They did not find any features that would compete with the prominence of the Sgr tidal stream. Nevertheless, M giants are not as good standard candles as RR Lyrae stars and

are more sensitive to metallicity effects. It is therefore worthwhile to examine the distribution of candidate RR Lyrae stars in areas of sky that were not explored until now.

In Figure 7 we examine magnitude-angle diagrams for candidates selected using  $D_{ug}^{\text{Min}} = 0.10$  and  $D_{gr}^{\text{Min}} = 0.20$ , along the great circle (within  $\pm 5^\circ$ ) marked by the short-dashed line in Figure 5 and defined by node =  $95^\circ$  and inclination =  $65^\circ$  relative to the celestial equator. (For more details about great circle coordinates, see Pier et al. 2003.) The structure seen in this figure is not nearly as prominent as that shown in the middle panel of Figure 6, where the candidate RR Lyrae stars were selected by the same criteria. Two possible overdensities are visible at the longitude of  $\sim 225^\circ$  and  $r \sim 16$  and the longitude of  $\sim 150^\circ - 165^\circ$  and  $r \sim 19 - 20$ . The latter is supported by the distribution of candidate RR Lyrae stars selected from repeated SDSS observations (and is probably associated with the Sgr tidal stream; Ivezić et al. 2004b), whereas such data are not available in the region of the former overdensity. We are currently investigating the distribution of candidate variable stars selected by comparing POSS and SDSS measurements (Sesar et al. 2005) in order to derive more reliable conclusions about the halo substructure in that region.

### 3.4. A Smoothed Representation of the Data

An advantage of the data representation used in Figures 6 and 7 (magnitude-angle diagrams) is its simplicity—only “raw” data are shown, without any postprocessing, and thus it is straightforward to compare the results from different surveys. On the other hand, it may be argued that the identification of overdensities discussed above is somewhat subjective, and their significance is not quantitatively estimated. Furthermore, the magnitude scale is logarithmic, and thus the spatial extent of structures is heavily distorted. In order to avoid these shortcomings, we have developed<sup>7</sup> a Bayesian self-adaptive method for estimating density traced by a point distribution, described in detail in Appendix B.

Figure 8 shows the RR Lyrae spatial density (multiplied by the cube of the Galactocentric radius) computed using this method for data displayed in Figures 6 and 7. The comparison of the variability-selected QUEST sample and color-selected SDSS sample (*top panels*) vividly demonstrates close correspondence of the implied halo substructure. Overall, the RR Lyrae spatial density follows an  $r^{-3}$  power law (see also Fig. 5 and discussion in Ivezić et al. 2004b). Significant deviations from the smooth  $r^{-3}$  power-law density are visible at Galactocentric radii beyond  $\sim 30$  kpc in all regions with available data, including those outside the Sgr tidal stream plane. However,

the overdensities in that plane are the most prominent ones. SDSS will eventually enable the construction of such maps for about  $\frac{1}{4}$  of the sky.

## 4. DISCUSSION

The robust recovery of the known halo substructures with the color selection proposed here suggests that it will be possible to constrain the halo structure out to  $\sim 70$  kpc from the Galactic center in the entire area imaged by the SDSS, and not only in the multiply observed regions. This method may result in discoveries of more Sgr dwarf debris in currently unexplored parts of the sky, which would be important to understand the evolution of the disruption of this galaxy. If events similar to the accretion and disruption of the Sgr dwarf have occurred with other galaxies, this technique has a good chance of discovering their signatures. Therefore, before large-scale variability surveys, such as Pan-STARRS (Kaiser et al. 2002) and the Large Synoptic Survey Telescope (Tyson 2002) become available, the candidate RR Lyrae stars selected using SDSS colors can be used for statistical studies of the halo substructure. The selection efficiency of  $\sim 60\%$ , with a completeness of about 30%, should be sufficient to uncover the most prominent features.

Of course, it is likely that any stream or other halo substructure found with the single-epoch color selection method proposed here will have particular kinematic and metal abundance distributions. A complete study of the origin and properties of such features will certainly require a follow-up (e.g., light curves for confirmation of bona fide RR Lyrae stars and for period determination, spectroscopy for radial velocity and metallicity measurements). Since RR Lyrae stars are fairly bright, such a follow-up may be executed using even modest-sized telescopes.

Observations for the QUEST RR Lyrae Survey were obtained at the Llano del Hato National Observatory, which is operated by CIDA for the Ministerio de Ciencia y Tecnología of Venezuela. Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation (NSF), the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max Planck Institute for Astronomy, the Max Planck Institute for Astrophysics, New Mexico State University, Princeton University, the United States Naval Observatory, and the University of Washington. A. K. V. and R. Z. were partially funded by the NSF under grant AST 00-98428. Ž. I. thanks Princeton University for generous financial support of this research.

<sup>7</sup> Mathematically, this problem is identical to the well-studied case of estimating continuous spatial density distribution of galaxies from redshift surveys (e.g., Schaap & van de Weygaert 2000). However, the samples discussed here are sparser than a typical galaxy redshift survey, and the method described in Appendix B offers a more accurate density estimation than commonly used methods, without a degradation in the spatial resolution.

## APPENDIX A

### A QUERY TO EXTRACT CANDIDATE RR LYRAE STARS FROM THE SDSS DATA RELEASE 1 DATABASE

The samples of candidate RR Lyrae stars discussed here can be extracted from the SDSS Data Release 1 database using the following query. (This particular implementation uses the EMACS interface; lines beginning with `--` are comments.)



---

```

-Flag
definitions
declare set @BRIGHT = dbo.fPhotoFlags('BRIGHT')
@BRIGHT
bigint
declare set @SATURATED = dbo.fPhotoFlags('SATURATED')
@SATURATED
bigint
declare set @bad_flags = (@SATURATED | @BRIGHT)
@bad_flags
bigint
- RR Lyrae selection cuts
- the complete sample (returns 21,426 stars)
declare @DugMin float set @DugMin = -0.05
declare @DgrMin float set @DgrMin = 0.06
- 3,643 stars for 0.10/0.20 selection, and 896 for 0.15/0.23

```

---

```

select
ra, dec, extinction_r,
psfMag_u, psfMag_g, psfMag_r, psfMag_i, psfMag_z,
psfMagErr_u, psfMagErr_g, psfMagErr_r, psfMagErr_i, psfMagErr_z
- or any other measured parameter
from
star
Where
- Check flags
(flags & @bad_flags) = 0 and nchild = 0 and
- RR Lyrae cuts
- brightness cut
psfMag_r > 14.0 and psfMag_r < 20.0 and
- color cuts
(psfMag_u - extinction_u) - (psfMag_g - extinction_g) > 0.98 and
(psfMag_u - extinction_u) - (psfMag_g - extinction_g) < 1.30 and
(psfMag_r - extinction_r) - (psfMag_i - extinction_i) > -0.20 and
(psfMag_r - extinction_r) - (psfMag_i - extinction_i) < 0.25 and
(psfMag_i - extinction_i) - (psfMag_z - extinction_z) > -0.25 and
(psfMag_i - extinction_i) - (psfMag_z - extinction_z) < 0.25
and (psfMag_u - extinction_u) - (psfMag_g - extinction_g) +
0.67*((psfMag_g - extinction_g) -
(psfMag_r - extinction_r)) -1.07 > @DugMin
and (psfMag_u - extinction_u) - (psfMag_g - extinction_g) +
0.67*((psfMag_g - extinction_g) -
(psfMag_r - extinction_r)) -1.07 < 0.35
and 0.45*((psfMag_u - extinction_u) - (psfMag_g - extinction_g)) +
(psfMag_g - extinction_g) - (psfMag_r
- extinction_r) -0.12 > @DgrMin
and 0.45*((psfMag_u - extinction_u) - (psfMag_g - extinction_g)) +
(psfMag_g - extinction_g) - (psfMag_r - extinction_r) -0.12 < 0.55
- end of Where clause

```

---

## APPENDIX B

### A BAYESIAN SELF-ADAPTIVE METHOD FOR ESTIMATING DENSITY TRACED BY A POINT DISTRIBUTION

#### B1. THE ESTIMATION OF DENSITY FIELD TRACED BY A POINT DISTRIBUTION

A continuous density distribution traced by a finite sample of discrete points can be estimated in a number of ways (e.g.,

Schaap & van de Weygaert 2000 and references therein). It is becoming increasingly popular to use the  $N$ th nearest neighbor distance,  $d_N$ , to estimate the local density (e.g., Gomez et al. 2003).<sup>8</sup> In this method, originally proposed by Dressler (1980), the volume number density,  $n_0$ , is estimated as

$$n_0 = \frac{N}{V_D(d_N)}, \quad (\text{B1})$$

---

<sup>8</sup> A similar method estimates density by counting points within a fixed volume (e.g., Blanton et al. 2003).

where  $V_D(d)$  is evaluated according to the problem dimensionality,  $D$  [e.g., for  $D = 3$ ,  $V_3(d) = 4\pi d^3/3$ , and for  $D = 2$ ,  $V_2(d) = \pi d^2$ ]. The simplicity of this estimator is a consequence of the assumption that the underlying density field is locally constant. The error of this estimate is  $N^{1/2}/V_D(d)$ , and the fractional accuracy thus increases with  $N$  at the expense of the spatial resolution. Here we show that the accuracy of this approach can be increased without a degradation in the spatial resolution by considering distances to *all*  $N$  nearest neighbors instead of only the distance to the  $N$ th neighbor.

## B2. DESCRIPTION OF THE METHOD

Distances to *all*  $N$  nearest neighbors contain information about the local density, and this information can be easily incorporated into a density estimator within the Bayesian probability framework (e.g., Press 1997). For a given set of distances to the  $N$  nearest neighbors,  $(d_k, k = 1, N)$ , the posterior probability density distribution for the local density,  $n_0$ , at the position from which the distances are measured is

$$p(n_0|\{d_k; k = 1, N\}, I) \propto p(d_N|n_0; I)p(n_0|\{d_k; k = 1, N-1\}, I), \quad (\text{B2})$$

where  $p(d_N|n_0; I)$  is the probability density distribution for the distance to the  $N$ th neighbor, given the local density  $n_0$  and prior information  $I$ . (The proportionality indicates that the integral of the right-hand side should be normalized to unity.) The advantage of the method proposed here comes from accounting for additional information contained in the distribution of the first  $N-1$  neighbors and encoded in  $p(n_0|\{d_k; k = 1, N-1\}, I)$ .

From recurrent application of equation (B2), it follows that

$$p(n_0|\{d_k; k = 1, N\}, I) \propto p(n_0|I) \prod_{k=1}^N p(d_k|n_0, I). \quad (\text{B3})$$

Assuming a uniform prior probability  $p(n_0|I)$ , the only other ingredient that needs be specified is  $p(d_k|n_0; I)$ . Analogously to the methods that treat only the  $N$ th nearest neighbor, we assume that the density is locally constant and proceed to derive  $p(d_k|n_0; I)$ .

If  $\mu$  randomly distributed points are expected within a volume  $V$ , then the probability that  $k$  points are found within the same volume is given by the Poisson distribution (e.g., Ripley 1988)

$$p(k|\mu) = \frac{\mu^k e^{-\mu}}{k!}. \quad (\text{B4})$$

The number of expected points,  $\mu = n_0 V_D(d)$ , can be conveniently parametrized as  $(d/d_0)^D$ , where  $d_0$  is the characteristic (mean) distance between two points [determined from  $n_0 V_D(d_0) = 1$ ]. The probability that the distance to the  $k$ th neighbor is between  $d$  and  $d + \delta d$  is the same as the probability that there are exactly  $k-1$  neighbors enclosed by  $d$  and follows from equation (B4). (For a treatment of non-Poisson distributions, see White 1979.) With the change of variables from  $\mu$  to  $d$ , the probability density distribution for the distance to the  $k$ th nearest neighbor becomes

$$p(d_k|d_0) = \frac{D e^{-(d_k/d_0)^D}}{d_0(k-1)!} \left(\frac{d_k}{d_0}\right)^{Dk-1}. \quad (\text{B5})$$

The most probable distance to the  $k$ th neighbor,  $d_k^{\max}$ , is given by  $(d_k^{\max}/d_0)^D = (Dk-1)/D$ , and the expectation value is  $\langle d_k \rangle = d_0 k^{1/D}$ .

For illustration,  $p(d_k|d_0)$  as a function of  $d_k/d_0$  is plotted for  $D = 2$  and  $k = 2, 4$ , and  $8$  in the top panel in Figure 9. The simple estimator given by equation (B1) can be derived from the distance to the  $N$ th nearest neighbor,  $d_N$ , by adopting  $\langle d_N \rangle = d_N$ , leading to  $n_0 = (d_N/d_0)^D/V_D(d_N) = N/V_D(d_N)$ .

In the two-dimensional case adopted here to analyze the distribution of candidate RR Lyrae stars (because the SDSS Data Release 1 footprint consists of two elongated strips; see Fig. 5), the probability density distribution for  $d_0$ , which is related to the local density via  $n_0 = \pi d_0^2$ , is thus determined from<sup>9</sup>

$$p(d_0|\{d_k; k = 1, N\}) \propto \prod_{k=1}^N \frac{2e^{-(d_k/d_0)^2}}{d_0(k-1)!} \left(\frac{d_k}{d_0}\right)^{2k-1}. \quad (\text{B6})$$

A few steps in a typical realization of this computation are shown in the middle panel in Fig. 9, where  $p(d_0)$  is evaluated for  $N = 2, 4, 8$ , and  $12$ , for a random sample with true  $d_0 = 1$ . Note how the posterior probability distribution approaches the true value and becomes narrower as  $N$  increases. We find that after  $N \sim 8$  the mean of the probability distribution becomes stable and only the distribution width decreases proportionally to  $N^{-1/2}$ . By repeating this computation 40,000 times for random subsamples with  $d_0 = 1$ , we have established that the distribution of the mean of the posterior probability distribution is similar to a typical posterior probability distribution. That is, the width of the posterior probability distribution is a reliable estimator of the errors in the determination of the mean posterior density.

The bottom panel in Figure 9 further illustrates the improvement in the local density estimate due to incorporating distances to all  $N$  nearest neighbors, instead of only the last one (again for a sample with  $D = 2$  and true  $d_0 = 1$ ). The dashed curve shows the probability density distribution for local density based on only the eighth nearest neighbor. The dot-dashed curve shows the probability density distribution based on all seven nearest neighbors, computed using equation (B6). This distribution is used as the prior  $p(n_0|\{d_k; k = 1, N-1\}, I)$  in equation (B2) to evaluate the final probability density distribution for local density based on all eight nearest neighbors, shown by the solid line (i.e., the dot-dashed and dashed curves are multiplied and renormalized to obtain the solid curve). The difference between using all eight neighbors and only the eighth one is vividly illustrated by the large differences between the solid and dashed curves. Using all eight nearest neighbors, the density is estimated with about 10% fractional accuracy. For comparison, the mean density fractional error is about 35% when using only the eighth neighbor. This improvement of about a factor of 3.5 in the accuracy of density estimate persists for larger  $N$ . Equivalently, for the same density estimator accuracy, one can achieve the same improvement of a factor of 3.5 in the obtained resolution.

<sup>9</sup> Eq. (B6) can be recast in a form that allows much simplified computation of the expectation value for the local density and its uncertainty (such that the dependences on  $d_k$  and  $d_0$  are separated and the final expressions involve only a summation of  $d_k^D$ , without a need to evaluate the full posterior probability distribution [P. Wozniak 2004, private communication]).

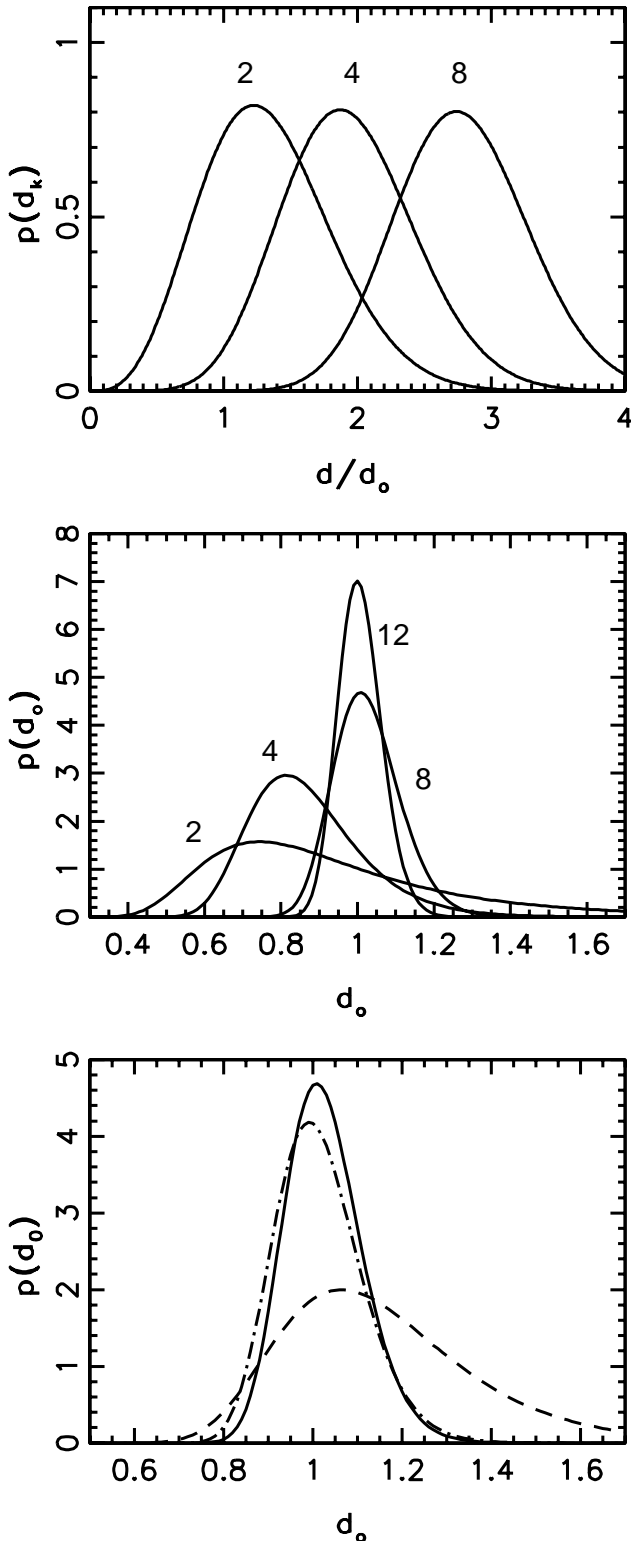


FIG. 9.—*Top*: The  $k$ th nearest neighbor distance distribution for a random two-dimensional sample with  $k = 2, 4$ , and  $8$ . The distances are scaled by the mean distance between two sources,  $d_0$ . *Middle*: Evolution of the posterior probability density distribution evaluated using eq. (B6) for  $k = 2, 4, 8$ , and  $12$  and a sample with true  $d_0 = 1$ . The dashed line in the bottom panel shows a typical probability density distribution for local density based on only the eighth nearest neighbor, in a random sample with  $d_0 = 1$ . The dot-dashed curve shows the probability density distribution based on all seven nearest neighbors, computed using eq. (B6). This distribution is used as the prior  $p(n_0|d_k, k = 1, N - 1; I)$  in eq. (B2) to evaluate the final probability density distribution for local density based on all eight nearest neighbors, shown by the solid line.

### B3. THE PERFORMANCE TESTS

To test the method described in § B2, we generate a mock sample of RR Lyrae stars with the same distance limits (from 5 to 70 kpc) as the candidate sample described in § 3. These numerical values, as well as the units, are of no consequence for the method performance and are simply adopted to ease the comparison with the corresponding results for the observed sample (shown in Fig. 8). Two realizations of the mock sample, one with 9651 stars and another with 190 stars, are shown in the top panels in Figure 10. The latter realization is used to test the method, and the former is shown to outline the adopted three components of the mock sample:

1. Uniform background (with an arbitrary density, controlled by the sample size).
2. A component centered on  $(X, Y) = (-25, 25)$ , with the number density falling off with the distance from that point,  $r$ , proportionally to  $r^{-1}$ , within a circle with a radius of 20. The mean density of this component is 2.8 times higher than the background density, resulting in 31 points (see the top right panel in Fig. 10). The corresponding total signal-to-noise ratio relative to the background is  $9.3 = (31 \times 2.8)^{1/2}$ .
3. A component with constant density inside an ellipse centered on  $(35, -10)$ , with axes equal to 36 and 18 (along  $x$  and  $y$ , respectively), and further constrained to  $10 < x < 60$ . The density of this feature is only 1.37 higher than the background density (i.e., the total density within the feature is 2.37 higher than the background) and with 23 stars in the sample corresponds to a signal-to-noise ratio of 5.6.

When the number of points that sample the underlying density distribution is sufficiently large,<sup>10</sup> such as in the top left panel in Figure 10, practically any method from the literature, including simple Gaussian smoothing, would suffice to uncover the three components in the mock sample. However, when the number of points is small, such as in the top right panel in Figure 10, it is much harder to recognize and characterize the underlying components. As an example, the middle panels demonstrate that the performance of Gaussian smoothing is not satisfactory. The middle left panel shows the result of a Gaussian smoothing with  $\sigma = 5$ . This smoothing length is approximately equal to the mean distance between the points and in some regions results in artificial underdensities (blue regions, e.g., around the lower left edge) and overdensities, thus implying spurious substructure. When the smoothing length is doubled, most of these spurious features disappear, but now the two real features are oversmoothed. In particular, the  $r^{-1}$  signature of the top left component is not recovered (see below). The number of points in the sample is simply too small (190) for Gaussian smoothing to recover the underlying structure.

The performance of the method proposed here is illustrated in the bottom panels. The shown density estimates are obtained using distances to eight and 12 nearest neighbors, for the left and right panel, respectively. Whereas results superior to Gaussian smoothing are already achieved using eight nearest neighbors (*bottom left*), the extension to 12 nearest neighbors (*bottom right*) removes all the spurious noise visible as slight overdensities in the background level outside the two features (e.g., near the upper right edge).

We used a 5 kpc wide strip parallel to the  $x$ -axis and centered on  $y = 25$  to measure how well the density profile of the upper

<sup>10</sup> The mean distance between the points has to be smaller than the characteristic size of the density features.

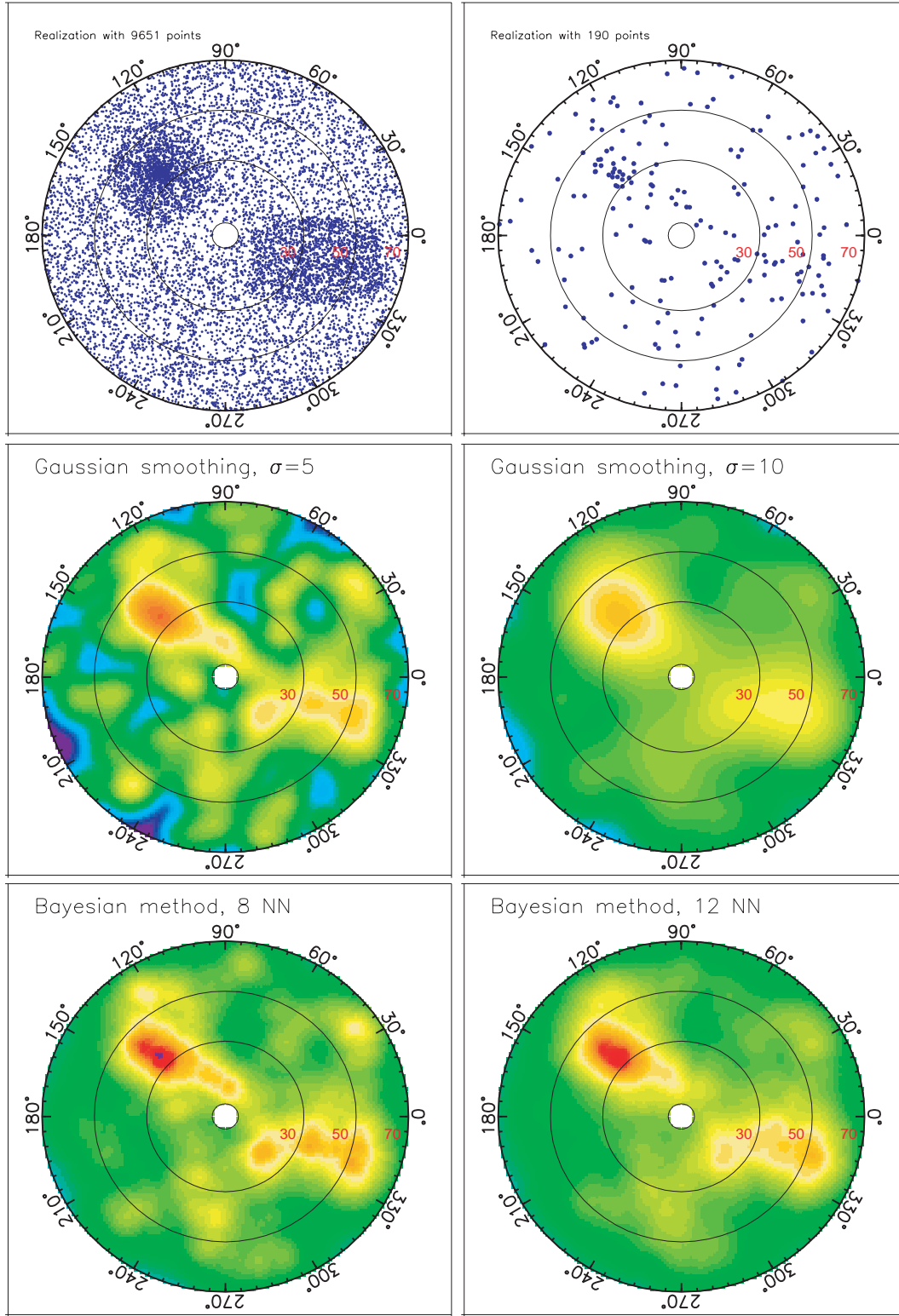


FIG. 10.—Tests of the proposed method for estimating the density traced by a point distribution. The color scheme in the bottom four panels represents the logarithm of the estimated number density with a dynamic range of 1000 (from light blue to red), for the sparsely sampled distribution shown in the top right panel. The underlying distribution is illustrated in the top left panel, using a  $\sim 50$  times larger sample. The green color corresponds to the background density outside the two features visible in the top left panel. The middle two panels show results based on a Gaussian smoothing kernel with two smoothing sizes, and the bottom two panels show results based on the Bayesian self-adaptive method proposed here. For more details please see Appendix B.

left feature is reproduced. The Bayesian method properly recovers the  $r^{-1}$  density distribution (with an error for the power-law index of  $\sim 0.1$ ) from the outer feature radius (20 kpc) to about 2–3 kpc from its center. Within the feature core the estimated density is nearly constant (because of lack of stars in the mock sample). For comparison, the Gaussian smoothing with  $\sigma = 10$  kpc produces a nearly flat core all the way to 12 kpc from the

feature center and thus fails to recognize the underlying  $r^{-1}$  density profile.

We note that an estimate based on only, for example, the eighth nearest neighbor is clearly superior to Gaussian smoothing, but not as good as when using all eight neighbors. The main difference is increased noise on small spatial scales, which results in spurious substructure in the two dominant features.

## REFERENCES

- Abazajian, K., et al. 2003, *AJ*, 126, 2081  
 Blanton, M. R., et al. 2003, *ApJ*, 594, 186  
 Dressler, A. 1980, *ApJ*, 236, 351  
 Finlator, K., et al. 2000, *AJ*, 120, 2615  
 Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748  
 Gomez, P. L., et al. 2003, *ApJ*, 584, 210  
 Gould, A., & Popowski, P. 1998, *ApJ*, 508, 844  
 Gunn, J. E., et al. 1998, *AJ*, 116, 3040  
 Helmi, A. 2002, *Ap&SS*, 281, 351  
 Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, *AJ*, 122, 2129  
 Ivezić, Ž., et al. 2000, *AJ*, 120, 963 (100)  
 ———. 2003, *Mem. Soc. Astron. Italiana*, 74, 978  
 ———. 2004a, *Astron. Nachr.*, 325, 583  
 ———. 2004b, in *ASP Conf. Ser. 317, Milky Way Surveys: The Structure and Evolution of Our Galaxy*, ed. D. Clemens, R. Shah, & T. Brainerd (San Francisco: ASP), in press (astro-ph/0309074)  
 ———. 2005, in *ASP Conf. Ser. 327, Satellites and Tidal Streams*, ed. F. Prada, D. Martinez-Delgado, & T. J. Mahoney (San Francisco: ASP), in press (astro-ph/0309075)  
 Johnston, K. V., Hernquist, L., & Bolte, M., 1996, *ApJ*, 465, 278  
 Kaiser, N., et al. 2002, *Proc. SPIE*, 4836, 154  
 Kurucz, R. L. 1993, *Model Atmospheres* (Strasbourg: CDS), <http://vizier.cfa.harvard.edu/viz-bin/Cat?VI/39>  
 Layden, A. C., Hanson, R. B., Hawley, S. L., Klemola, A. R., & Hanley, C. J. 1996, *AJ*, 112, 2110  
 Lenz, D. D., Newberg, H. 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, *Proc. SPIE*, 4836, 350  
 Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, *ApJ*, 599, 1082  
 Mayer, L., Moore, B., Quinn, T., Governato, F., & Stadel, J. 2002, *MNRAS*, 336, 119  
 Pier, J. R., Munn, J. A., Hindsley, R. B., Hennesy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, *AJ*, 125, 1559  
 Press, W. H. 1997, in *Unsolved Problems in Astrophysics*, ed. J. N. Bahcall & J. P. Ostriker (Princeton: Princeton Univ. Press), 49  
 Richards, G. T., et al. 2001, *AJ*, 121, 2308  
 Ripley, B. D. 1988, *Statistical Inference for Spatial Processes* (Cambridge: Cambridge Univ. Press)  
 Schaap, W. E., & van de Weygaert, R. 2000, *A&A*, 363, L29  
 Schlegel, D., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525  
 Sesar, B., et al. 2005, *AJ*, submitted (astro-ph/0403319)  
 Smith, H. A. 1995, *RR Lyrae Stars* (Cambridge: Cambridge Univ. Press)  
 Smith, J. A., et al. 2002, *AJ*, 123, 2121  
 Stoughton, C., et al. 2002, *AJ*, 123, 485  
 Suntzeff, N. B., Kinman, T., & Kraft, R. P. 1991, *ApJ*, 367, 528  
 Tyson, J. A. 2002, *Proc. SPIE*, 4836, 10  
 Vivas, A. K., & Zinn, R. 2003, *Mem. Soc. Astron. Italiana*, 74, 928  
 Vivas, A. K., et al. 2004, *AJ*, 127, 1158  
 ———. 2001, *ApJ*, 554, L33  
 White, S. M. D. 1979, *MNRAS*, 186, 145  
 Yanny, B., et al. 2000, *ApJ*, 540, 825  
 York, D. G., et al. 2000, *AJ*, 120, 1579