The Milky Way Tomography with SDSS: V. Dissecting Dust

Michael Berry[1], Željko Ivezić[1], Branimir Sesar[1], Mario Jurić[2], Jillian Bellovary[1], Keira J. Brooks[1], Douglas Finkbeiner[3], David Schlegel[4], Kevin R. Covey[3], Lynne Jones[1], Constance M. Rockosi[5], Brian Yanny[6], Robert H. Lupton[7], Gillian R. Knapp[7], James E. Gunn[7], and who else?

July 31, 2009

ABSTRACT

We utilize SDSS photometry for 245 million stars to simultaneously obtain best-fit main-sequence stellar energy distribution and the amount of dust extinction towards each star. Using a subsample of 24 million stars with 2MASS photometry, which enables more robust results, we show that SDSS photometry is sufficient to break degeneracies between intrinsic stellar color, dust amount, and dust properties. These fits enable detailed and robust studies of the dust properties, and of the stellar spatial distribution at low galactic latitudes. Our results are in good agreement with the SFD dust maps at high galactic latitudes, and constrain the ratio of total to selective absorption to $R_V = 3.0 \pm 0.2$. At low galactic latitudes ($|b| < 5°$), we demonstrate that the SFD map cannot be reliably applied without accounting for the fact that most stars are embedded in dust, rather than behind it as is the case at high galactic latitudes. We find indications that sometimes the SFD map overestimates the dust extinction even when distance effects are accounted for. In cases where such discrepancies are robustly detected, they seem correlated with the distribution of molecular gas. We make these best-fit parameters, as well as all the input data, publicly available.

Subject headings: methods: data analysis — stars: statistics — Galaxy: disk, stellar content, interstellar medium

---

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

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

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

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

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

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

[7]Princeton University Observatory, Princeton, NJ 08544

1. Introduction

SDSS has enabled detailed mapping of the stellar distribution at high galactic latitudes. In addition to large number of stars ($> 10^8$), the key advantage of SDSS is reasonably accurate distance determination ($\sim$ 10%) based on photometric parallax approach. For example, using this method Jurić et al. (2008, hereafter J08) mapped the spatial distribution of stars over the distance range from 100 pc to 10 kpc. In addition to quantifying a smooth background density distribution using exponential disk and power-law halo model, they found a number of local overdensities embedded in disk and halo.

J08 study was based on SDSS data at high galactic latitudes ($|b| > 30°$). Meanwhile, the second phase of SDSS has delivered imaging data for ten $\sim$2.5 degree wide strips that cross the Galactic plane (the so-called SEGUE data). At least in principle, these data can be used to test J08 models much closer to the mid-plane, and to search for evidence of effects such as disk warp and disk flare.

However, at low galactic latitudes sampled by SEGUE data, there are severe problems with the extinction corrections (based on the maps from Schlegel, Finkbeiner & Davis 1998; hereafter SFD). When the full SFD extinction correction is applied, the resulting color-magnitude and color-color diagrams have dramatically different morphology than observed at high galactic latitudes. Models developed by J08 suggest that these problems are due to the fact that stars are embedded in the dust layer, rather than behind it, as is true for most stars at high galactic latitudes. Therefore, in order to fully exploit SEGUE data, distance to stars and the amount of dust extinction have to be determined simultaneously. Our additional motivation for quantifying stellar and dust distribution close to the plane is to inform the planning of LSST survey, which is considering deep multi-band coverage of the Galactic plane (Ivezić et al. 2008). Of course, in addition to studying the spatial distribution of stars, the constraints on dust properties are interesting in their own right.

The amount of dust can be constrained by measuring dust extinction, typically at UV, optical and near-IR wavelengths, or dust emission at far-IR wavelengths. The most widely used dust map (SFD) is derived from observations of dust emission at 100 μm and 240 μm, and has angular resolution of $\sim$6 arcmin. It has been claimed that SFD maps overestimate the dust column by 20-30% when the dust extinction in the r band, A_r, exceeds 0.5 mag (e.g. Arce & Goodman 1999), possibly due to confusion with point sources. In addition, a generic shortcoming of the emission-based methods is that they provide no constraint on the three-dimensional distribution of dust; instead, only the total amount of dust along the line of sight is measured.

Some recent extinction-based determinations of dust distribution used SDSS, 2MASS

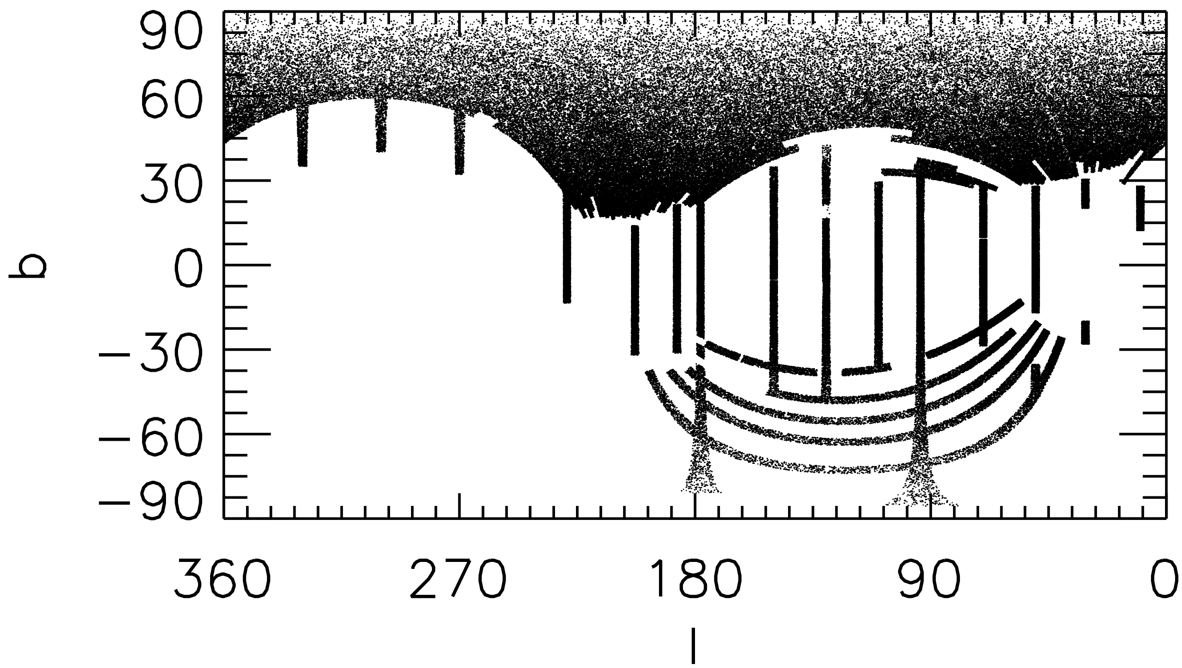


Fig. 1.— The sky coverage for SDSS Data Release 7, used in this study. The ten SEGUE stripes cross $b = 0$.

and a few other photometric surveys to map the dust distribution close to the galactic plane. For example, ...

Add paper outline at the end.

2. Data and Methodology

Describe SDSS data, and a few other details

This is a lot of data: 5-band photometry for 245 million stars. The distribution on the sky is shown in Figure 1.

2.1. SDSS/SDSS-2MASS Data Set

(Need to talk more with Branimir)

The combined catalog contains sources from SDSS DR7.2 and 2MASS whereas the second catalog has only sources from SDSS DR7.2. Stars in the SDSS only catalog are also filtered to have galactic latitudes less than 30. SDSS sources in the catalog are not DEBLENDED_AS_MOVING, SATURATED, BLENDED, BRIGHT, nor NODEBLEND, and have nchild == 0 and rModelMag < 21. 2MASS sources have rd_flag == 222, bl_flag == 111, and cc_flag == 0 per Covey et al. (2007) recommendations. 2MASS sources are also filtered to have r-band magnitudes less than 21. The SDSS-2MASS combined catalog contains 25 million matched SDSS-2MASS sources with a matching radius of 1.5 arcsec.

Stars with u-band errors greater than 1.5 magnitudes have their u-band magnitude set to 999.9 and their u-band error set to 9999.9 so as to have a negligible impact on the fitting procedure. Stars with errors in the g, r, i, z, J, H, and K-bands greater than 0.5 have their respective magnitudes and errors set to 999.9 and 9999.9. Furthermore, a minimum error was set to 0.02 magnitudes in all bands.

The Vega-based 2MASS photometry is translated to AB system following Finlator et al. (2000)

$$\begin{aligned} J_{AB} &= J_{2MASS} + 0.89 \\ H_{AB} &= H_{2MASS} + 1.37 \\ K_{AB} &= K_{2MASS} + 1.84 \end{aligned} \tag{1}$$

Note that these corrections have no impact on fitting.

2.2. Illustration of the Method

We make two basic assumptions. First, we assume that the median stellar locus quantified by Covey al al. (2007) is a good description of stellar colors at all galactic latitudes. Second, we assume that the normalized dust extinction curve, A_λ/A_r, can be described as a function of single parameter, R_V. Therefore, for a given set of measured colors, four in SDSS-only case, and seven in SDSS-2MASS case, we fit three free parameters: stellar model, m, dust amount, A_r, and R_V.

When the number of measured colors is small, or when the sampled wavelength range is not sufficiently wide, the best-fit solutions can be degenerate. Figure 2 illustrates an example of degenerate solutions in the $r-i$ vs. $g-r$ space, and how degeneracies are broken when the $i-z$ color is added to the data. Because the direction of the reddening vector in the $i-z$ vs. $r-i$ color-color diagrams is essentially independent of R_V, the measured $r-i$ and

$i-z$ colors provide robust constraints for m and A_r, irrespective of R_V. The addition of the measured $g-r$ color then constrains R_V.

The stellar locus in the $i-z$ and $r-i$ color-color diagram and the reddening vector are not perpendicular and thus there is non-zero covariance between the best-fit m and A_r values. The addition of other bands, e.g. 2MASS bands to SDSS bands, alleviates this covariance somewhat but not completely. We quantify this effect using Monte Carlo simulations below.

2.3. Covey et al. SEDs

ZI

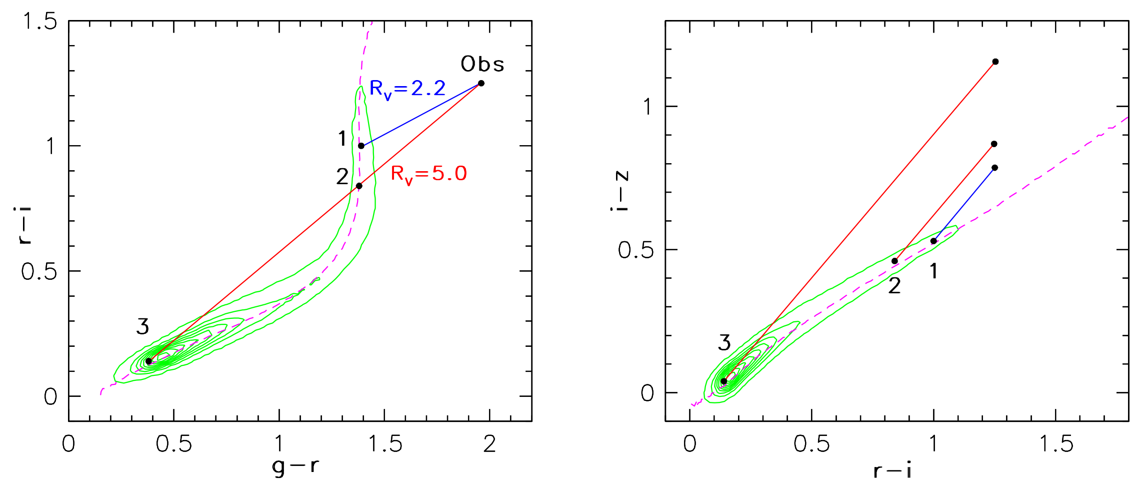


Fig. 2.— An illustration of the constraints on intrinsic stellar colors, extinction in the r band, A_r, and the ratio of total to selective extinction, R_V. In both diagrams, the linearly-spaced contours show the main stellar locus as observed at high galactic latitudes. The dashed lines mark the median stellar locus from Covey et al. (2007). In the left panel, the dot marked "Obs" represents a hypothetical observation. Depending on adopted R_V, as marked, different combinations of intrinsic stellar colors (i.e., the position along the stellar locus) and A_r are consistent with the observed $g-r$ and $r-i$ colors. Multiple solutions are possible even for a fixed value of R_V. The three solutions marked 1-3 correspond to (R_V,A_r)= 1:(2.2,1.0), 2:(5.0,2.2), and 3:(5.0,6.0). As shown in the right panel, these degeneracies can be broken if the $i-z$ color is also available: the three (R_V,A_r) combinations have different reddened $i-z$ colors. Note that reddening vectors in the right panel are parallel despite very different R_V values.

The model space includes 228 steps along the stellar locus, parametrized by the $g - i$ color ($-0.25 < g - i < 4.50$). Dominated by main sequence stars, but good description of giants, too. It fails big time for unresolved pairs of white and red dwarfs, quasars, white dwarfs and other very hot stars, L/T dwarfs, etc. However, main sequence stars are expected to contribute more than 95% of the sample at the faint magnitude levels probed by SDSS, and the fitting failures can be easily recognized as large χ^2 outliers.

2.4. R_V Values

ZI

Waiting on Mike to finish 3D fits...

Need to incorporate stuff from Meyer et al. poster: using the position of the stellar locus at high galactic latitudes, they found $R_V = 3.05 \pm 0.05$. This was our motivation to first attempt 2-dimensional (m and A_r) fits with $R_V = 3.1$.

2.5. Best-fit Models and χ^2 Minimization

The best-fit stellar model and dust extinction are found by comparing the observed SEDs to the empirical SEDs quantified by Covey et al. with varying amounts of dust extinction. This comparison is based on colors and minimizes χ^2 defined as

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{c_i^{obs} - c_i^{mod}}{\sigma_i} \right), \tag{2}$$

where c_i^{obs} are N observed adjacent (e.g., $u - g$, $g - r$, etc.) colors ($N = 4$ for SDSS-only dataset, and $N = 7$ for SDSS-2MASS dataset), and the model colors are constructed using extinction-corrected magnitudes

$$m_\lambda^{corr} = m_\lambda^{obs} - A_\lambda, \tag{3}$$

with $\lambda = (ugriz[JHK])$, and extinction correction is a two-parameter function

$$A_\lambda = C_\lambda(R_V)\, A_r. \tag{4}$$

The errors, σ_i, are computed from photometric errors quoted in catalogs, with a floor of 0.02 mag added in quadrature to account for plausible systematic errors (such as calibration errors). When the quoted errors exceed 0.5 mag (1.5 mag in the u band), they are set to 999.9 (that is, such a data point is effectively not used in fitting).

For each star, all 228 Covey et al. models are tried, with dust extinction values in the range $0 \leq A_r \leq 3$ and 0.02 mag wide steps. If the A_r value with minimum χ^2 is greater than 2, then the trials are extended to the $A_r < 5$ range, and if the new A_r value with minimum χ^2 is greater than 3, then the trials are extended to the maximum value $A_r = 10$.

Once the minimum χ^2 is located, χ^2_{min}, an ellipse is fit to the section of the χ^2 surface defined by $\chi^2 < \chi^2_{min} + 6.17$ (i.e., within 2σ deviation for 2 degrees of freedom):

$$\chi^2(m, A_r) = a(m - m^*)^2 + b(m - m^*)(A_r - A_r^*) + c(A_r - A_r^*)^2 \tag{5}$$

were m is the model index, and m^* and A_r^* are the values corresponding to χ^2_{min}. Using the best-fit parameters a, b and c, the (marginalized) model and A_r errors can be computed from

$$\sigma_m = \left(a - \frac{b^2}{4c}\right)^{-\frac{1}{2}} \tag{6}$$

$$\sigma_A = \left(c - \frac{b^2}{4a}\right)^{-\frac{1}{2}} \tag{7}$$

The χ^2 surface for stars with $\chi^2_{min} > 200$ is not fit with an ellipse and such stars are instead marked as bad fits.

2.6. Distance Estimates

Given the best-fit model, the distance to each star is computed using the best-fit $g - i$ color, extinction-corrected r band magnitude, and expressions for $Mr(g - i, [Fe/H])$ from Ivezić et al. (2008). Given that for the vast majority of stars metallicity is not available, and that most stars at low galactic latitudes are disk stars, we adopt $[Fe/H] = -0.4$ for all stars. These expressions are valid only for main sequence stars. For example, giants will have grossly underestimated distances (and they can be recognized as stars with much large best-fit extinction values than those of nearest neighbor stars).

2.7. Monte Carlo Simulations

In order to test the implementation of χ^2 minimization algorithm, and to study the dependence of best-fit parameter uncertainties on photometric errors, the amount of extinction, and the intrinsic stellar color, we perform Monte Carlo simulations.

In the first test, we study the variation of best-fit parameters with photometric errors, where the latter are generated using gaussian distribution and four different widths: (0.01, 0.02, 0.04, 0.08) mag. The noiseless "observed" magnitudes for a fiducial star with intrinsic color $g - i = 1.95$ and $A_r = 1.5$, are convolved with photometric noise generated independently for each band, and the resulting noisy colors are used in fitting. The errors in best-fit models and A_r are illustrated in Figures 3 and 4.

HERE ADD CONCLUSIONS AFTER MIKE COMPUTES RMS FOR PANELS IN FIGS. 3 and 4.

More or less, errors in best-fit model scale with photometric errors, but we need to see if this scaling is linear and give the scaling coefficient.

In the second test, we have investigated the covariance between the best-fit model and A_r values. Figure 5 shows the χ^2 surface for a blue and a red star, and for two values of A_r, when only SDSS bands are used in fitting and gaussian noise with $\sigma = 0.02$ mag is assumed. The best-fit model-A_r covariance is larger for the bluer star, in agreement with the behavior illustrated in Figure 2 (the angle between the reddening vector and the stellar locus is smaller for the blue part of the locus, than for the red part). When the 2MASS bands are added, the χ^2 surface is essentially unchanged.

Table 1: SDSS AND SDSS-2MASS DATA FILES $R_V = 3.1$

Definition	stars	size (mb)	stars	size (mb)
$\|b\| < 30, l$ 50	7,628,624	2,200	1,533,211	700
$\|b\| < 30, l$ 70	6,317,564	1,900	1,427,507	600
$\|b\| < 30, l$ 90	4,404,358	1,300	1,238,009	600
$\|b\| < 30, l$ 110	3,449,763	1,000	1,060,742	500
$\|b\| < 30, l$ 130	2,325,644	700	721,862	300
$\|b\| < 30, l$ 150	2,484,827	700	873,794	400
$\|b\| < 30, l$ 178	2,294,412	700	788,832	400
$\|b\| < 30, l$ 187	2,548,694	700	878,777	400
$\|b\| < 30, l$ 200	2,740,520	800	824,923	400
$\|b\| < 30, l$ 230	3,030,631	900	828,242	400
total	37,225,027	10,900	10,175,899	4,700
$\|b\| < 30$, other	8,478,425	3,100	2,513,240	1,600
$30 < b < 45$,	8,755,061	3,200	3,428,794	2,100
$45 < b, l < 180$	7,279,906	2,700	2,891,935	1,800
$45 < b, 180 < l$	5,802,229	2,100	2,208,236	1,400
$b < -30$	4,528,535	1,700	1,894,590	1,200
total	34,844,156	12,800	12,936,795	8,100
total	71,069,183	23,700	23,112,694	12,800

Table 2: SDSS AND SDSS-2MASS DATA FILES $1 < R_V < 8$

Definition	stars	size (mb)	stars	size (mb)
$\|b\| < 30, l$ 50	7,628,509	2,500	1,512,202	800
$\|b\| < 30, l$ 70	6,316,690	2,100	1,409,616	800
$\|b\| < 30, l$ 90	4,401,948	1,500	1,214,575	600
$\|b\| < 30, l$ 110	3,447,995	1,200	1,042,985	600
$\|b\| < 30, l$ 130	2,292,590	800	705,990	400
$\|b\| < 30, l$ 150	2,483,536	800	864,815	500
$\|b\| < 30, l$ 178	2,292,155	800	776,839	400
$\|b\| < 30, l$ 187	2,546,961	800	872,125	500
$\|b\| < 30, l$ 200	2,738,176	900	819,377	400
$\|b\| < 30, l$ 230	3,030,371	1,000	827,087	400
total	37,178,931	12,400	10,045,621	5,400

Note. —

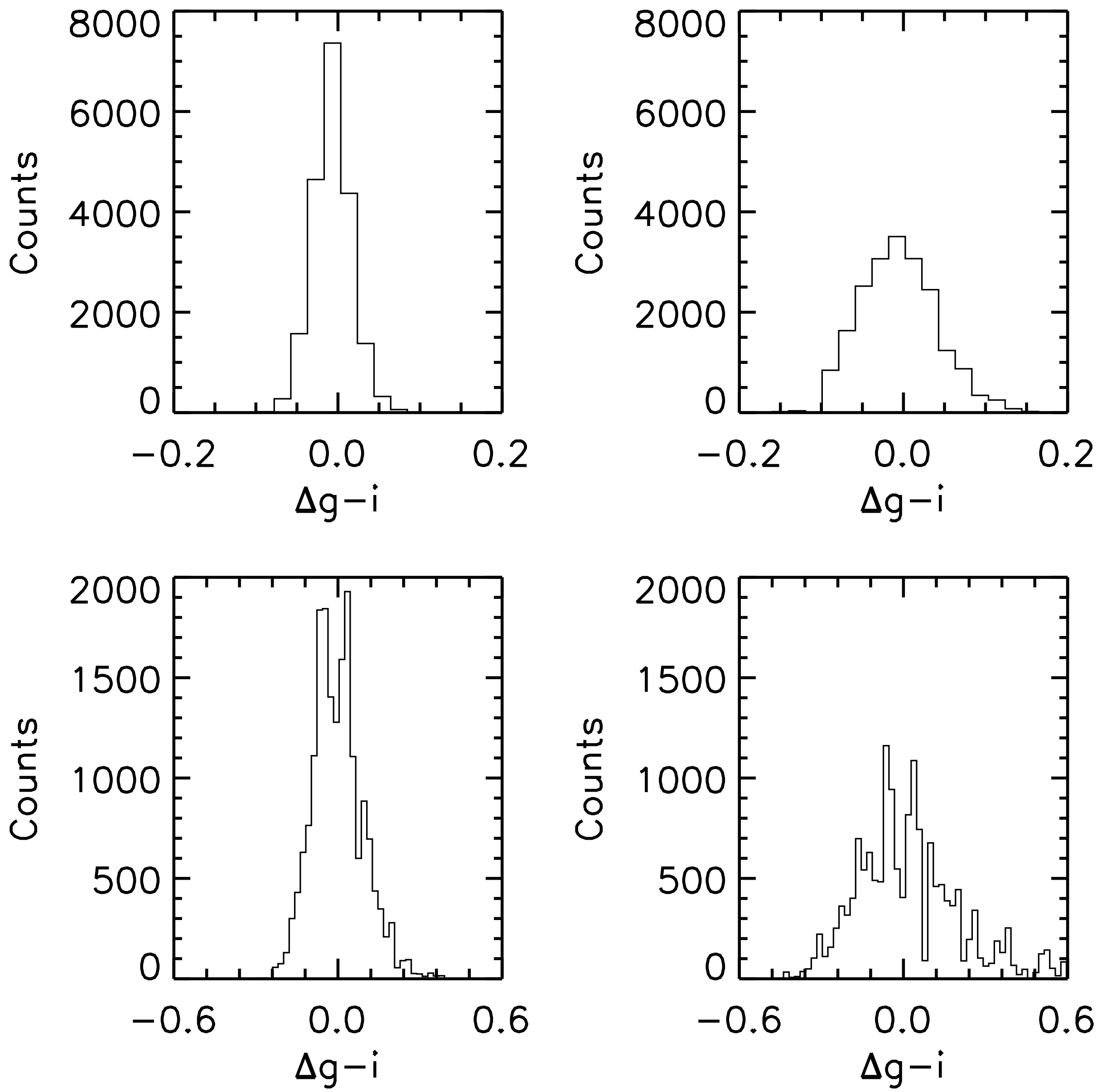


Fig. 3.— A Monte Carlo study of best-fit model errors as a function of photometric errors, for a star with g-i=1.95 and Ar=1.5. The photometric errors are generated from gaussian distributions with widths equal to 0.01 (top left), 0.02 (top right), 0.04 (bottom left) and 0.08 (bottom right). The widths of the shown distributions are xxxx, clockwise, from the top left panel.

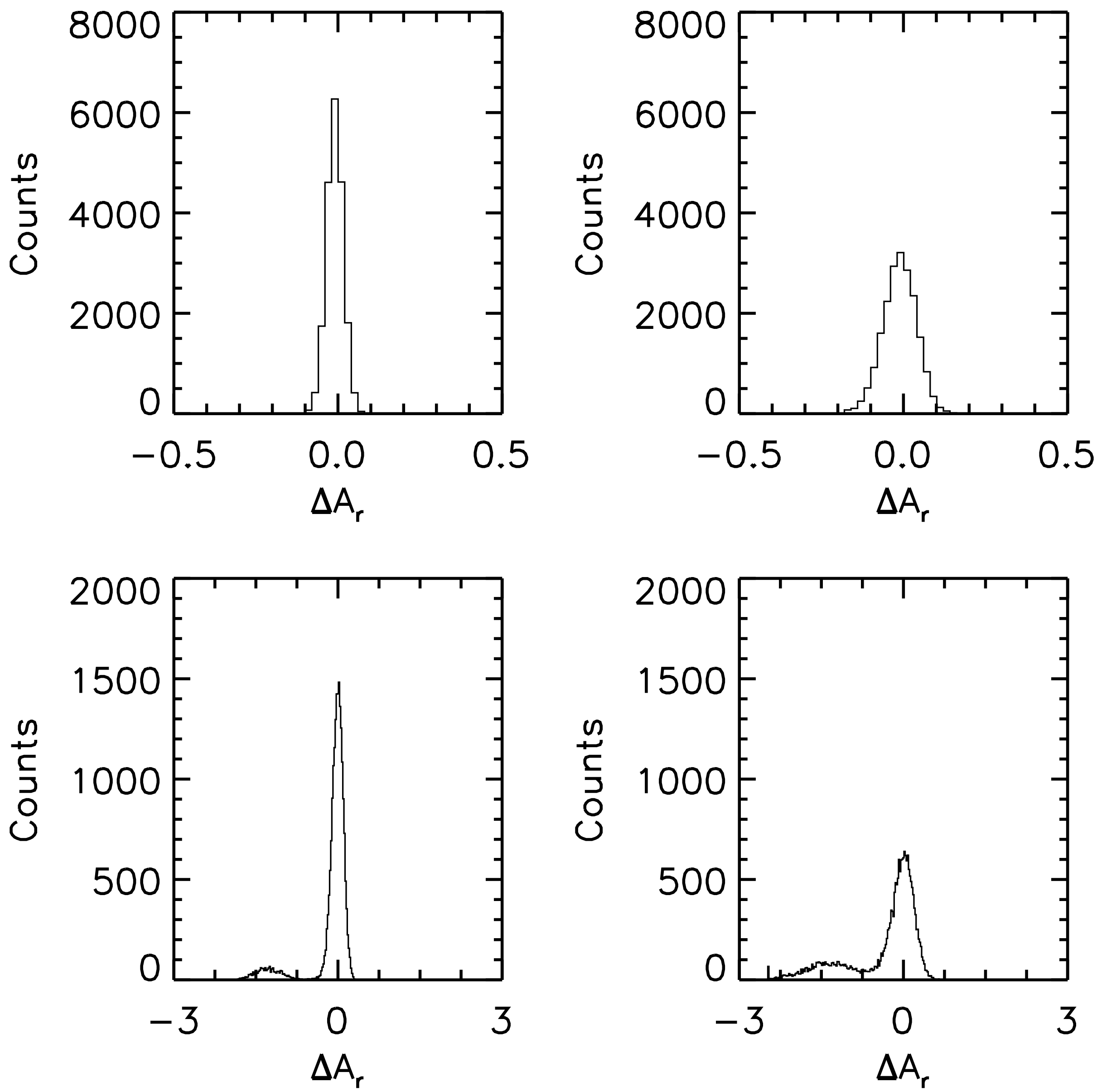


Fig. 4.— Analogous to Figure 2, except that the errors in best-fit Ar are shown. The widths of the shown distributions are xxxx, clockwise, from the top left panel.

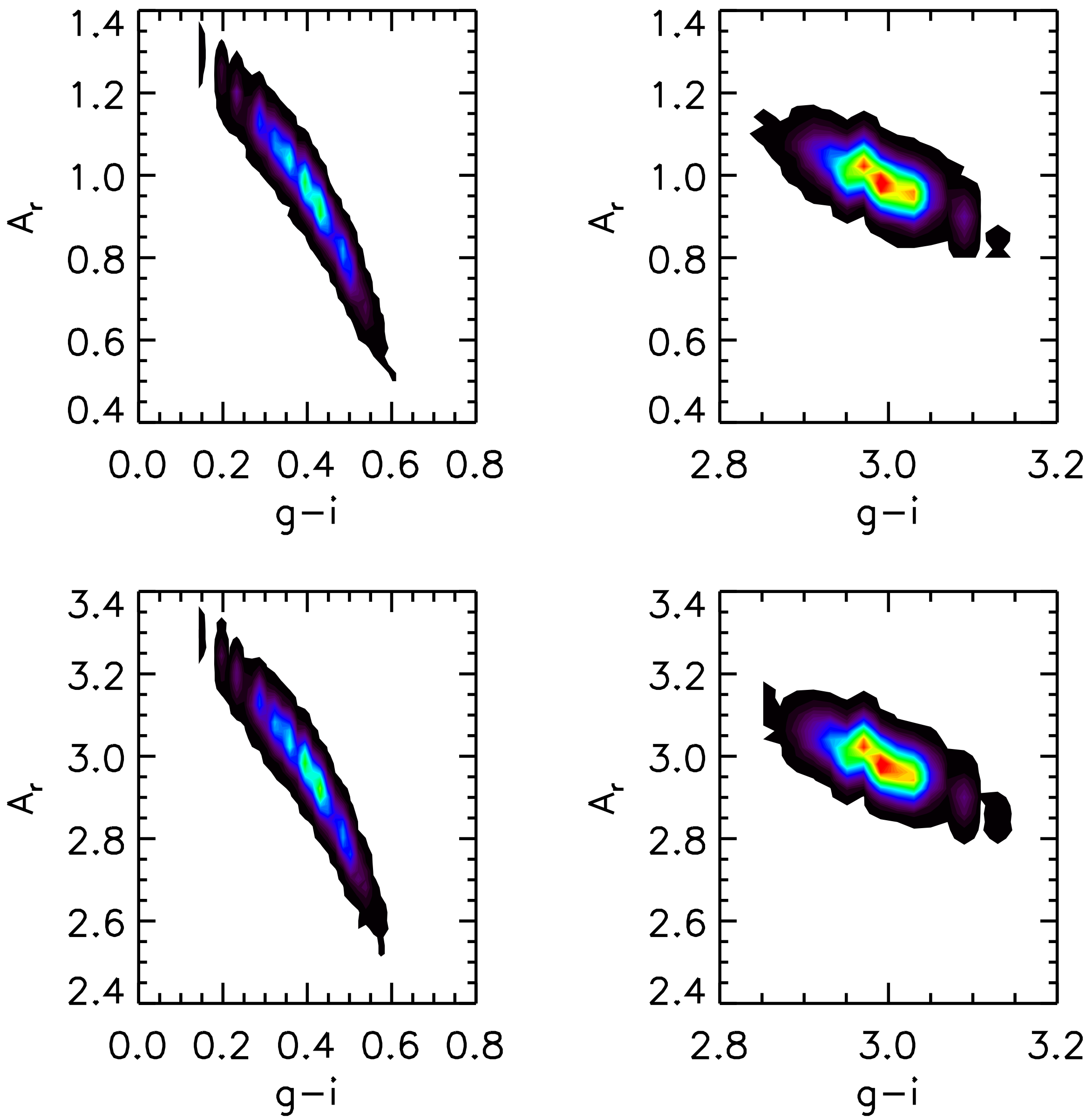


Fig. 5.— The distribution of best-fit values for A_r and $g-i$ for two different stars (left column: a blue star with true $g-i$=0.4; right column: a red star with true $g-i$=3.0) and two different extinction values (top panels: $A_r = 1$; bottom panels: A_r=3). Photometric errors in the $ugriz$ bands are generated using gaussian distributions with σ=0.02 mag (uncorrelated between different bands). Note that the A_r vs. $g-i$ covariance is larger for the blue star).

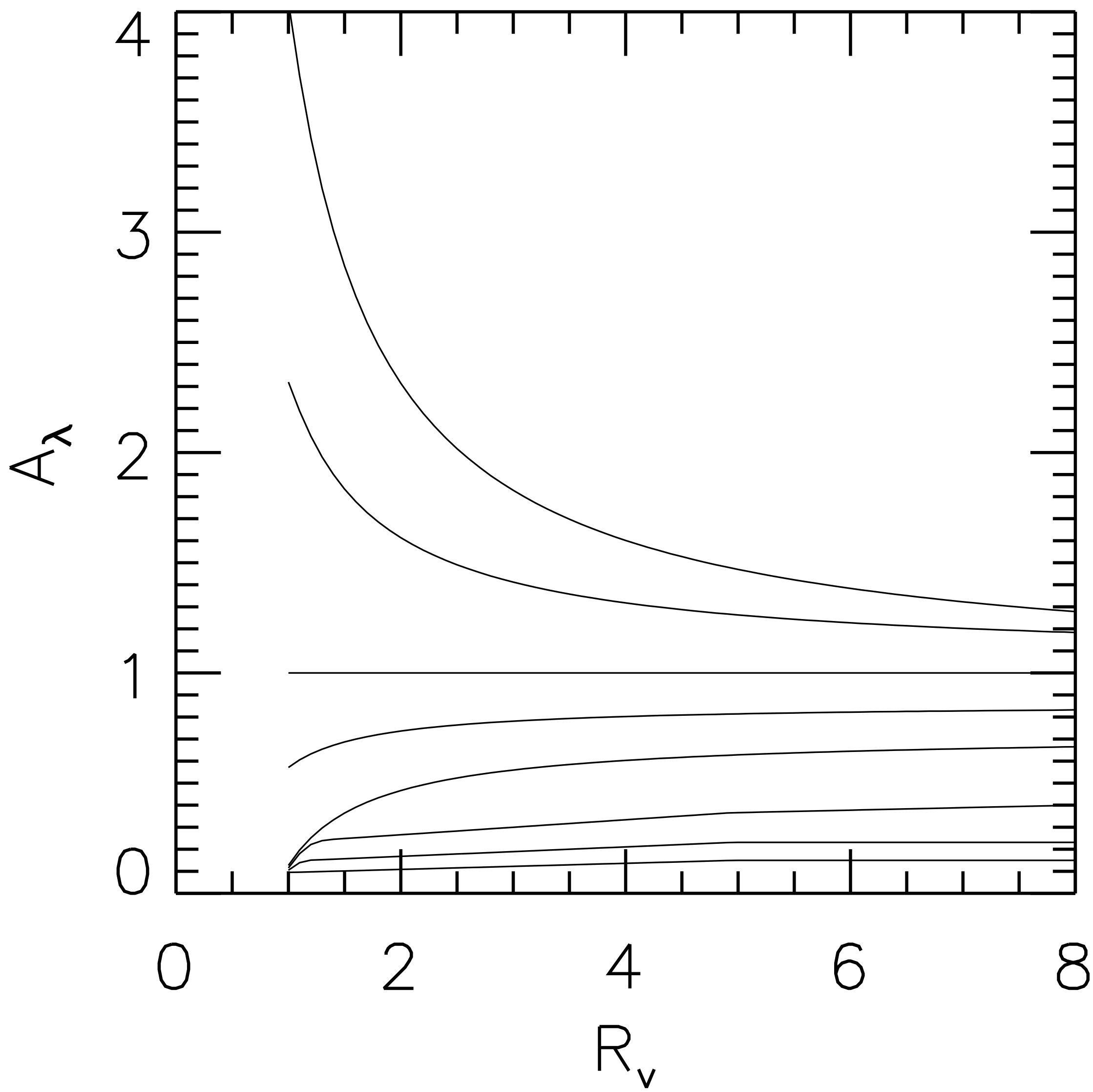


Fig. 6.— The A_λ/A_r ratio, computed by Doug eons ago, shown as a function of R_V, for $\lambda = (ugrizJHK)$, top to bottom.

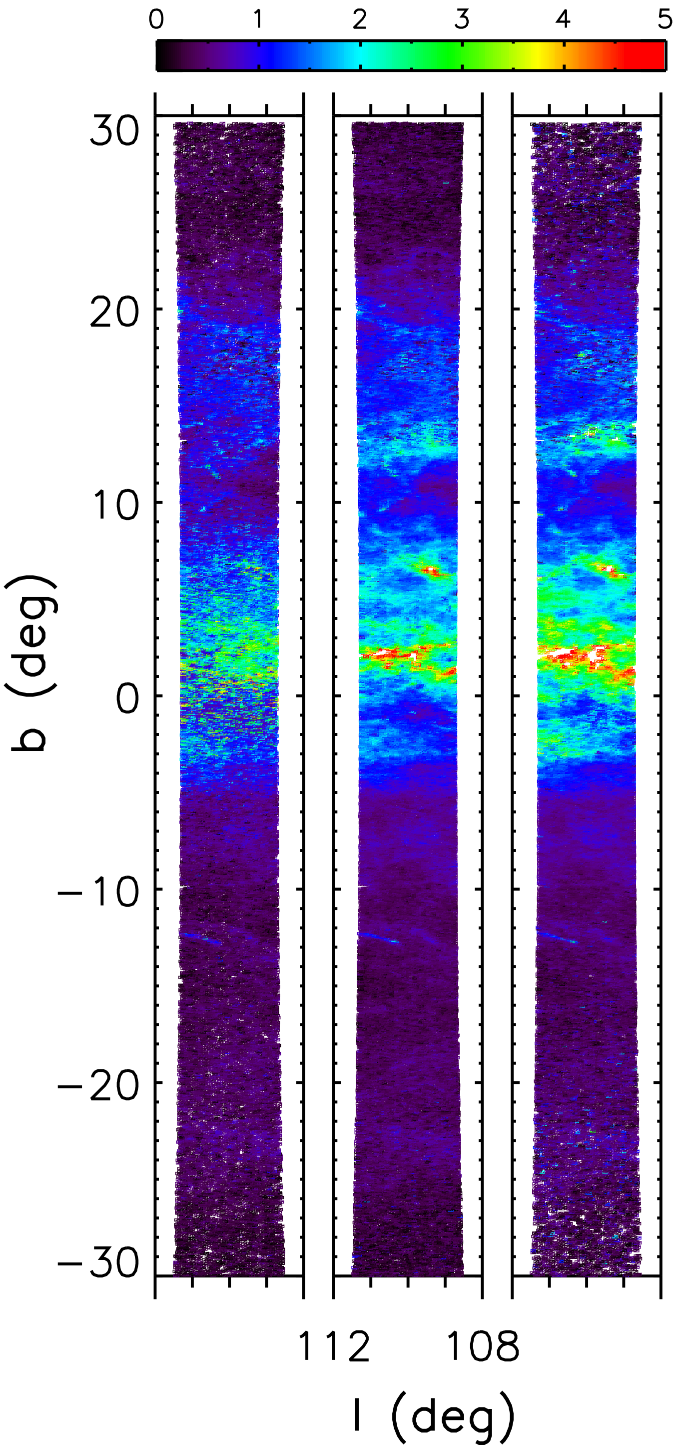


Fig. 7.— The best-fit A_r based on SDSS data and assuming $R_V = 3.1$ for stars selected from three distance ranges: 0.3-0.6 kpc (left), 1-1.5 kpc (middle) and 2-2.5 kpc (right). It is assumed that all stars are on main sequence when estimating distances. The legend on top shows the color-code, and each pixel shows the median A_r (the pixel size is 4x4 arcmin^2). Note that A_r increases with distance.

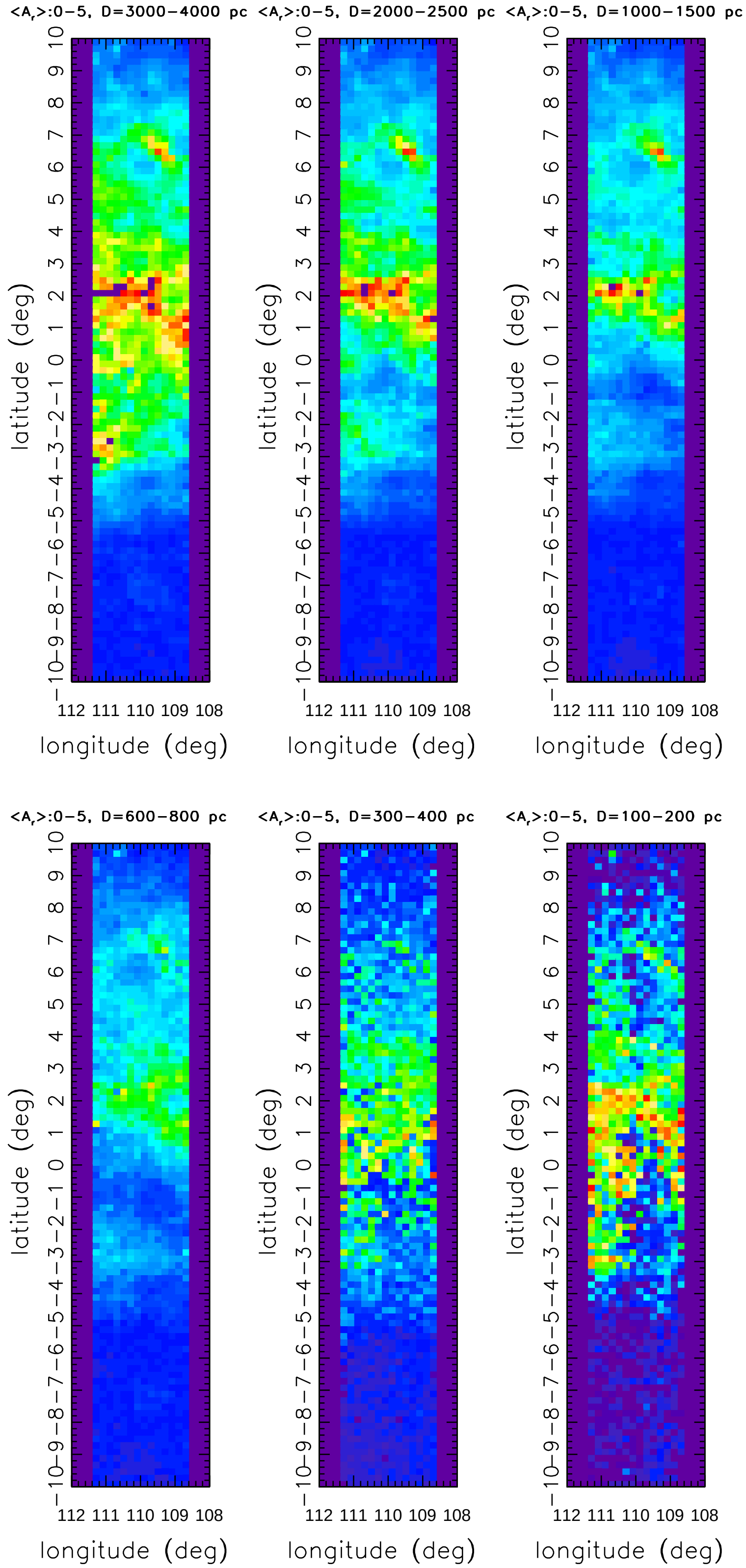


Fig. 8.— Similar to Figure 7, except with finer distance steps and made using slightly different visualization code. The high extinction values in the closest slice (bottom right panel) are most likely due to distant giants with underestimated distances.

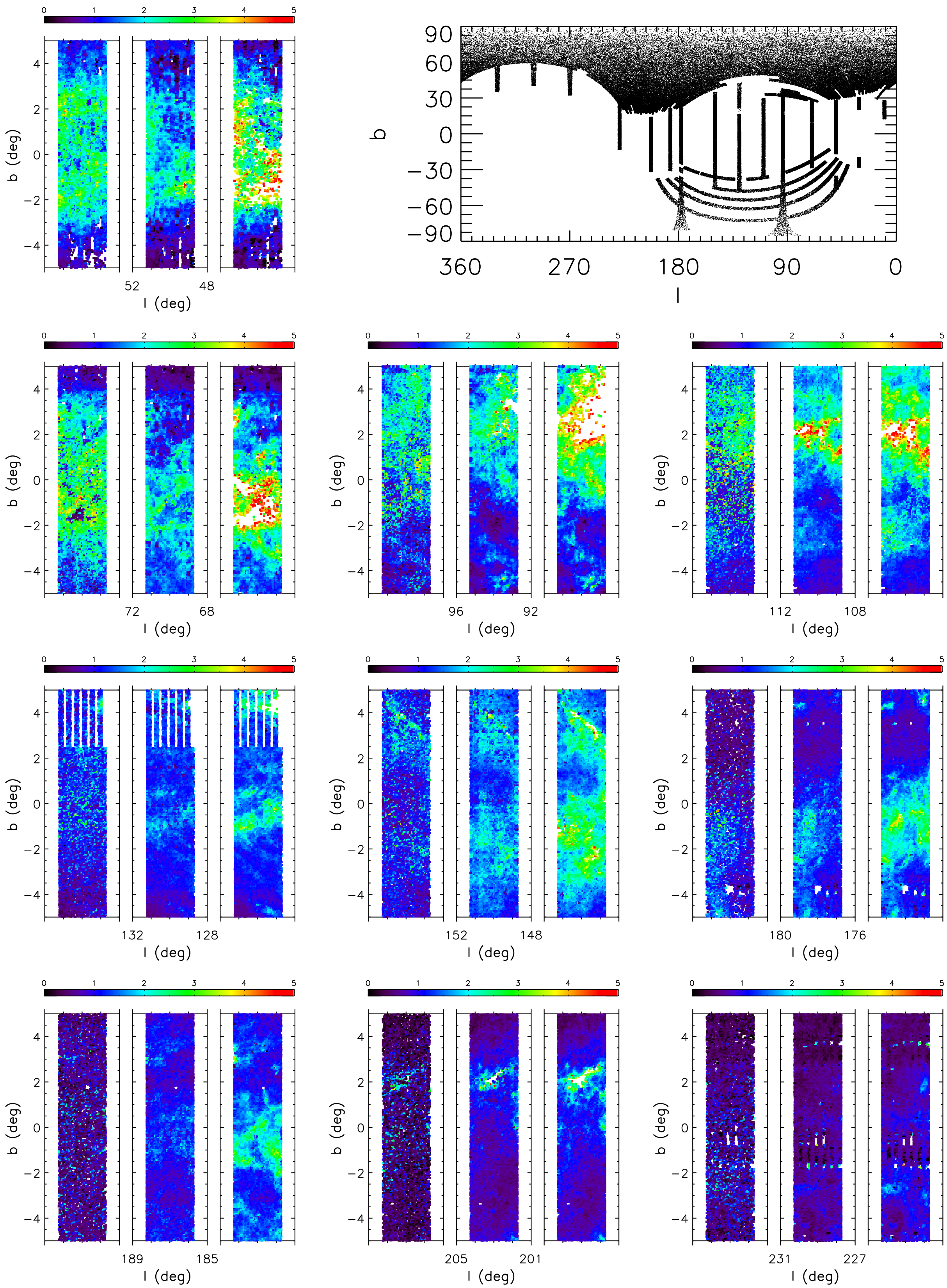


Fig. 9.— Similar to Figure 7, except that only the $|b| < 5°$ region is displayed, and all ten analyzed SEGUE stripes are shown. The top right shows the sky coverage of the analyzed data.

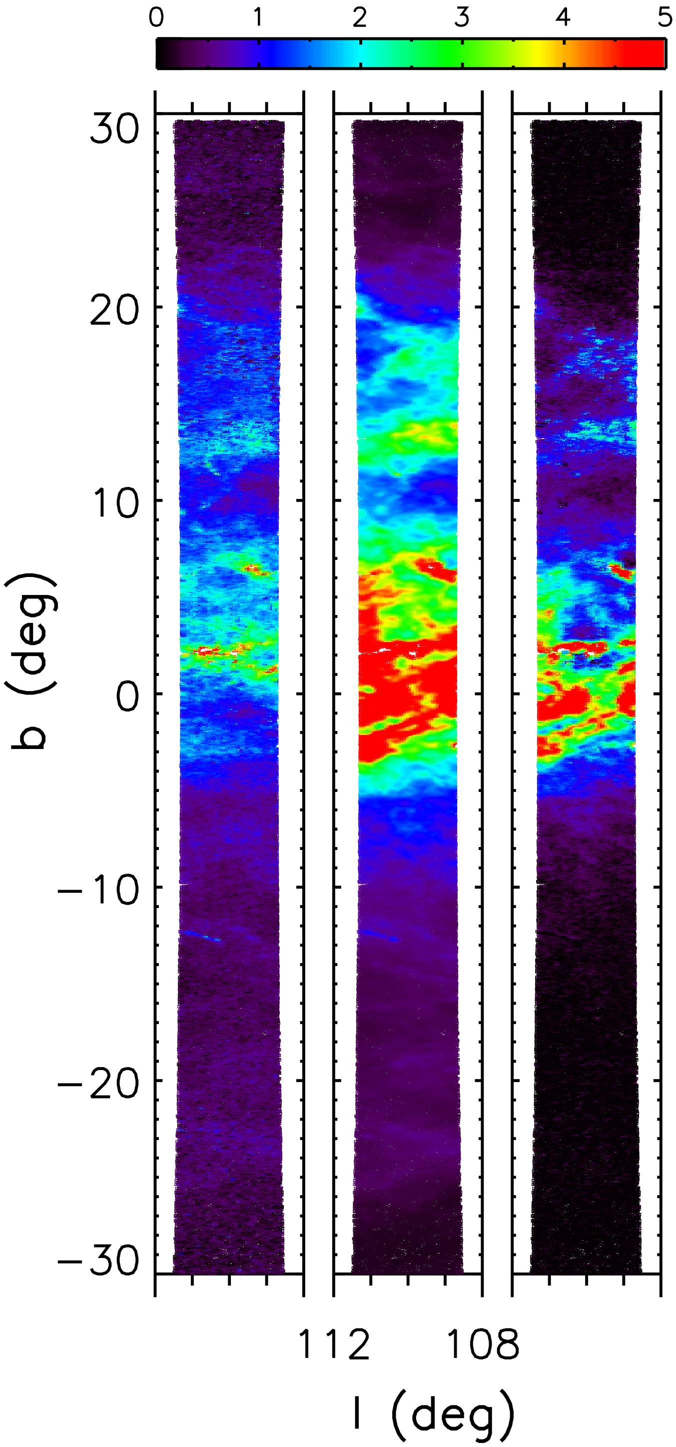


Fig. 10.— Analysis of the differences in A_r (right panel) between best-fit values (left panel) and the SFD values (middle panel). The legend on top shows the color-code, and each pixel shows the median A_r (or the median difference in the right panel) for stars with $\chi^2_{pdf} < 3$ and distance in the range 0.8-1.2 kpc (SDSS-based fits). If the SFD maps are correct, then the structure seen in the right panel must be more distant than ~1 kpc.

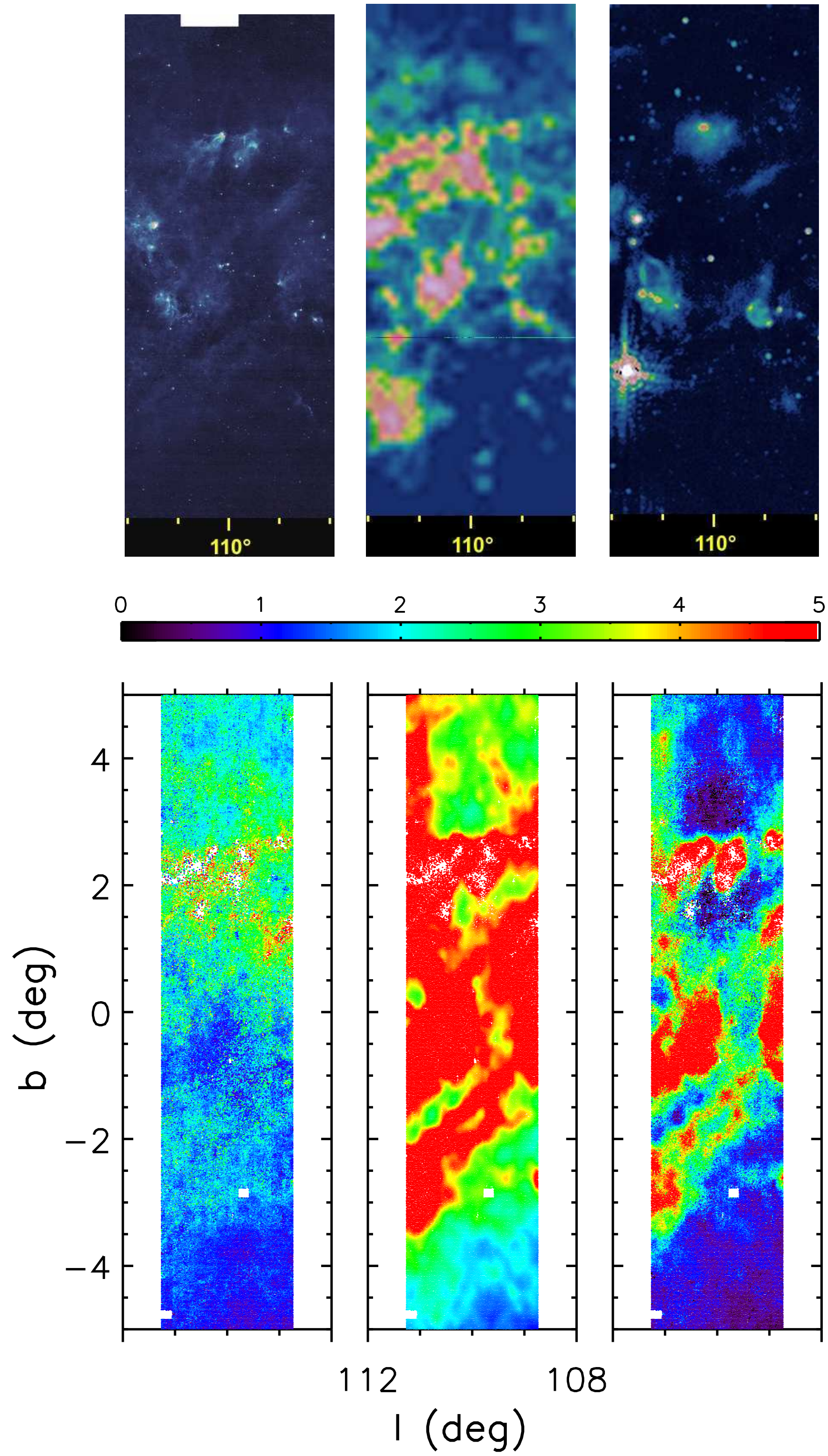


Fig. 11.— A comparison of the differences in A_r (right panel) between best-fit values and the SFD values (the lower three panels are the central parts of panels shown in Figure 10. The top three panels show mid-IR (left), CO (middle) and radio continuum (right) maps on approximately the same scale (add references).

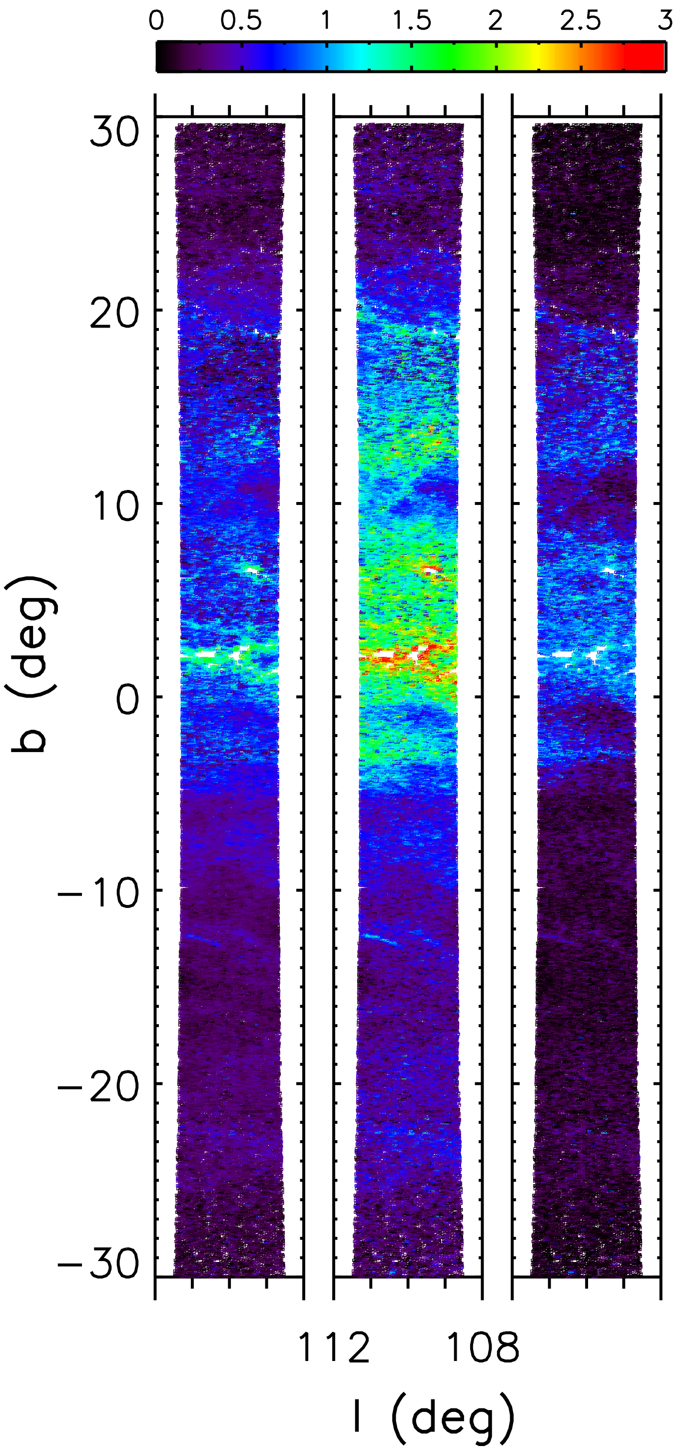


Fig. 12.— Analysis of the differences in A_r (right panel) between fits assuming $R_V = 2$ (left panel) and fits assuming $R_V = 4$ (middle panel). The legend on top shows the color-code, and each pixel shows the median A_r (or the median difference in the right panel) for stars with $\chi^2_{pdf} < 3$ (SDSS-based fits).

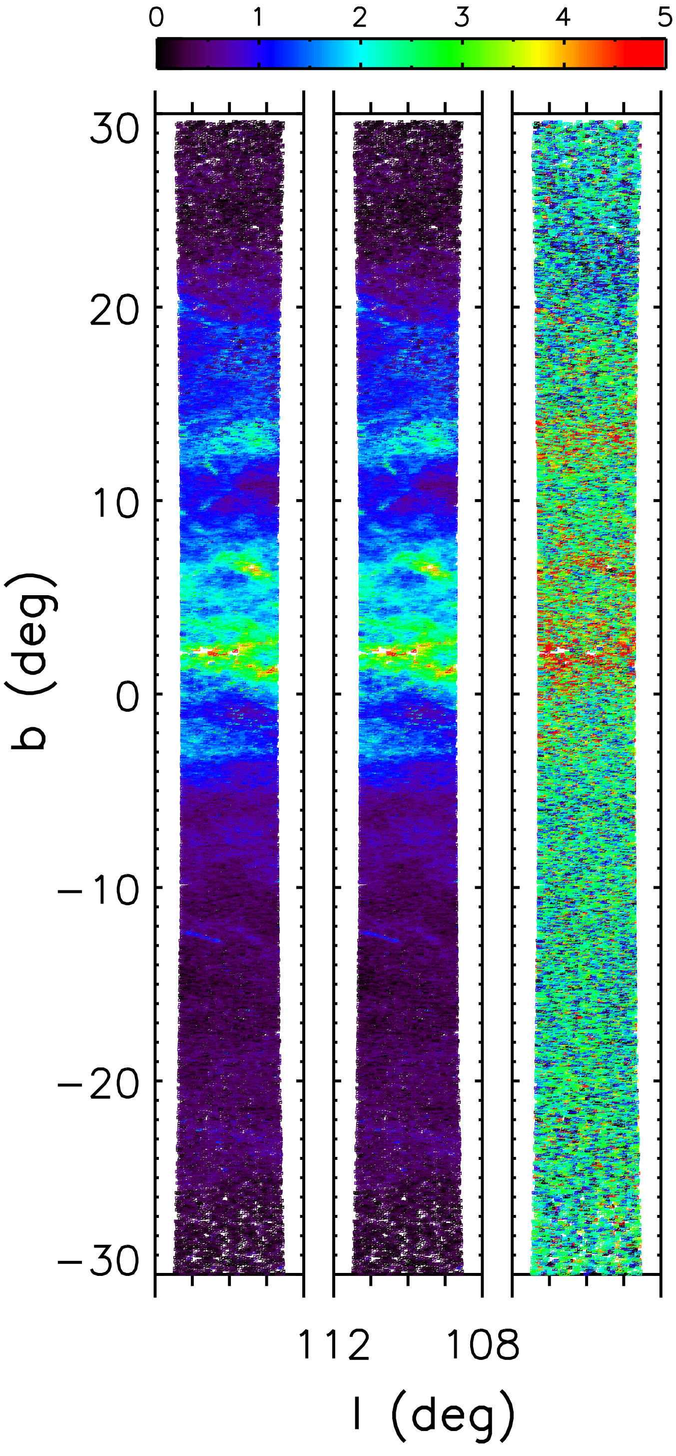


Fig. 13.— Analysis of the differences in A_r (right panel) between fits based only on SDSS data (left panel) and fits based on both SDSS and 2MASS data (middle panel). The legend on top shows the color-code for the left and middle panels, and each pixel shows the median A_r for stars with $\chi^2_{pdf} < 3$. The color code for the right panel uses the same palette, but the limits are ± 0.1 mag and each pixel shows the median difference in A_r between the two fits.

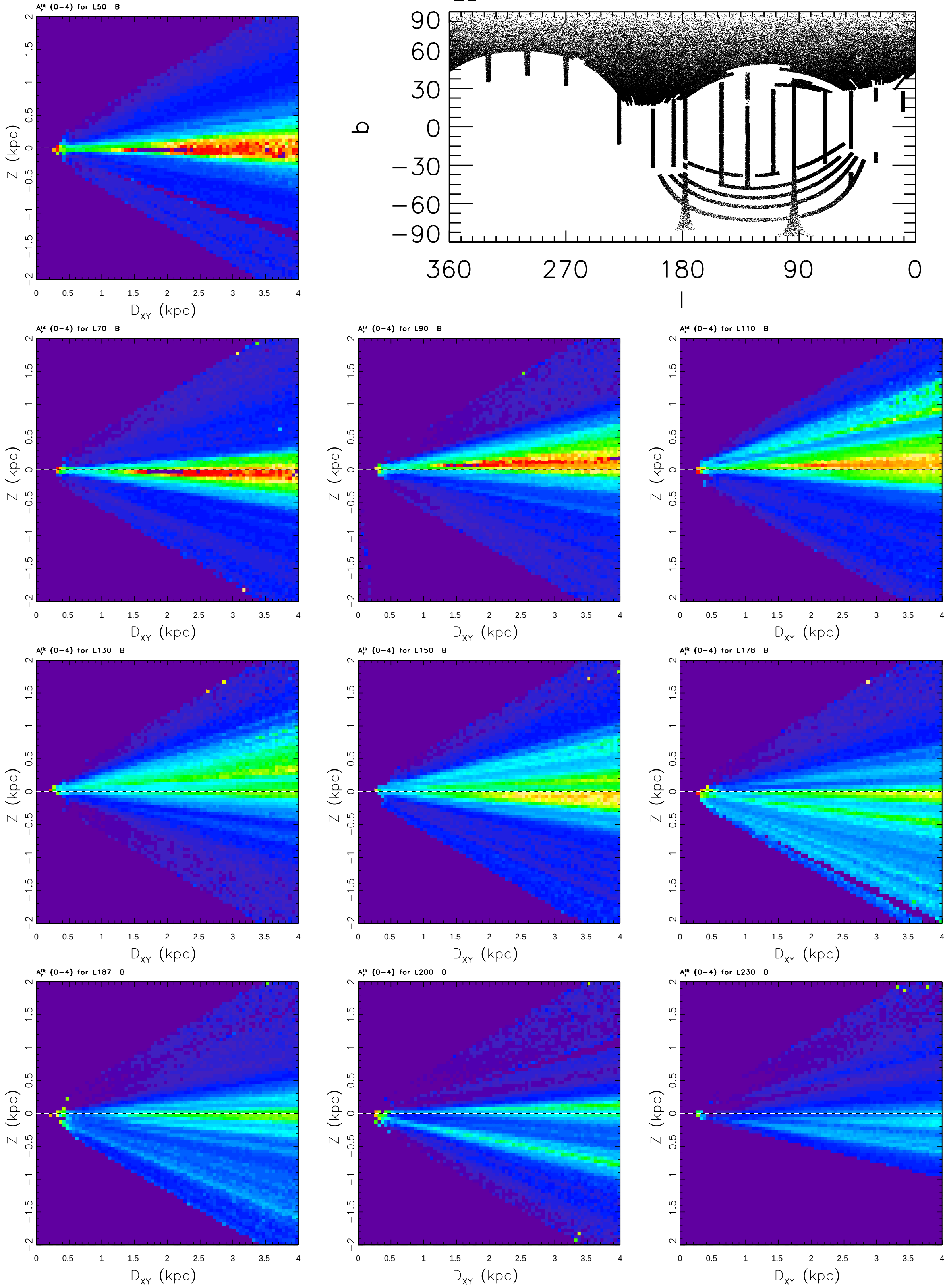


Fig. 14.— The best-fit A_r as a function of distance from the Galactic plane, Z, and distance along the plane, D_xy. Each pixel shows the median A_r between the observer and that point (i.e. this is **not** a cross-section of 3D dust map!). The top right shows the sky coverage of the analyzed data.

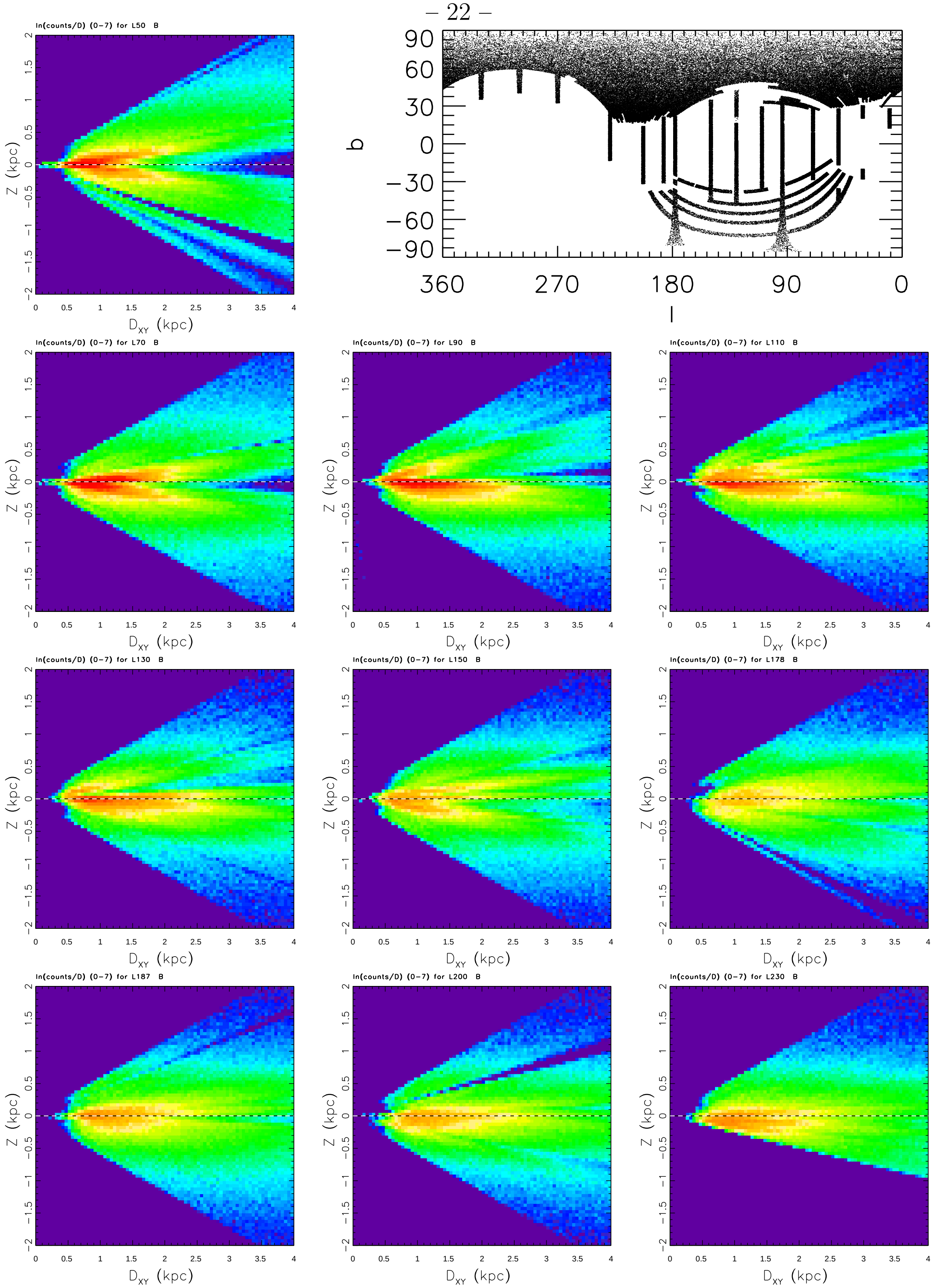


Fig. 15.— Similar to Figure 14, except that the local volume density is shown (arbitrarily normalized). The top right shows the sky coverage of the analyzed data.

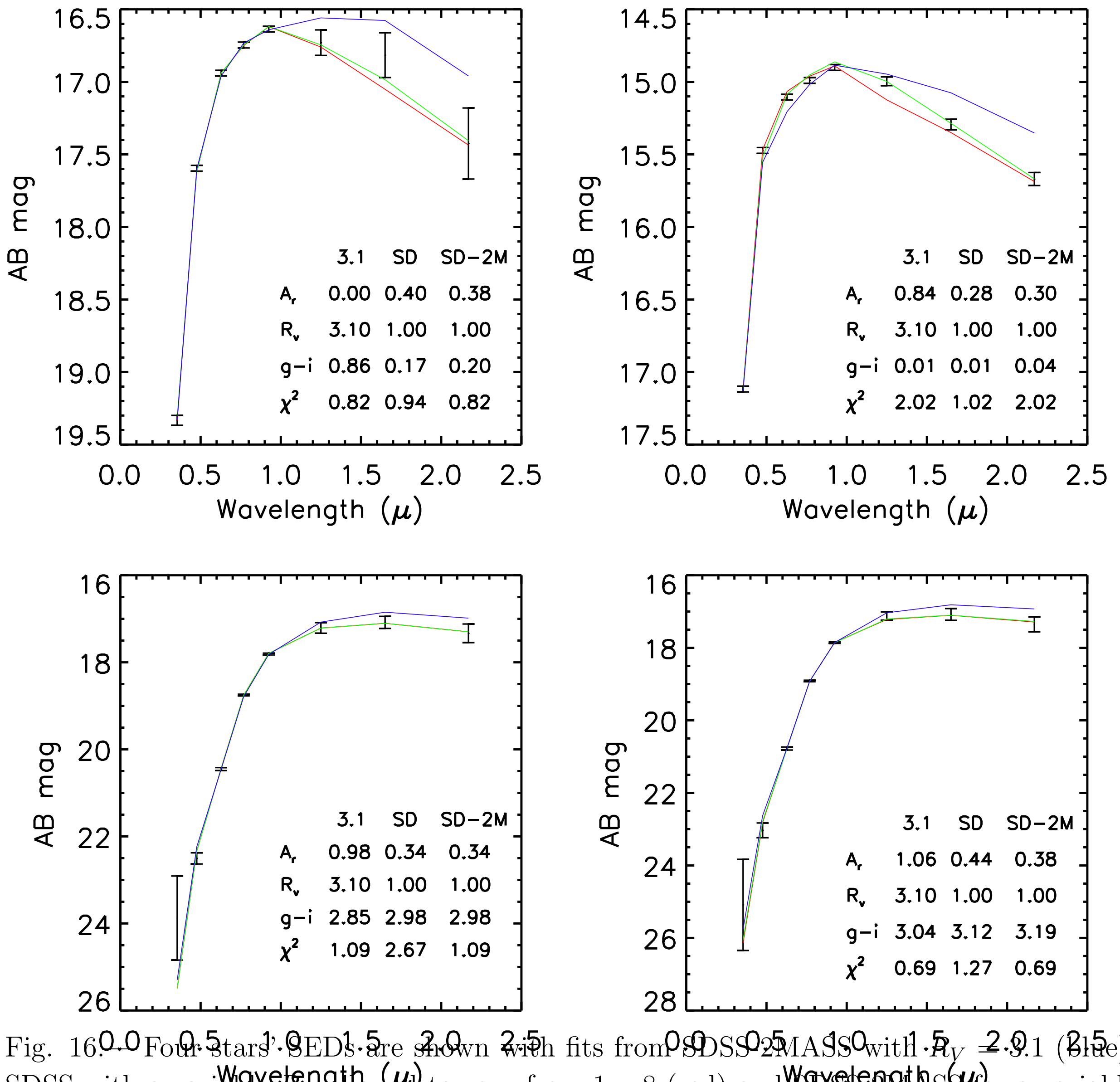


Fig. 16.— Four stars' SEDs are shown with fits from SDSS-2MASS with $R_V = 3.1$ (blue), SDSS with a variable R_V allowed to vary from $1 - 8$ (red) and SDSS-2MASS for a variable R_V (green).

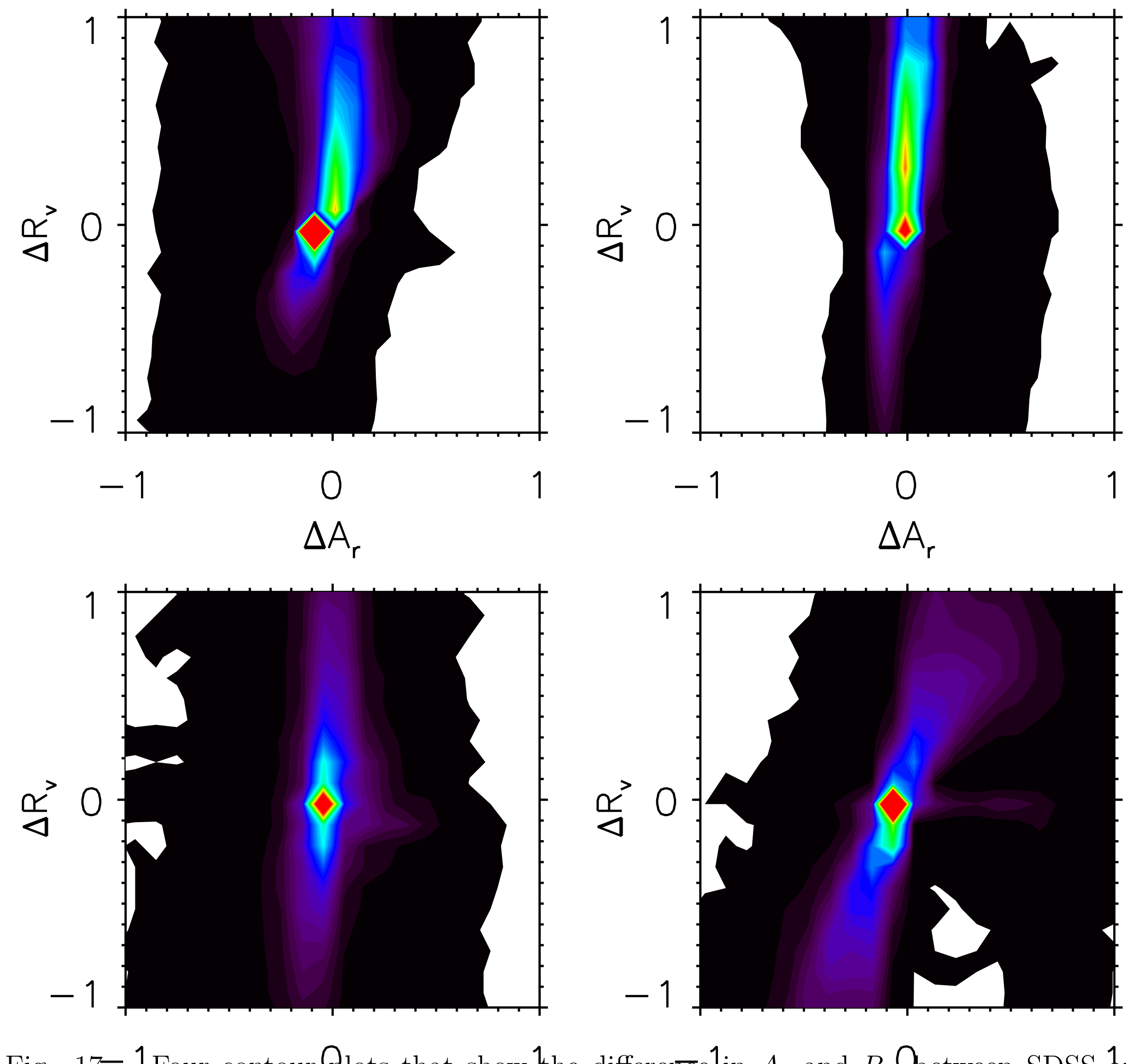
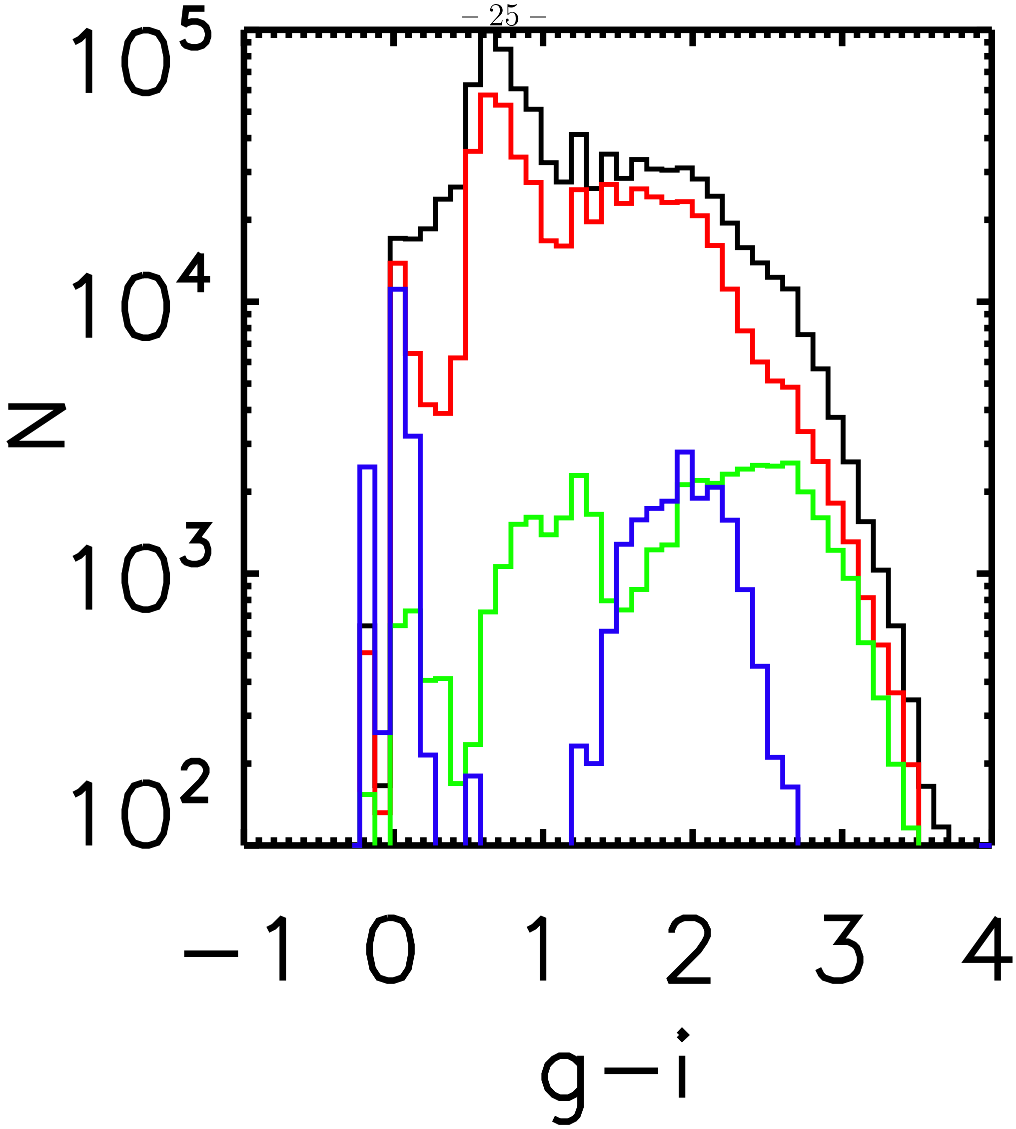
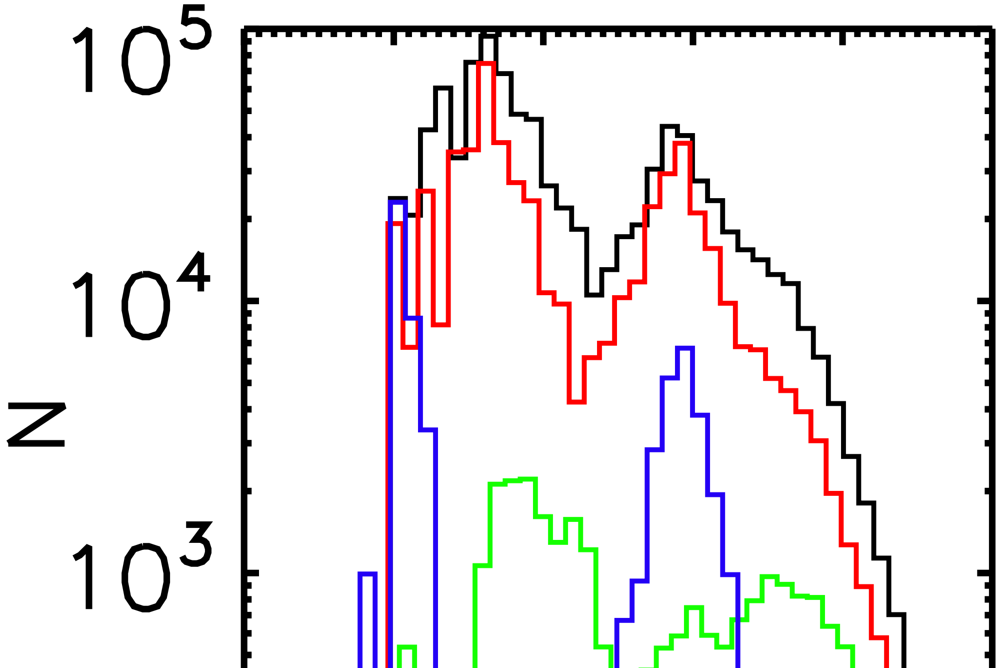


Fig. 17.— Four contour plots that show the difference in A_r and R_V between SDSS and SDSS-2MASS fits when R_V is allowed to vary ($\Delta A_r = A_r$(SDSS)-A_r(SDSS-2MASS)). The plots are organized by increasing $g-i$ quartiles from lowest to highest going clockwise from the upper left (g-i<0.549, 0.549<g-i<0.85, 0.85<g-i<1.75, 1.75<g-i).


10^5
10^4
10^3
10^2
N
−1
0
1
2
3
4
g−i


10^5
10^4
10^3
N

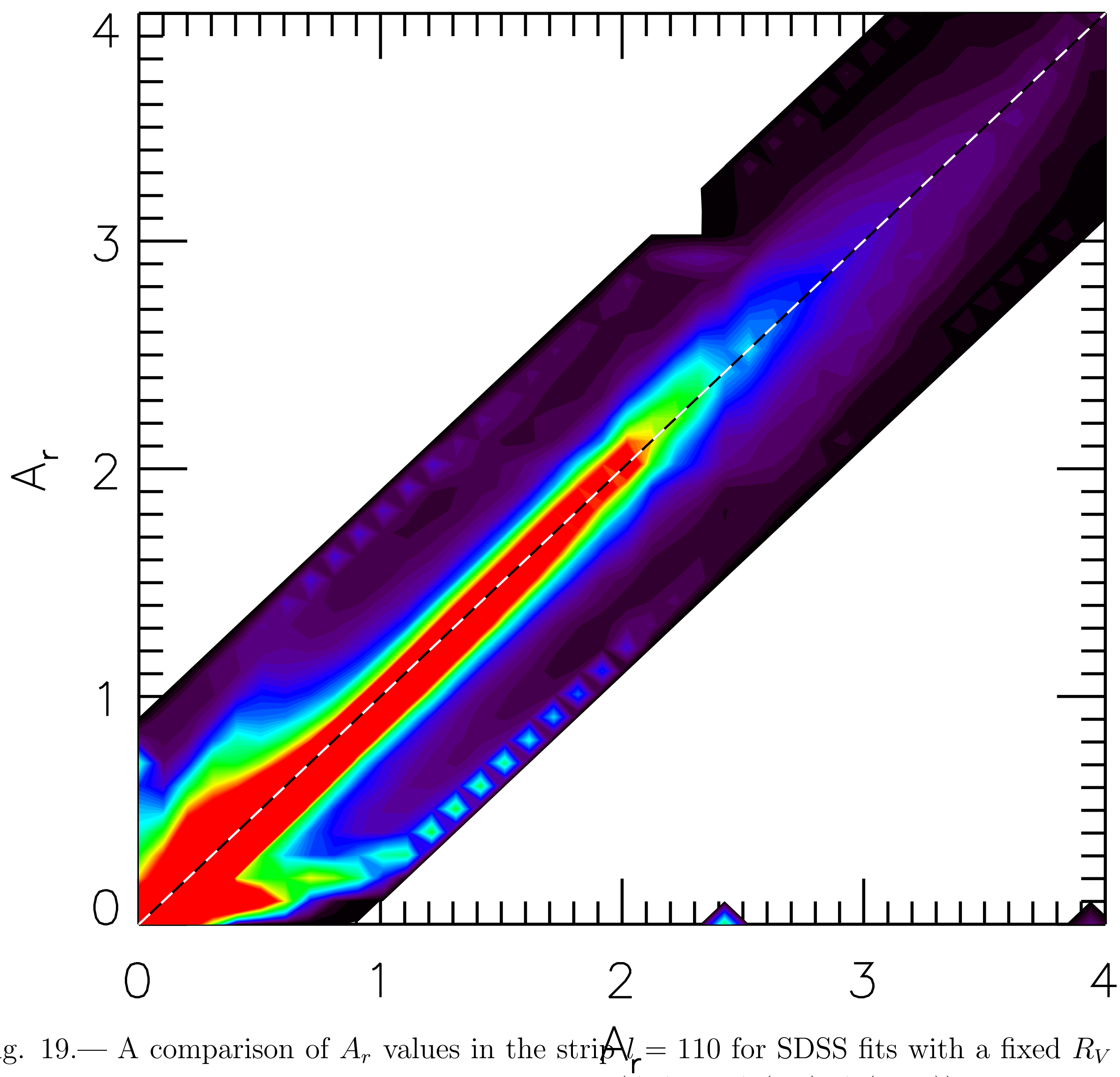


Fig. 19.— A comparison of A_r values in the strip $l = 110$ for SDSS fits with a fixed R_V of 3.1 on the x-axis and a variable R_V on the y-axis ($\Delta A_r = A_r(3.1)$-$A_r(fR_V)$).

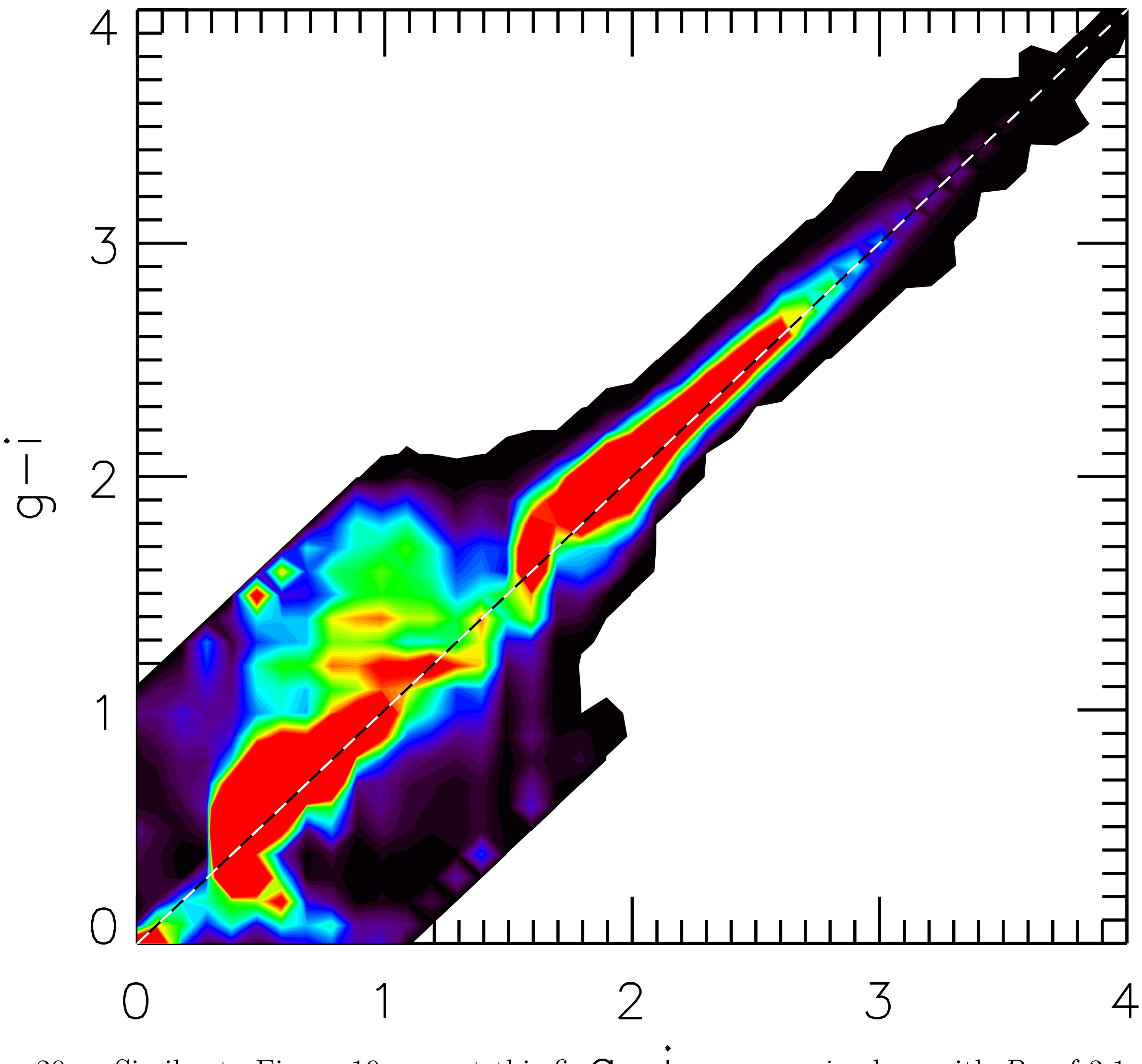


Fig. 20.— Similar to Figure 19, except this figure compares $g-i$ values with R_V of 3.1 on the x-axis and a variable R_V on the y-axis.

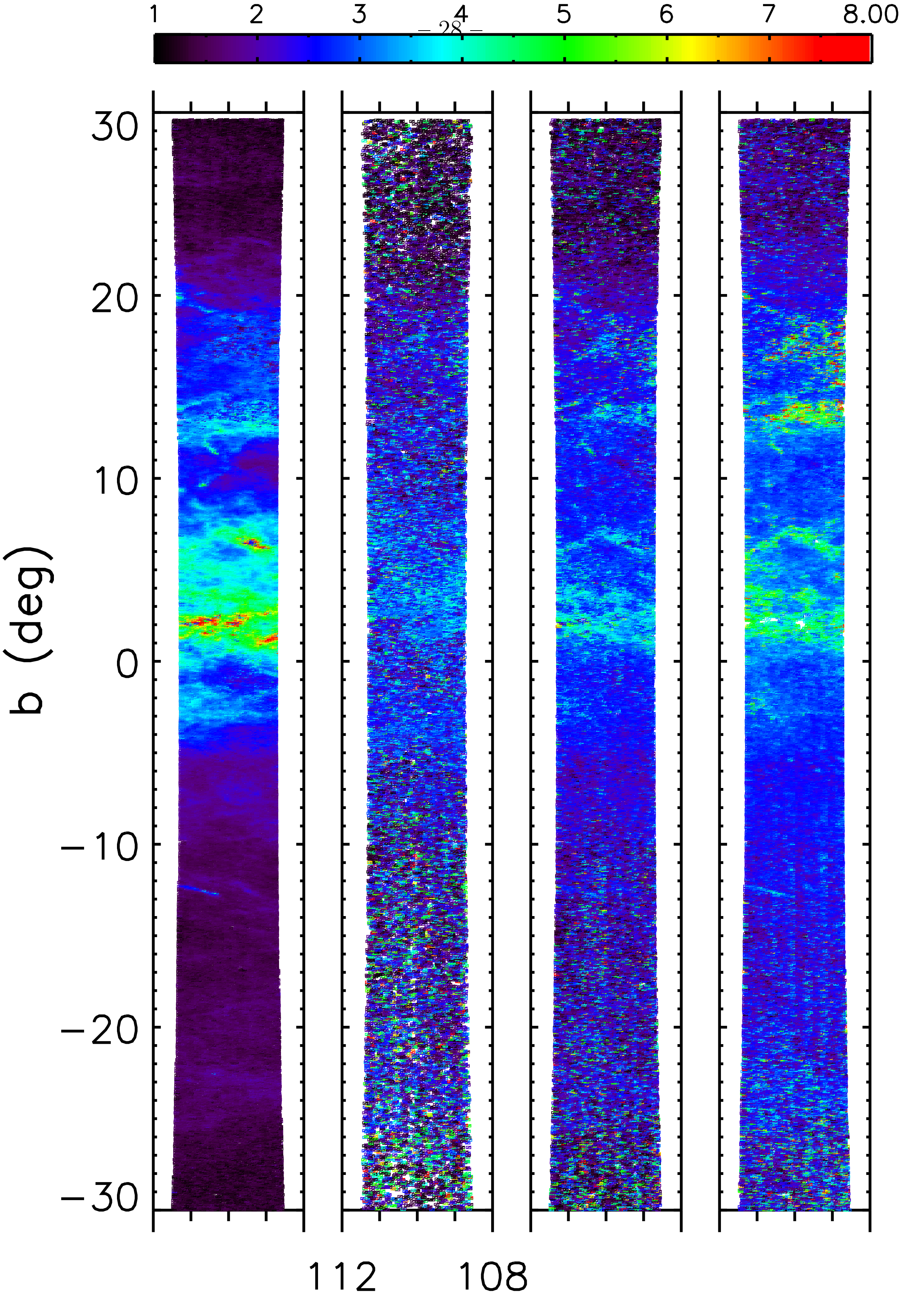


Fig. 21.— The first panel displays the median A_r values on a scale of 1-5 for stars with χ^2 <3 using SDSS-2MASS data for the SEGUE strip l=110. The next three panels display the median R_V values in distance cuts of 0.1-0.5kpc (left), 0.5-1kpc (middle), and 1<kpc (right). The legend on top shows the color-code, and each pixel shows the median A_r (the pixel size is 4x4 arcmin^2).

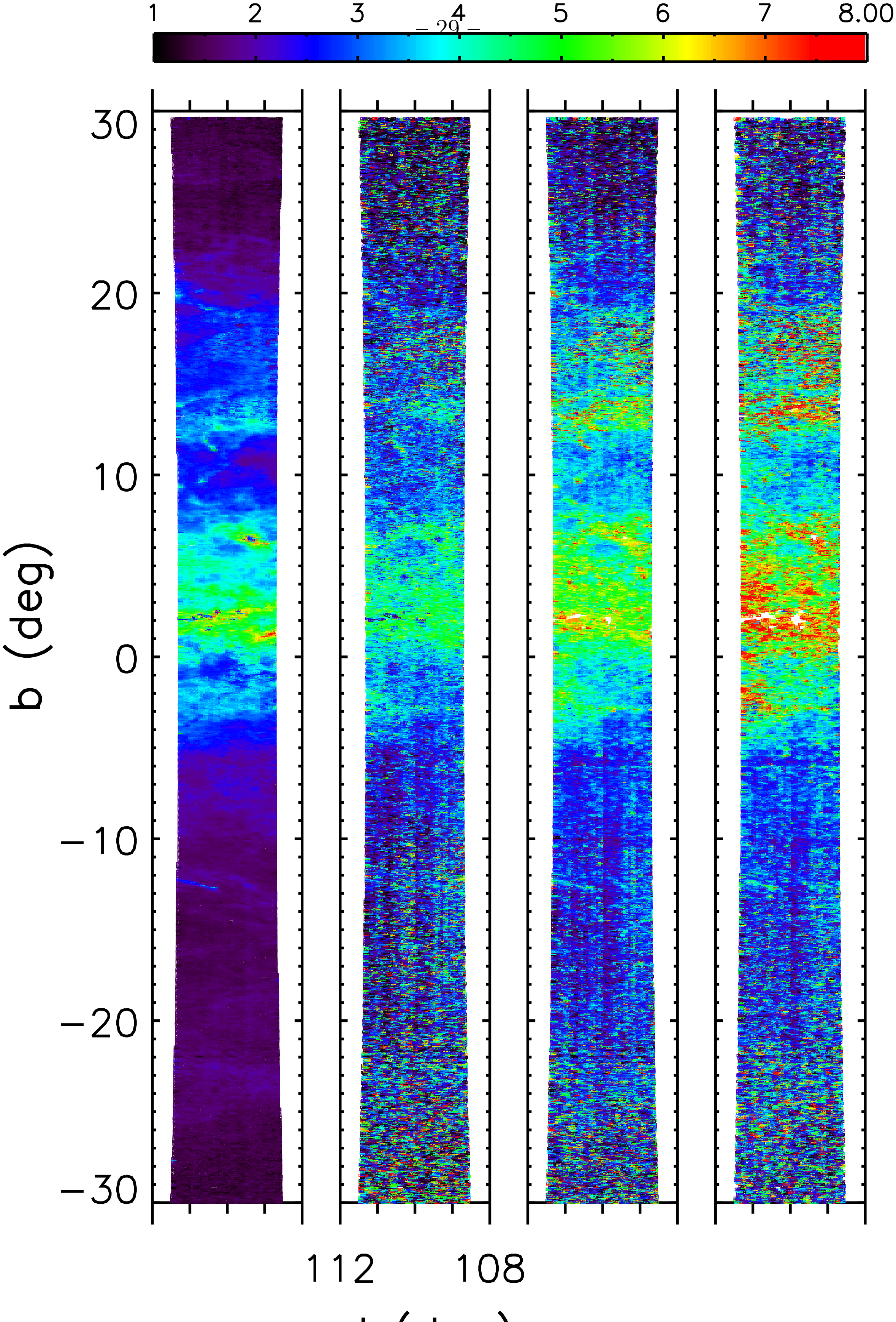


Fig. 22.— Similar to Figure 21, except this map shows fits using only SDSS bands (median A_r on the far left) and has distance cuts of 0.5-1kpc (left), 1-2kpc (middle), and 2-3kpc (right).