The Milky Way Tomography with SDSS: V. Dissecting Dust

Version from July 9, 2010

Michael Berry[1,2], Željko Ivezić[1], Branimir Sesar[1], Mario Jurić[3], Eddie Schlafly[3], Jillian Bellovary[1], Douglas Finkbeiner[3], Keira J. Brooks[1], Kevin R. Covey[3], Rob Gibson[1], Lynne Jones[1], Andrew J. Connolly[1], Adam Kowalski[1], Sarah Loebman[1], Rok Roškar[1], Amy Kimball[1], Nicholas A. Bond[2], David Schlegel[4], Jeffrey A. Munn[5], Timothy C. Beers[6], Julianne Dalcanton[1], Constance M. Rockosi[7], Brian Yanny[8], Heidi J. Newberg[9], Steven R. Majewski[10], Robert H. Lupton[11], Gillian R. Knapp[11], James E. Gunn[11], J. Allyn Smith[12], Mamoru Doi,[13], Masayuki Tanaka[14], Masataka Fukugita[15], Steve Kent[8], Tom R. Quinn[1], Suzanne Hawley[1], Scott Anderson[1], John Barentine[16], Howard Brewington[16], Mike Harvanek[16], Scott Kleinman[16], Jurek Krzesinski[16], Dan Long[16], Atsuko Nitta[16], Stephanie Snedden[16], Donald P. Schneider[17]

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

[2]Physics and Astronomy Department, Rutgers University Piscataway, NJ 08854-8019, U.S.A.

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

[6]Dept. of Physics & Astronomy and JINA: Joint Institute for Nuclear Astrophysics, Michigan State University, East Lansing, MI 48824, USA

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

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

[9]Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180

[10]Department of Astronomy, University of Virginia, P.O. Box 400325, Charlottesville, VA 22904-4325

[11]Princeton University Observatory, Princeton, NJ 08544

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

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

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

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

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

[17]Department of Astronomy and Astrophysics, The Pennsylvania State University, Uni-

Received ______________; accepted ______________

versity Park, PA 16802

ABSTRACT

We use SDSS photometry for 73 million stars to simultaneously obtain best-fit main-sequence stellar energy distribution and the amount of dust extinction along the line of sight towards each star. Using a subsample of 23 million stars with 2MASS photometry, whose addition enables more robust results, we show that SDSS photometry alone is sufficient to break degeneracies between intrinsic stellar color, dust amount, and dust properties. These fits enable detailed studies of the dust properties and its spatial distribution, and of the stellar spatial distribution at low galactic latitudes ($|b| < 30°$). Our results are in good agreement with the extinction normalization given by the Schlegel, Finkbeiner & Davis (1998, SFD) dust maps at high northern galactic latitudes, but indicate that the SFD extinction map appears to be consistently overestimated by about 20% in the southern sky. The constraints on the shape of the dust extinction curve across the SDSS and 2MASS bandpasses.disfavor the reddening law of O'Donnell (1994), but support the models by Fitzpatrick (1999) and Cardelli et al. (1989). For the latter, we find a ratio of the total to selective absorption to be $R_V = 3.01 \pm 0.05$(random)± 0.1 (systematic) over most of the high-latitude sky. At low galactic latitudes ($|b| < 5°$), we demonstrate that the SFD map cannot be reliably used to correct for extinction because most stars are embedded in dust, rather than behind it, as is the case at high galactic latitudes. We present evidence that sometimes the SFD map grossly overestimates the dust extinction at low galactic latitudes even when these distance effects are accounted for. In cases where such discrepancies are large, they seem correlated with the distribution of molecular gas. We analyze three-dimensional maps of the best-fit R_V and show that it can reach values as high as 5-6 in some low-latitude regions with large amounts of dust. We make these best-fit parameters, as well as all the input

SDSS and 2MASS data, publicly available.

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

1. INTRODUCTION

From our vantage point inside the disk of the Milky Way, we have a unique opportunity to study an $\sim L^*$ spiral galaxy in great detail. By measuring and analyzing the properties of large numbers of individual stars, we can map the Milky Way in a nine-dimensional space spanned by the three spatial coordinates, three velocity components, and the three main stellar parameters – luminosity, effective temperature, and metallicity. In a series of related studies, we used data obtained by the Sloan Digital Sky Survey (York et al. 2000) to study in detail the distribution of tens of millions of stars in this multi-dimensional space. In Jurić et al. (2008, hereafter J08) we examined the spatial distribution of stars in the Galaxy, in Ivezić et al. (2008b, hereafter I08) we extended our analysis to include the metallicity distribution, and in Bond et al. (2009, hereafter B10) we investigated the distribution of stellar velocities. In Jurić et al. (2010, in prep) we estimate stellar luminosity functions for disk and halo, and describe an empirical Galaxy model and corresponding publicly available modelling code that encapsulate these SDSS-based results.

All of the above studies were 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, see Yanny et al. 2009). At least in principle, these data can be used to extend the above analysis much closer to the Galaxy 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 interstellar dust extinction corrections. High-latitude SDSS data are typically corrected for interstellar extinction using maps from Schlegel, Finkbeiner & Davis (1998; hereafter SFD). When the full SFD extinction correction is applied to low-latitude data, 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 predominantly due to the fact that stars are embedded in the dust layer, rather than behind it (the latter is an excellent approximation for most stars at high galactic latitudes). This conclusion is also supported by other Galaxy models, such as Besançon (Robin et al. 2003) and TRILEGAL (**?**). Therefore, in order to fully exploit SEGUE data, both the intrinsic colors of a given star and the amount of dust extinction along the line of sight to the star have to be known. Distances to stars, which can be derived using appropriate photometric parallax relations (see I08), would then enable mapping of the stellar spatial distribution, and the ISM dust distribution and dust extinction properties are interesting in their own right (**?**, and references therein). Additional motivation for quantifying stellar and dust distribution close to the plane is to inform the planning of the LSST survey, which is considering deep multi-band coverage of the Galactic plane[1] (Ivezić et al. 2008).

The amount of dust can be constrained by measuring dust extinction and/or reddening, typically at UV, optical and near-IR wavelengths, or by measuring 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 an angular resolution of ∼6 arcmin (the temperature correction applied to IRAS 100 μm data is based on DIRBE 100 μm and 240 μm data, and has a lower angular resolution of ∼1 degree. See SFD for more details). It has been found that SFD map overestimates the dust column by 20-30% when the dust extinction in the SDSS r band, $A_r \sim 0.87 A_V$, exceeds 0.5 mag (e.g., Arce & Goodman 1999). Such an error may be due to confusion of the background emission by emission from point sources. A generic shortcoming of the far-IR emission-based methods is that they cannot provide constraints on the three-dimensional distribution of dust; instead, only the total amount of dust along the line of sight to infinity is measured. In addition, they

[1]See also the LSST Science Book available from www.lsst.org.

provide no constraints for the wavelength dependence of extinction at UV, optical and near-IR wavelengths.

In this work, we estimate dust reddening for each detected star by simultaneusly fitting its observed optical/IR spectral energy distribution (SED) using an empirical library of intrinsic reddening-free SEDs, and a reddening curve described by the standard parameter $R_V = A_V/E(B-V)$, and the dust extinction along the line of sight in the SDSS r band, A_r. This method can be used to estimate the three-dimensional spatial distributions of both stars and dust. The dataset and methodology, including simulation-based tests of the adopted algorithm are described in §2. A preliminary analysis of results is presented in §3, and the results are summarized and discussed in §4.

.

2. DATA AND METHODOLOGY

All datasets used in this work are defined using SDSS imaging data for unresolved sources. Objects that are positionally associated with 2MASS sources are a subset of the full SDSS sample. We first briefly describe the SDSS and 2MASS surveys, and then describe the selection criteria and the fitting algorithm.

2.1. SDSS Survey

The characteristics of the SDSS imaging and spectroscopic data relevant to this work (Fukugita et al. 1996; Gunn et al. 1998; Hogg et al. 2001; Smith et al. 2002; Stoughton et al. 2002; Pier et al. 2003; Ivezić et al. 2004; Tucker et al. 2006; Gunn et al. 2006; Abazajian et al. 2009; Yanny et al. 2009) are described in detail in the first two papers in the series

(J08, I08). Here we only reiterate that the survey photometric catalogs are 95% complete to a depth of $r \sim 22$, with photometry accurate to ∼0.02 mag (both absolute and rms error) for sources not limited by Poisson statistics. Sources with $r < 20.5$ have astrometric errors less than 0.1 arcsec per coordinate (rms; Pier et al. 2003), and robust star/galaxy separation is achieved above $r \sim 21.5$ (Lupton et al. 2001).

The latest public SDSS Data Release 7 contains photometric and astrometric data for 357 million unique objects[2], detected in 11,663 sq. deg. About half of these objects are unresolved.

2.2. 2MASS Survey

The Two Micron All Sky Survey used two 1.3 m telescopes to survey the entire sky in near-infrared light (Skrutskie et al. 1997; Cutri et al. 2003). Each telescope had a camera was equipped with three 256× 256 arrays of HgCdTe detectors with 200 pixels and observed simultaneously in the J (1.25 μm), H (1.65 μm), and K_s (2.17 μm, hereafter K) bands. The detectors were sensitive to point sources brighter than about 1 mJy at the 10σ level, corresponding to limiting magnitudes of 15.8, 15.1, and 14.3, respectively (Vega based; for corrections to AB magnitude scale see below) . Point-source photometry is repeatable to better than 10% precision at this level, and the astrometric uncertainty for these sources is less than 0.2 arcsec. The 2MASS catalogs contain positional and photometric information for about 500 million point sources and 2 million extended sources.

[2]For more details, see http://www.sdss.org/dr7/

2.3. The Main-Sample Selection

The main sample is selected from the SDSS Data Release 7 using the following two main criteria:

1. Unique unresolved sources: binary processing flags DEBLENDED_AS_MOVING, SATURATED, BLENDED, BRIGHT, and NODEBLEND must be false, and parameter nCHILD=0.
2. The r-band magnitudes (uncorrected for extinction) must satisfy $r < 21$.

yielding 73 million stars (for an SQL query used to select the main sample see Appendix A). The distribution of selected sources on the sky is shown in Figure 1.

For isolated sources, the $r < 21$ condition ensures that photometric errors are typically not larger than 0.05 mag (see Fig. 1 in Sesar et al. 2007). For sources with $r < 19$, the errors reach their systematic limit of ∼0.02 mag. However, for sources in complex environments errors can be much larger, and sometimes reported errors are unreliable. If the cataloged photometric error is larger than 0.5 mag in the $griz$ bands, or larger than 1.5 mag in the u band, that data point is not used in the analysis (formally, we reset the magnitudes to 999.9 and their errors to 9999.9 in publicly available files, see Appendix B).

2.4. SDSS-2MASS Subsample

Following Covey et al. (2007), acceptable 2MASS sources must have 2MASS quality flags $rd\_flag == 222$, $bl\_flag == 111$, and $cc\_flag == 0$, and selected 2MASS sources are positionally matched to SDSS sources with a distance cutoff of 1.5 arcsec. The combined SDSS-2MASS catalog contains ∼23 million sources. The wavelength coverage of the SDSS and 2MASS bandpasses are shown in Fig. 3 in Finlator et al. (2000). The distributions of

SDSS-2MASS sources in various color-color and color-magnitude diagrams are discussed in detail by Finlator et al. (2000) and Covey et al. (2007). We emphasize that practically all sources in an SDSS-2MASS point source sample defined by a K-band flux limit are sufficiently bright to be detected in all other SDSS and 2MASS bands. For orientation, main sequence stars selected by the condition $K < 14.3$ are closer than approximately 1-2 kpc (with red stars much closer than blue stars).

Similarly to the treatment of SDSS photometry, for stars with reported errors in the J, H, and K bands greater than 0.5 mag, we reset magnitudes and errors to 999.9 and 9999.9, respectively. The Vega-based 2MASS photometry is translated to SDSS-like AB system following Finlator et al. (2000)

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

Note that these corrections have no impact on fitting and the results (because the same corrections are applied to models and observations and thus cancel out, see below), but are convenient when visualizing SEDs.

2.5. Model Assumptions and Fitting Procedures

There are two empirical results that form the basis of our method. First, the stellar locus in the multi-dimensional color space spanned by SDSS and 2MASS colors is nearly one dimensional (because for most stars the effective temperature has much more effect on colors than other physical parameters, such as age and metallicity). The locus position reflects basic stellar physics and is so well defined that it has been used to test the quality of SDSS photometry (Ivezić et al. 2004), as well as to calibrate new photometric data

(?). Second, the *shape* of the dust extinction curve can be described as a one-parameter family, usually parametrized by $R_V = A_V/E(B-V)$ (????). Using this parametrization, extinction in an arbitrary photometric bandpass λ is equal to

$$A_\lambda = C_\lambda(R_V)\, A_r, \tag{2}$$

where A_r is extinction in the SDSS r band, and $C_\lambda(R_V)$ describes the shape of the extinction curve. Hence, the observed colors can be fit using only three free parameters: the position along the locus, R_V, and A_r. Some caveats to this statement, such as the fact that not all unresolved sources are found along the locus (e.g., quasars and unresolved binary stars), and that even for fixed dust properties A_r and A_λ depend on the source spectral energy distribution, are discussed in quantitative detail further below.

The best-fit empirical stellar model from a library described in §2.6, and the dust extinction according to a $C_\lambda(R_V)$ parametrization described in §2.7, are found by minimizing χ^2_{pdf} defined as

$$\chi^2_{pdf} = \frac{1}{N-2}\sum_{i=1}^{N}\left(\frac{c_i^{obs} - c_i^{mod}}{\sigma_i}\right)^2, \tag{3}$$

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). The model colors are constructed using extinction-corrected magnitudes

$$m_\lambda^{corr} = m_\lambda^{obs} - A_\lambda, \tag{4}$$

with $\lambda = (ugriz[JHK])$, resulting in

$$c^{mod} = c^{lib}(t) + [C_{\lambda 2}(R_V) - C_{\lambda 1}(R_V)]\; A_r. \tag{5}$$

Here $\lambda 1$ and $\lambda 2$ correspond to two adjacent bandpasses which define colors c^{mod} and c^{lib}. Hence, by minimizing χ^2_{pdf}, we obtain the best-fit values for three free parameters: R_V, A_r, and the model library index, t (intrinsic stellar color, or position along the locus). Once these parameters are determined, the overall flux normalization (i.e. apparent magnitude offset) is determined by minimizing χ^2_{pdf} for the fixed best-fit model.

We minimize χ^2_{pdf} by brute force method. All 228 library SEDs are tried, with dust extinction values in the range $0 \leq A_r \leq 10$, and with 0.02 mag wide steps. This is not a very efficient method, but the runtime on a multi-processor machine was nevertheless much shorter, in both human and machine time, than post-fitting analysis of the results. We investigate the impact of R_V by producing two sets of best-fit t and A_r. First, we use fixed $R_V = 3.1$, and then allow R_V to vary in the range 2-6, with 0.1 wide steps. The results for the two cases are compared and analyzed in the next section.

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), as well as for the finite locus width. In principle, σ_i could be varied with the trial library SED to account for the varying width of the stellar locus. We have not implemented this feature because it does not dominate the systematic errors.

For a given R_V value (whether constant, or a grid value in the free R_V case), once the minimum χ^2, χ^2_{min}, is located, 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):

$$\chi^2(t, A_r|R_V) = a(t-t^*)^2 + b(t-t^*)(A_r - A_r^*) + c(A_r - A_r^*)^2 \tag{6}$$

were t is the model index, and t^* and A_r^* are the best-fit 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

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

$$\sigma_A = \left(c - \frac{b^2}{4a}\right)^{-\frac{1}{2}} \tag{8}$$

Note that the b coefficient controls the covariance between t^* and A_r^*. 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. Covey et al. Stellar SEDs

Covey et al. (2007) have quantified the main stellar locus in the $ugrizJHK$ photometric system using a sample of ∼600,000 point sources detected by SDSS and 2MASS. They tabulated the locus position and width as a function of the $g-i$ color, for 228 $g-i$ values in the range $-0.25 < g-i < 4.50$. We adopt this locus parametrization as our empirical SED library.

This $g-i$ parametrization reflects the fact that the stellar effective temperature, which by and large controls the $g-i$ color, is more important than other physical parameters, such as age (gravity) and metallicity, in determining the overall SED shape (for a related discussion and principal component analysis of SDSS stellar spectra see McGurk, Kimball & Ivezić 2009). The adopted $g-i$ range includes the overwhelming majority of all unresolved SDSS sources, and approximately corresponds to MK spectral types from early A to late M. Due to the 2MASS flux limits, the stellar sample analyzed by Covey et al. does not include faint blue stars (those with $r \gtrsim 16$ for $g-r < 0.6$; see Fig. 4 in Finlator et al. 2000). Only faint blue SDSS stars are dominated by low-metallicity halo stars (see Fig. 3 in I08), and thus the Covey et al. locus corresponds to predominantly metal-rich main sequence stars ($[Fe/H] > -1$).

The adopted model library cannot provide a good fit for SEDs of unresolved pairs of white and red dwarfs (Smolčić et al. 2004), hot white dwarfs (Eisenstein et al. 200x), and quasars (Richards et al. 200x). Systematic photometric discrepancies at the level of a few hundredths of a magnitude are also expected for K and M giants, especially in the u band (Helmi et al. 200x). Similar discrepancies are expected for metal-poor stars (I08). Nevertheless, all these populations contribute only a few percent of the full sample (Finlator et al. 2000, J08), and in most cases can be recognized by resulting large values of χ^2_{min}. At least in principle, additional libraries can be used to fit observed SEDs of sources that have large χ^2_{min} for SEDs of main sequence stars (such as quasars, unresolved binary stars, blue horizontal branch stars, and white dwards). This additional analysis has not been attempted here.

2.7. Parametrization of Dust Properties

In order to implement the fitting method described in §2.5, the shape of the extinction curve (C_λ, see eq. 2) must be characterized. C_λ in the SDSS bands was initially computed (prior to the beginning of the survey, to enable spectroscopic targeting) using the standard parametrization of the extinction curve (Cardelli, Clayton & Mathis 1989; O'Donnell 1994) with $R_V = 3.1$. The resulting values (C_λ=1.87, 1.38, 0.76, 0.54, with $\lambda = u, g, i, z$) are commonly adopted to compute the extinction in the SDSS bands, together with A_r given by the SFD map via $A_r = 2.75E(B-V)$.

A preliminary analysis of the position of the stellar locus in the SDSS-2MASS color space suggested that these values need to be changed somewhat (Meyer et al. 2005). Here we revisit the Meyer et al. analysis using an improved SDSS photometric catalog from the so-called stripe 82 region[3] (Ivezić et al. 2007). The SDSS-2MASS subset of that catalog includes 102,794 sources unresolved by SDSS, and which also have a 2MASS source with $K < 14.3$ within 1.5 arcsec. The results of our analysis provide an updated set of

[3]Available from http://www.astro.washington.edu/users/ivezic/sdss/catalogs/stripe82.html

C_λ coefficients, which are then used to select a dust extinction model for generating the required $C_\lambda(R_V)$ dependence. Similarly to the recent analysis by Schlafly et al. (in prep, hereafter Sch2010), we find that the O'Donnell (1994) model can be rejected, and adopt the CCM dust extinction law (?).

2.7.1. Determination of the locus shifts

The interstellar extinction reddens the stellar colors and shifts the position of the *whole* stellar locus at high galactic latitudes, where practically all stars are hidden behind the dust screen. At high galactic latitudes, distances to an overwhelming majority of stars are larger ($\gtrsim$ 100 pc) than the characteristic scale height of the interstellar dust layer (∼70 pc, J08). Both the amount of reddening and its wavelength dependence can be determined by measuring the locus position and comparing it to the locus position corresponding to a dust-free case. The latter can be determined in regions with very small extinction ($A_r \sim 0.05$) where errors in the SFD extinction map as large as 20% would still be negligible.

We measure the locus position in the seven-dimensional SDSS-2MASS color space using an extended version of the "principal color" method developed by Ivezić et al. (2004) to track the SDSS photometric calibration quality. We utilize six independent two-dimensional projections spanned by the $r-K$ and $\lambda-r$ colors, where $\lambda = u, g, i, z, J$ and H (see Fig. 2). Since the extinction in the 2MASS K band is small and fairly model and R_V-independent ($A_K/A_r = 0.133$ for $R_V = 3.1$, with a ∼10% variation over the range of plausible R_V and dust models, as discussed further below), the locus shifts in the $r-K$ direction provide robust constraints for A_r. For example, a 10% uncertainty in the A_K/A_r ratio results in only 1.5% uncertainty in A_r determined from a given $A_r - A_K$ value. We determine these shifts iteratively, starting with A_r given by the SFD map, and adjusting A_r until the observed and corrected $r-K$ color distributions agree in a maximum likelihood sense. This determination of A_r is very similar to the "blue tip" method introduced in Sch2010. The main two differences are due to the addition of 2MASS data. First, the low-metallicity faint blue stars are not included in the sample analyzed here. Such stars could systematicaly influence the locus morphology and reddening estimates; nevertheless, our results are in good agreement with the Schlafly et al. results, as discussed below. Second, the availability of the K magnitudes enables a robust and straightforward determination of A_r, and without any consideration of the SFD map.

After A_r is estimated from the $r-K$ color shifts, the locus shifts in the $\lambda - r$ directions then provide constraints for the extinction wavelength dependence, C_λ. We measure these shifts using principal colors, P_1 and P_2, with P_1 parallel to the blue part of the stellar locus, and P_2 perpendicular to it (see the top left panel in Fig. 2 for illustration of the principal axes, and for a comparison of the locus orientation with the direction of the standard reddening vectors). The blue part of the stellar locus at the probed faint magnitudes ($14 < r < 17$) includes mostly thick disk stars with distances of the order 1 kpc or larger, which are thus beyond all the dust.

We measure the position of the blue part of the locus in each $\lambda - r$ vs. $r-K$ diagram using stars with $1.5 < r-K < 2.5$ (approximately; the range is enforced using the $P_1(\lambda)$ color). The blue part of the locus is parametrized as

$$P_1(\lambda) = \cos(\theta_\lambda)\,(r-K) + \sin(\theta_\lambda)\,(\lambda - r) + c_1(\lambda) \tag{9}$$

and

$$P_2(\lambda) = -\sin(\theta_\lambda)\,(r-K) + \cos(\theta_\lambda)\,(\lambda - r) + c_2(\lambda). \tag{10}$$

The best-fit angle θ_λ found using stripe 82 dataset is equal to (61.85, 33.07, 14.57, 23.47, 34.04, 43.35) deg. for $\lambda = (u, g, i, z, J, H)$. The values of c_1 and c_2 are completely arbitrary; we set $c_1 = 0$, and determine $c_2(\lambda)$ by requiring that the median value of $P_2(\lambda)$ color is 0

(c_2=0.463, 0.434, 0.236, 0.424, −0.048, −0.019, for u, g, i, z, J, H, respectively). Given the locus shift $\Delta P_2(\lambda)$, and A_r determined from the $r-K$ color offset (or alternatively from the ΔP_1 offsets), the corresponding A_λ can be determined from

$$C_\lambda \equiv \frac{A_\lambda}{A_r} = 1 + \tan(\theta_\lambda)(1 - \frac{A_K}{A_r}) + \frac{1}{\cos(\theta_\lambda)} \frac{\Delta P_2(\lambda)}{A_r}. \tag{11}$$

Assuming a constant A_K/A_r ratio, it is straightforward to compute the error of this estimate.

The locus position must be measured over a sky area where the amount of dust and dust properties can be assumed constant. The smaller the area, the more robust is this assumption. However, the chosen area cannot be arbitrarily small because the error in the locus position, and thus the C_λ error, is inversely proportional to the square root of the star counts. Within the analyzed stripe 82 region, the counts of SDSS-2MASS stars in the blue part of the stellar locus never drop below 70 stars/deg^2. We bin the data using 4 degree wide bins of R.A. (with $|Dec|$ <1.27 deg., an area of ∼10 deg^2 per bin), which guarantees that random errors in A_λ never exceed ∼2% (even for the u band, and a factor of few smaller in other bands). In addition, we consider four larger regions: the high-latitude northern sky with $b > 45°$, split into $l < 180°$ and $l > 180°$ subregions, a northern strip defined by $30° < b < 45°$, and a southern strip defined by $-45° < b < -30°$.

2.7.2. Interpretation of the locus shifts and adopted dust extinction model

We find that the variations in the shape of the extinction curve acrros the 28 R.A. bins from Stripe 82 region are consistent with measurement errors. The values of C_λ obtained for the whole Stripe 82 region are listed in the first row in Table 1. Practically identical coefficients are obtained for the southern strip defined by $-45° < b < -30°$. The extinction curve values for the northern sky are consistent with the southern sky. One of the largest discrepancies is detected in a region from the northern strip defined by $30^\circ < b < 45^\circ$ and $0^\circ < l < 10^\circ$, and these values are listed in the second row in Table 1. Nevertheless, the north vs. south differences are not large, and using models described below, correspond to an R_V variation of about 0.1.

Much larger north vs. south differences are detected when comparing the best-fit A_r values to the values given by the SFD map. The accuracy of the A_r determined here is about 3-10%, depending on the amount of dust. We find that the SFD A_r values are consistently larger by about 20% than the values determined here across the southern hemisphere. Interestingly, no such discrepancy is detected across the northern sky, to within measurent errors of $\sim$5%. In several isolated regions, the discrepancies can be much larger. For example, in a region defined by $-45^\circ < b < -30^\circ$ and $157^\circ < l < 160^\circ$, the SFD values appear overestimated by 50% (the median value of A_r given by the SFD map is 1.3). These results are similar to those presented in Sch2010, where the spatial variation of errors in the SFD map and their possible causes are discussed in more detail.

We adopt the C_λ values determined for Stripe 82 region (the first row in Table 1) to select a dust extinction law used in subsequent fitting of SEGUE data. Using the same assumptions and code as Sch2010, we compute dust extinction curve for three popular models, and for three different input stellar spectral energy distributions. As can be seen in Figure 3, the differences between the models are much larger than the impact of different underlying spectra.

A comparison of the observational constraints and model predictions is summarized in Figure 4. Following Sch2010, we use ratios of the reddening values for this comparison. The differences in the extinction curve shape between the southern and northern sky determined here are similar to the difference from the Sch2010 results, and are consistent with estimated measurement uncertainties. As evident, the O'Donnell (1994) model predicts unacceptable values of the $(A_r - A_i)/(A_i - A_z)$ ratio for all values of R_V. The other two models are in fair agreement with the data. Due to a slight offset of the Sch2010 measurements, they argued that the CCM model (?) is also unsatisfactory, though the discrepancy was not as large as in the case of the O'Donnell (1994) reddening law.

Although none of the models shows a perfect agreement with the data, discrepancies are not large. To further illustrate the constraints from different bands, we determine the best-fit R_V and its uncertainty in each band using the CCM model. If a model is acceptable, the constraints from different bands have to be statistically consistent. As shown in Figure 5), this is indeed the case, and we obtain the best-fit $R_V = 3.01 \pm 0.05$. The systematic error of this estimate, implied by the variation of the extinction curve shape across the analyzed regions is about 0.1. The corresponding figure for the F99 (?) reddening law looks similar, with the best-fit $R_V = 3.30 \pm 0.1$, while for the O'Donnell (1994) model, $R_V = 3.05 \pm 0.05$. However, for the latter, the predicted extinction in the i band is inconsistent with the rest of the bands at about 2σ level (see Figure 6). The predicted values of the extinction curve for all three models, using their individual best-fit R_V, are listed in Table 1.

For the rest of our analysis, we generate $C_\lambda(R_V)$ values using the CCM law and an F star spectral energy distribution (6500 K). The adopted curves are shown in Figure 7, and a few representative values are listed in Table 2.

2.8. Illustration of the Method and Fitting Degeneracies

To summarize, we make two basic assumptions when analyzing observed SEDs of low-latitude stars (SEGUE strips). First, we assume that the median stellar locus in SDSS and 2MASS bandpasses, as quantified by Covey al al. (2007) at high galactic latitudes, 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, $R_V = A_V/E(B-V)$. 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 (position along the nearly one-dimensional locus), t, dust amount, A_r, and R_V.

When the number of measured colors is small, when the color errors are large, or when the sampled wavelength range is not sufficiently wide, the best-fit solutions can be degenerate. The main reason for this degeneracy is the similarity of the stellar locus orientation and the direction of the dust reddening vector. Figure 8 illustrates an example of degenerate solutions in the $r-i$ vs. $g-r$ color-color diagram, 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 diagram is essentially independent of R_V, the measured $r-i$ and $i-z$ colors provide robust constraints for t and A_r, irrespective of R_V. The addition of the measured $g-r$ color to $r-i$ and $i-z$ colors then constrains R_V.

Since the stellar locus in the $i-z$ and $r-i$ color-color diagram and the reddening vector are not perpendicular, there is non-zero covariance between the best-fit t 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.

2.9. Tests of the Method

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 $g-i = 1.95$ (roughly at the "knee" of the stellar locus in the $r-i$ vs. $g-r$ color-color diagram) 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 9 and 10.

The errors in the best-fit stellar SED, parametrized by the $g-i$ color, are about twice as large as the assumed photometric errors. When photometric errors exceed about 0.03 mag, the best-fit A_r distribution becomes bimodal. Therefore, even the addition of the red passbands is insuficient to break the stellar color–reddening degeneracy when the photometry is inaccurate.

In the second test, we have investigated the covariance between the best-fit model and A_r values. Figure 11 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 8 (the angle between the reddening vector and the stellar locus is smaller for the blue part of the locus, than for the red part). The A_r vs. $g-i$ covariance does not strongly depend on assumed A_r. When the 2MASS bands are added, the morphology of the χ^2 surface is essentially unchanged.

These tests shows that the implementation of the χ^2 minimization produces correct results, and that the accuracy of SDSS and 2MASS photometry is sufficiently accurate (for most sources) to break degeneracy between the dust reddening and intrinsic stellar color.

3. ANALYSIS OF THE RESULTS

We apply the method described in the preceeding Section four times. We fit separately the full SDSS dataset (73 million sources) using only SDSS photometry, and the SDSS-2MASS subset (23 million sources) using both SDSS and 2MASS photometry. We first consider a fixed C_λ extinction curve determined for Stripe 82 region (the coefficients listed in the first row in Table 1), and refer to it hereafter as the "fixed $R_V = 3.1$" case (although the best-fit CCM model corresponds to $R_V = 3.01 \pm 0.05 \pm 0.1$). These fixed-$R_V$ fits are obtained for the entire dataset, including the high galactic latitude regions where the extinction is too small to reliably constrain the shape of the extinction curve using data for individual stars. In order to investigate the variation of R_V in high-extinction low galactic latitude regions, we use the CCM C_λ curves discussed in Section 2.7.2 (and shown in Figure 7). We only consider the ten SEGUE stripes, limited to the latitude range $|b| < 30^\circ$, which include 37 million sources in the full SDSS dataset, and 10 million sources in the SDSS-2MASS subset.

The resulting best-fit parameter set is rich in content and its full scientific exploitation is far beyond the scope of this paper. The main purpose of the preliminary analysis presented below is to illustrate the main results and demonstrate their reliability, as well as to motivate further work – all the data and the best-fit parameters are made publicly available, as described in Appendix B.

3.1. Fixed R_V Case

We first consider results based only on SDSS photometry and with the dust extinction curve ("fixed R_V" case). A comparison of SDSS-only and SDSS-2MASS results is discussed further below.

3.1.1. The Northern Galactic Cap Region

Given small A_r (the median A_r is $\sim$0.1 mag), we do not expect stable solutions when R_V is a free parameter. Even for a fixed extinction curve, the results for individual stars can have uncertainties as large as A_r given by the SFD map. Nevertheless, by taking a median value for typically hundred stars, a map can be produced that closely resembles the SFD map (Figure 13). Add details about pixels and about bias in the best A_r (must have $r < 18$ for a small bias; increasing errors as shown by Sesar et al.).

The residuals show structure reminiscent of the SDSS scanning pattern (Figure 13). Can we place a limit on zeropoint errors? Does the patter persists when 2MASS data is added?

Stripping (for SDSS, 15¡r¡18) corresponds to jump in Ar from 0.08 to 0.13

3.1.2. The SEGUE Stripes

A summary of the best-fit A_r for the ten SEGUE stripes, in the range $|b| < 5°$ and for three distance slices, is shown in Figure 14. Distances to stars are determined by assuming that all sources are main sequence stars, and using photometric parallax relation from I08 with $[Fe/H] = -0.4$ (with the best-fit intrinsic colors). An expected scatter in metallicity of 0.2-0.3 dex corresponds to about 10-15% uncertainty in distance. Although not all sources are main sequence stars (such as red giants, which have grossly underestimated distances, see below for discussion), their fraction is sufficiently large that the results are not strongly biased[4]. Furthermore, sources with SEDs significantly different from the main sequence SEDs are not included: the figures are constructed only with sources that have

[4]Here we will add analysis of Mario's simulated sample which supports this statement.

the best-fit $\chi^2_{pdf} < 3$. The distribution of the best-fit χ^2_{pdf} for all sources in all ten slices is shown in Figure 15. As evident, about 85% of all sources satisfy the $\chi^2_{pdf} < 3$ condition.

3.1.3. Analysis of the $l \sim 100^\circ$ Region

For detailed analysis of the best-fit results, we select a single fiducial slice with $l \sim 110^\circ$. The distribution of the best-fit A_r for this region is shown in Figure 16. As evident, the best-fit A_r increases with the stellar distance, although the two are determined independently (distance is computed a posteriori, from the best-fit apparent magnitude). More distance slices for the $l \sim 110^\circ$ region, on a finer grid, are shown in Figure 17.

An interesting behavior can be seen in Figure 17: while A_r increases with distance for bins more distant than 200 pc, the closest distance bin (100-200 pc) has a large fraction of stars with much larger A_r than seen in the 200-300 pc bin. Furthermore, the morphology of the large A_r distribution is similar to that for the most distant (3-4 kpc) bin (e.g., a cloud at $l \sim 111^\circ$ and $b \sim -3^\circ$. It turns out that the closest distance bin is "contaminated" by stars on the red giant branch. This conclusion is supported by their $g - i$ color distribution, which is narrow and peaked $g - i \sim 2$ (see Figure 26). SDSS colors of red giants are very similar to those of main sequence stars (to within a few hundreths of magnitude, e.g., see **?**), and it is not easy to distinguish them at high galactic latitudes using photometry alone. However, at low galactic latitudes considered here they are easily separated from main sequence stars because their best-fit A_r is much larger than expected for their corresponding *main sequence distance*. At $g - i \sim 2$, the absolute magnitude of main sequence stars is $M_r \sim 8.5$ (I08), while the tip of the K/M giant branch is at $M_r \sim -2.3$ (**?**). This absolute magnitude difference corresponds to a factor of ~100 in distance at the same apparent magnitude! Hence, red giants in the closest distance bin have grossly underestimated distances, with true distances of the order 10 kpc.

In other words, red giants can be easily identified as sources with good fits that are found above the A_r vs. distance curve traced by main sequence stars which dominate the sample, and at (main sequence) distances below 1 kpc.

Originally ZI figure, but Mike's look better: redo for the $|b| < 10°$ range, and distance slices 100-200, 200-300, 300-400, 2000-2500 and 3000-4000 pc. Point out red giants and refer to analysis of the spatial distribution of dust in Section 3.3

Discuss comparison with the SFD maps: Figure 18.

Difference is due to distant molecular clouds: Figure 19. Can measure MC distance! (using SFD as proxy, with LSST better, mention in Discussion)

3.2. Free R_V Case

What happens when RV is free? Explain Figure 20. Differences in A_r can be large.

Comparison of fixed and free R_V fits: differences in A_r in Figure 21. Mike, is the color map for star counts on log or linear scale? Analogous for the best-fit $g - i$: reference Figure 22.

3.2.1. Comparison of the SDSS-only and SDSS-2MASS Results

Comparison of SDSS and SDSS-2MASS results: A_r differences in Figure 24, A_r vs. R_V covariances in Figure 25, and $g - i$ histograms in Figure 26.

Three different types of best-fit SEDs: using only SDSS data with fixed $R_V = 3.1$, using only SDSS data with free R_V, and using both SDSS and 2MASS data with free R_V (green line): forward-reference Figure 28

The results for 3D R_V distribution: Figure 29 for the SDSS-2MASS subset and distance limit of $\sim$1 kpc, and for the full SDSS dataset with distance limit of $\sim$3 kpc in Figure 30.

3.3. The Spatial Distribution of Dust

The best-fit A_r as a function of distance from the Galactic plane, Z, and distance along the plane, D_{xy}, is shown in Figure 31. Each pixel in the maps shows the median A_r (over many stars from that pixel) between the observer and that point (i.e. this is not a cross-section of three-dimensional dust map). The extinction generally increases with distance for most lines of sight. The translation of these "integral" constraints on the dust distribution into a true three-dimensional distribution will be presented in a future publication (Schlafly et al., in prep.).

A few hot pixels close to the origin are likely due to red giants with understimated distances. Discuss the $g - i$ color distribution shown in Figure 26.

Extract red giants from all files and discuss their number and distance limit. Is there relevance for SDSS-III?

Dust scale height: a pretty plot, Figure 32, but exactly does it show? Mike, help!

3.4. The Spatial Distribution of Stars

The spatial distribution of stars is shown in Figure 33. Is this SDSS data, fixed R_V case? Plot made by ZI.

The fall-off of the stellar volume number density at distance beyond 1-2 kpc is due to the sample faint limit, and does not reflect the disk structure.

Comment on distance limits, exponential vs. sech2 disk profile.

4. SUMMARY AND DISCUSSION

This is the first analysis based on SDSS data that simultaneously estimates intrinsic stellar color and dust extinction along the line of sight for millions of stars detected in SEGUE survey.

4.1. Summary of the Methodology

Both in case of single stars that are projected onto the locus in the multi-dimensional color space, and for the shifts of the whole locus at high galactic latitudes, there is always one color constraint fewer than the number of bands. One way of thinking about this "missing" equation is that extinction is described by Ar and four (or seven) measures of the shape of the extinction curve, $C_\lambda = A_\lambda / A_r$.

There are three different approaches that we can use to "close" this system:

```
1) Ar comes from the SFD map. In this case, A_c=(C_m1-C_m2)*ArSFD, and from
   the resulting four equations we get four C_m. It is easy to show that
        Cu = 1  + (Aug+Agr)/ArSFD
        Cg = 1  + Agr/ArSFD
         Ci = 1  - Ari/ArSFD
        Cz = 1  - (Ari+Aiz)/ArSFD

If there are errors in ArSFD, the resulting Cm will look weird (see below).
But we still can ask questions such as "do we always get the same Cm?", and
```

"Is there correlation between Cm variation and the input ArSFD?". In particular, given that the nature gave us many lines of sight with varying ArSFD, we can assume some spatially invariant error model for the ArSFD map (say, a multiplicative factor) and determine its free parameters by minimizing the variance of Cm over the chosen piece of sky.

To investigate the impact of errors in ArSFD on this method, I used the CCM model with Rv=3.1, computed Aug etc using Ar=1, and assumed a multiplicative error in ArSFD. For a correction factor of 0.95, the best-fit Rv is 3.28 with the i and z band constraints biased to even higher values. For a correction factor of 0.9, the best-fit Rv is 3.48, with the i and z reddening values barely consistent with the CCM curve. For additive errors such that the true Ar is ArSFD+0.05, the best-fit Rv varies from 2.2 at ArSFD=0.1to 2.93 at for ArSFD=1.0.

2) use models that predict the four C_m values as a function of *single* parameter Rv. Given that now we have three "spare" constraints, we can also play games such as checking consistency of model predictions and thus selecting the "best" model. With this method, we can also directly test ArSFD, though only in a model-dependent way.

3) assume (fix) one value of C_m and solve for other C_m and Ar. This may sound crazy at first, but it works when you add 2MASS data. The K band is at 2.2 mic, which is almost as large as infinity. With both SDSS and

2MASS, Ar comes from r-K shifts, and relies on the fact that AK/Ar is small (0.132) and varies little with Rv and among models. For example, for Rv>2 all models predict AK/Ar variation of well within 20%, and this variation translates to only about 3% error when the r-K shift is interpreted as Ar-AK=(1-AK/Ar)*Ar=0.868*Ar. In this case, one has eight bands, makes 7 colors, and gets the Ar constraint and six C_m constraints.

How do we measure Aug, Agr, etc. using the stellar locus? If you think of the locus as an "image", then in principle we "slide" the image of the reddened sample to perfectly align with the "intrinsic" dereddened locus. We can do that in each 2D color-color diagram (in which case, e.g., the g-r shift better be the same in both g-r vs. u-g, and r-i vs. g-r diagrams), or we can determine all four shifts simultaneously in 4D space. And of course, you can look at 1D projections of each color and adjust the position of the blue tip as in your method.

If there were no astrophysical systematics, all this would be easy and different methods should produce identical results. But as we do have distance, age and metallicity effects, we need to be careful. For example, red stars can be as close as 100 pc and be within dust layer, and the blue tip is sensitive to age and metallicity of turnoff stars that define it. The idea behind PC2 method is to avoid distance effects by considering only blue stars, and to avoid age and metallicity effects by measuring shifts perpendicular to the locus. The reason for the latter is that the variation of age will "extend" or "shorten" the locus (shift the blue tip) but will not effect its position in the P2 direction. With metallicity, the effects are a little bit more complicated, but mostly confined to the u band. For blue stars, the g-r color is essentially a measure of Teff, with negligible dependence on metallicity. At a given g-r color, the u-g color depends on metallicity (becomes bluer, see the top right panel in fig.2 from Tomography II). For example, at g-r=0.3, the u-g color varies by about 0.2 mag as the metallicity varies from the median thick-disk value (-0.5) to the median halo value (-1.5). This shift is of course not parallel to the locus so it does have some effect on the PC2 distribution. However, already at g-r=0.5, the fraction of halo stars is sufficiently small that this effect is more or less negligible. Thus, in the range 0.5 < g-r < 1.2, only the reddening (and calibration errors, of course!) can shift the locus perpendicularly to its blue part (even in the u-g vs. g-r diagram). An added benefit is that the PC2 distributions are very narrow which more than offsets the fact that the reddening vectors are measured only along PC2 direction from the SNR point of view.

The PC2 colors are defined as linear combinations of colors:

PC2 = a*c1 + b*c2 + c

with coefficients a, b and c known (a and b come from the orientation of the locus, and c is set so that <PC2>=0). A measurement of the median PC2 offset is then

Delta(PC2) = a*Delta*(c1) + b*Delta(c2)

There are three SDSS color-color diagrams and thus three Delta(PC2) measurements, but as you pointed out, there are four colors! We again need to "close" the system by some ansatz.

When 2MASS data are available, the r-K distribution is used as an additional constraint to derive Delta(c). Indeed, the way we set it up is to always consider colors based on the r band, m-r vs. r-K. This way, the extinction in the m band is measured in only a single color-color diagram, and values in different bands are covariant only through the adopted r-K reddening value. When we look at the ratios, such as (Au-Ag)/(Ag-Ar), this covariance by and large cancels out (to the extent that AK/Ar=0.133 is a correct assumption).

When only SDSS data are available (finally answering your question!), one needs to adopt an additional locus parametrization. In the context of calibration testing for SDSS (where this PC method was developed), we used the so-called x color, which is essentially the g-r color for red stars defined by 0.8 < r-i > 1.6. By using this color, and the so-called w color (which is PC2 for blue stars in the r-i vs. g-r diagram), we are effectively doing the "locus image alignment" that I mentioned above. The measured offsets of these four colors (s is PC2(ug-gr), w is PC2(gr-ri), y is PC2(ri-iz), and x) can then be easily (four equations with four unknowns) transformed in the implied offsets of u-g, g-r, etc.

So, yes, you were right to raise this point as the system is not only based on three PC2 colors, but also on the x color! My discussion of potential age/metallicity effects on the blue tip position does not prove that these effects are large in practice. We will find out in detail when you compare the u-g, etc. offsets obtained from the blue tip method with the analogous offsets obtained using the PC method.

4.2. Future Surveys

The results presented here will be greatly extended by several upcoming large-scale, deep optical surveys, including the Dark Energy Survey (Flaugher 2008), Pan-STARRS (Kaiser et al. 2002), and the Large Synoptic Survey Telescope (Ivezić et al. 2008a). These surveys will extend the faint limit of the sample analyzed here by $4 - 6$ mag. These upcoming studies are thus certain to provide valuable new information about the dust and stellar distribution within the Galactic disk beyond the current limiting distance of a few kpc. Also, better seeing, multiple epochs.

Ž. Ivezić and B. Sesar acknowledge support by NSF grants AST-615991 and AST-0707901, and by NSF grant AST-0551161 to LSST for design and development activity. J. Dalcanton acknowledges NSF CAREER grant AST-02-38683. D. Schneider acknowledges support by NSF grant AST-06-07634. T.C. Beers acknowledges partial support from PHY 08-22648: Physics Frontier Center/Joint Institute for Nuclear Astrophysics (JINA), awarded by the U.S. National Science Foundation. We acknowledge the hospitality of the KITP at the University of California, Santa Barbara, where part of this work was completed (supported by NSF grant PHY05-51164). Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.

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. SQL Query Example

The following SQL query was used to select and download data for all SDSS stars with spectroscopic and proper-motion measurements (see http://casjobs.sdss.org/CasJobs).

```
SELECT
  round(p.ra,6) as ra, round(p.dec,6) as dec,
  p.run, p.camcol, p.field,            --- comments are preceded by ---
  round(p.extinction_r,3) as rExtSFD, --- r band extinction from SFD
  round(p.modelMag_u,3) as uRaw,   --- N.B. ISM-uncorrected model mags
  round(p.modelMag_g,3) as gRaw,   --- rounding up
  round(p.modelMag_r,3) as rRaw,
  round(p.modelMag_i,3) as iRaw,
  round(p.modelMag_z,3) as zRaw,
  round(p.modelMagErr_u,3) as uErr,
  round(p.modelMagErr_g,3) as gErr,
```

```
  round(p.modelMagErr_r,3) as rErr,
  round(p.modelMagErr_i,3) as iErr,
  round(p.modelMagErr_z,3) as zErr,
  (case when (p.flags & '16') = 0 then 1 else 0 end) as ISOLATED,
  ISNULL(round(t.pmL,3), -9999) as pmL, --- proper motion data are set to
  ISNULL(round(t.pmB,3), -9999) as pmB, --- -9999 if non-existent (NULL)
  ISNULL(round(t.pmRaErr,3), -9999) as pmErr  --- if pmErr < 0 no pm data
INTO mydb.dustSample
FROM phototag p LEFT OUTER JOIN propermotions t ON
  (p.objID = t.objID and t.match = 1 and t.sigra < 350 and t.sigdec < 350)
        --- quality cut on pm
WHERE
  p.type = 6 and              --- select unresolved sources
  (p.flags & '4295229440') = 0 and --- '4295229440' is code for no
                             --- DEBLENDED_AS_MOVING or SATURATED objects
  p.mode = 1  --- PRIMARY objects only, which implies
            --- !BRIGHT && (!BLENDED || NODEBLEND || nchild == 0)]
  p.modelMag_r < 21  --- adopted faint limit
--- the end of query
```

B. Data Distribution

Describe website

REFERENCES

Abazajian, K. N. et al. 2009, ApJS, 182, 543

Bond, N. et al. 2009, ApJ, in press (also arXiv:0909.0013)

Flaugher, B. 2008, in A Decade of Dark Energy: Spring Symposium, Proceedings of the conferences held May 5-8, 2008 in Baltimore, Maryland. (USA). Edited by Norbert Pirzkal and Henry Ferguson. http://www.stsci.edu/institute/conference/spring2008

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

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

—. 2006, AJ, 131, 2332

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

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

Ivezić, Z. et al. 2008a, ArXiv e-prints

Ivezić, Ž. et al. 2008b, ApJ, 684, 287

Jurić, M. et al. 2008, ApJ, 673, 864

Kaiser, N. et al. 2002, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4836, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. J. A. Tyson & S. Wolff, 154–164

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

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

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

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

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

Yanny, B. et al. 2009, AJ, 137, 4377

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

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

Table 1: Observational Constraints and Model Values for the Extinction Curve

Band:	u	g	i	z	J	H	K_s
S82	1.810	1.400	0.759	0.561	0.317	0.200	0.132
North	1.750	1.389	0.750	0.537	0.297	0.180	0.132
OD	1.813	1.406	0.783	0.562	0.325	0.205	0.132
F99	1.795	1.415	0.748	0.554	0.308	0.194	0.132
CCM	1.814	1.394	0.764	0.552	0.321	0.202	0.131

Note. — The first two rows list observational constraints for the shape of the extinction curve, $C_\lambda \equiv A_\lambda/A_r$. The first row corresponds to the so-called SDSS Stripe 82 region (defined by $300° <$ R.A. $< 60°$ and $|Dec| < 1.27°$), and the second row to a northern region defined by $30° < b < 45°$ and $0° < l < 10°$. The last three rows list model predictions computed for an F star spectrum and the best-fit value of R_V (OD=O'Donnell 1994: $R_V = 3.05$; F99=Fitzpatrick 1999: $R_V = 3.30$; CCM=Cardelli et al. 1989: $R_V = 3.01$).

Table 2: Adopted Extinction Coefficients, $C_\lambda(R_V)$

R_V	u	g	i	z	J	H	K_s
2.0	2.314	1.614	0.736	0.466	0.266	0.168	0.108
3.0	1.828	1.412	0.781	0.559	0.300	0.189	0.122
3.1	1.799	1.400	0.784	0.565	0.303	0.191	0.124
4.0	1.601	1.318	0.802	0.603	0.334	0.211	0.136
5.0	1.469	1.263	0.814	0.628	0.366	0.231	0.149

Note. — An illustration of the dependence of the adopted extinction curve, $C_\lambda \equiv A_\lambda/A_r$ on R_V (see also Figure 7).

Table 3: 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
grand total	71,069,183	23,700	23,112,694	12,800

Note. — Needs update!

Table 4: 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. — Needs update!

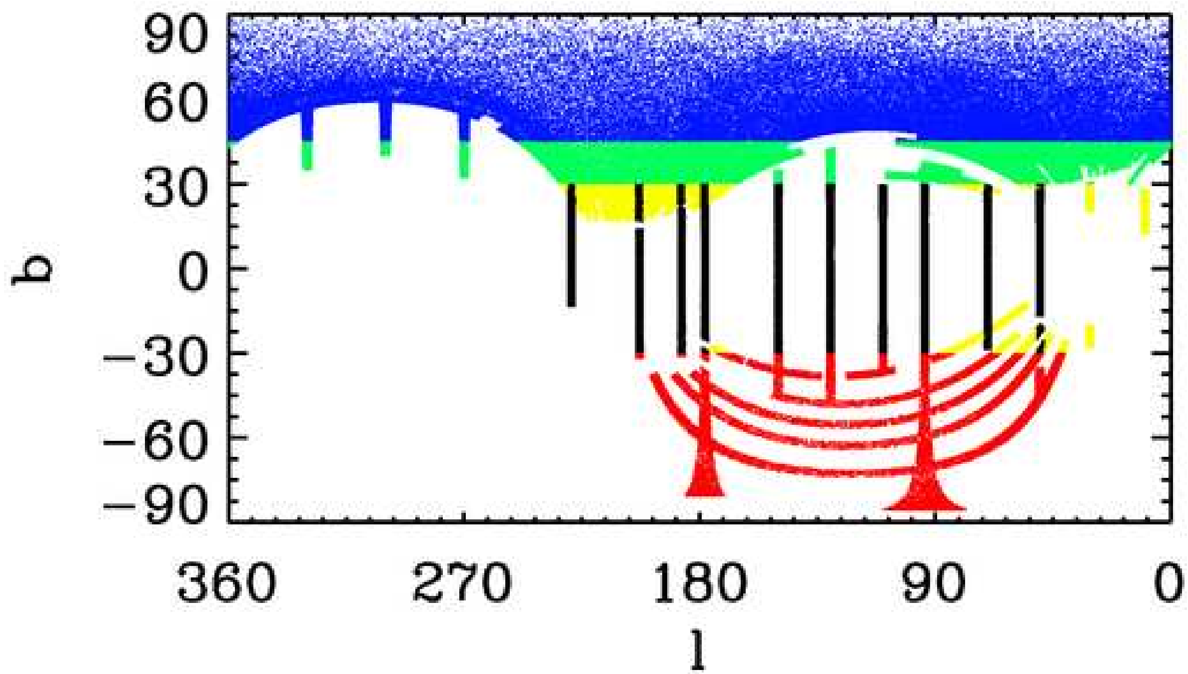


Fig. 1.— The sky coverage for SDSS Data Release 7, used in this study. The points show a small random subsample of the full sample of 73 million stars. The different colors represent the various data file sets (blue, $b > 45$; green, $45 > b > 30$; black, the 10 SEGUE strips; yellow $|b| < 30$, stars not in SEGUE strips; and red, $b < -30$).

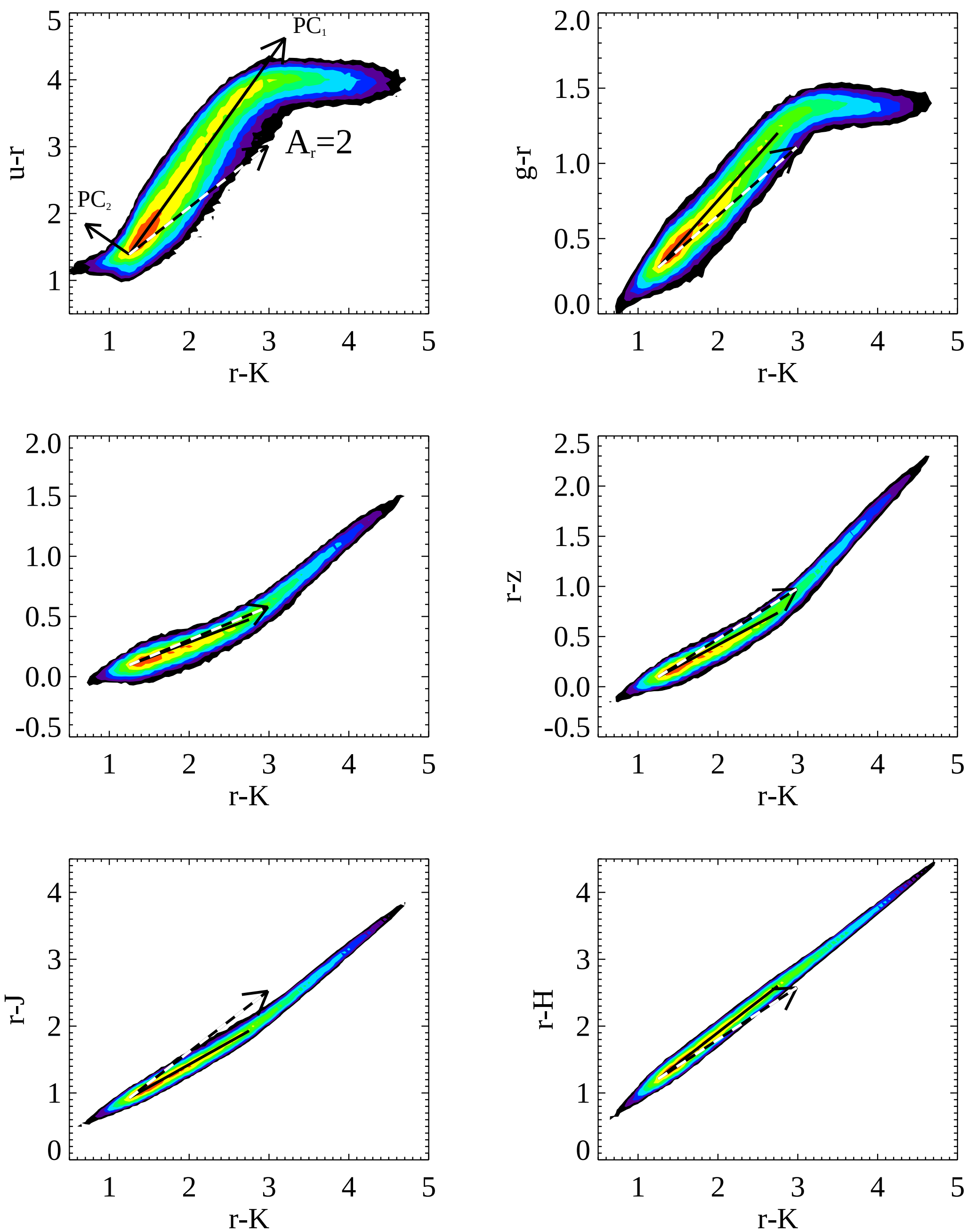


Fig. 2.— The distribution of unresolved SDSS sources with 2MASS detections in the $m - r$ vs. $r - K$ color-color diagrams, with $m = u, g, i, z, J$ and H. The source density is shown as color-coded maps, and it increases from black to green to red. The two arrows marked PC$_1$ and PC$_2$ in the top left panel illustrate the “principal color” axes discussed in text and used to track the locus shifts due to interstellar dust reddening. The dashed vector in each panel shows the reddening vector for $A_r = 2$ and standard $R_V = 3.1$ dust (**?**).

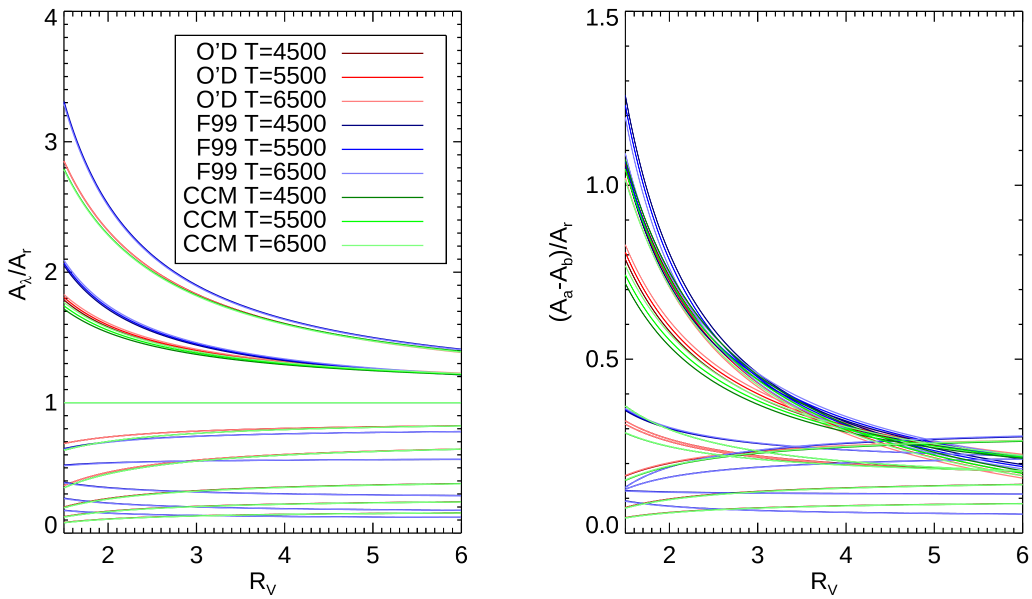


Fig. 3.— Model predictions for the extinction curve shape as a function of R_V for three different models: O'D (**?**), F99 (**?**), and CCM (**?**), evaluated for stars with three different effective temperatures (as listed in the legend, in K). The left panel shows $C_\lambda = A_\lambda/A_r$ for $\lambda = (u, g, r, i, z, J, H, K)$ (top to bottom, respectively) and the right panel is analogous, except that the ratios based on colors ($u-g$, $g-r$, $r-i$, $i-z$, $z-J$, $J-H$, and $H-K$) are shown. Note that most of the sensitivity to R_V comes from the blue bands (u and g).

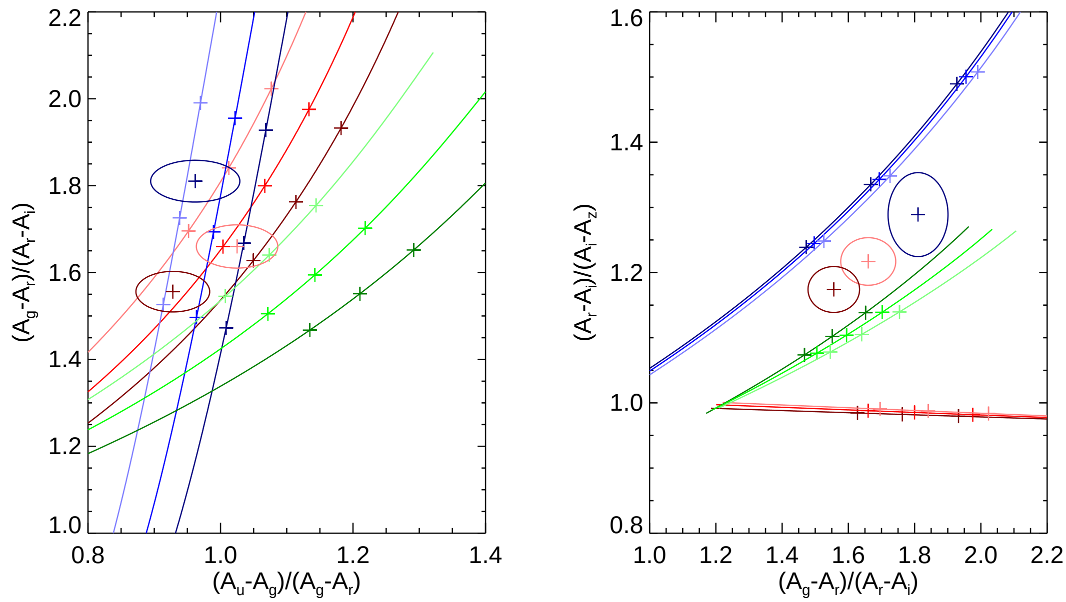


Fig. 4.— A comparison of the constraints on the extinction curve shape (the three plus symbols, with approximate $1-\sigma$ uncertainty limits shown as ellipses) and three model predictions (see Figure 3 for legend; the three crosses along the curves correspond to R_V=2.6, 3.1, and 3.6). The brown symbol corresponds to the Stripe 82 region (southern galactic hemisphere), the pink symbol to the northern galactic hemisphere, and the blue symbol is the constraint from the Schalfly et al. (2010) analysis. The F99 (**?**) and CCM (**?**) models are in fair agreeement with the data, while the OD model predicts unacceptable values of the $(A_r - A_i)/(A_i - A_z)$ ratio for all values of R_V (see also Figure 5).

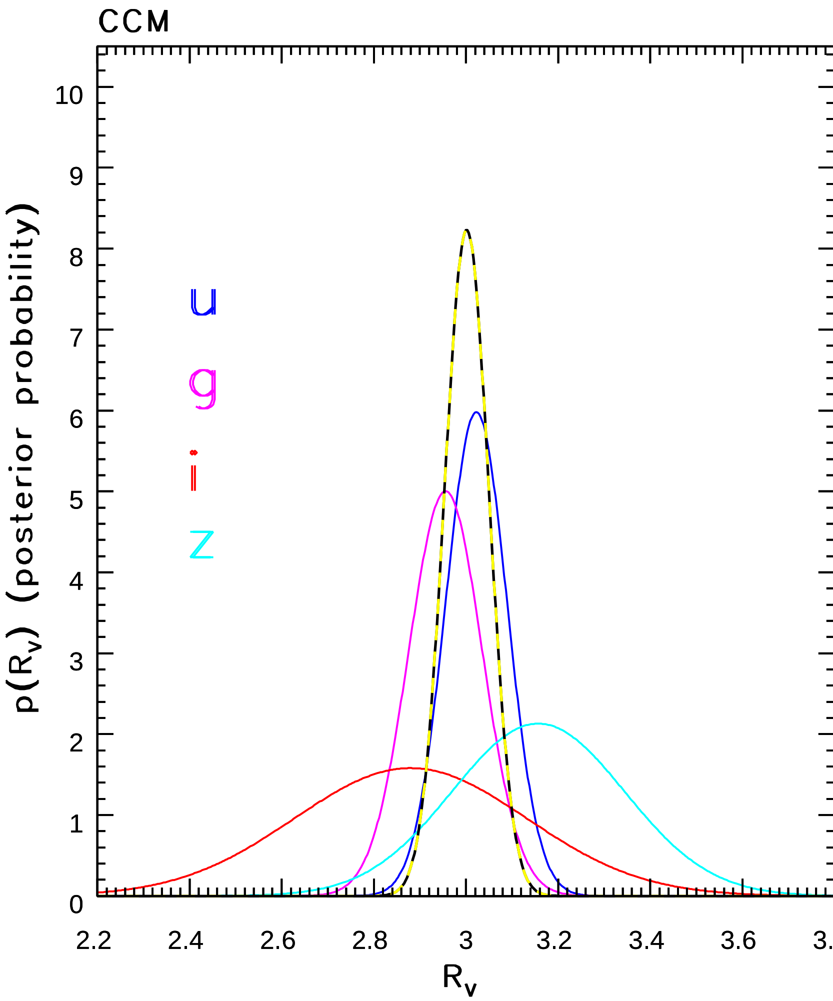


Fig. 5.— Constraints on R_V based on the CCM (**?**) dust reddening law. Only the SDSS bands, which provide the strongest constraints on R_V are shown (see the legend). The dashed line shows the overall constraint on R_V (posterior probability distribution for a flat prior), with the best-fit value of $R_V = 3.01 \pm 0.05$.

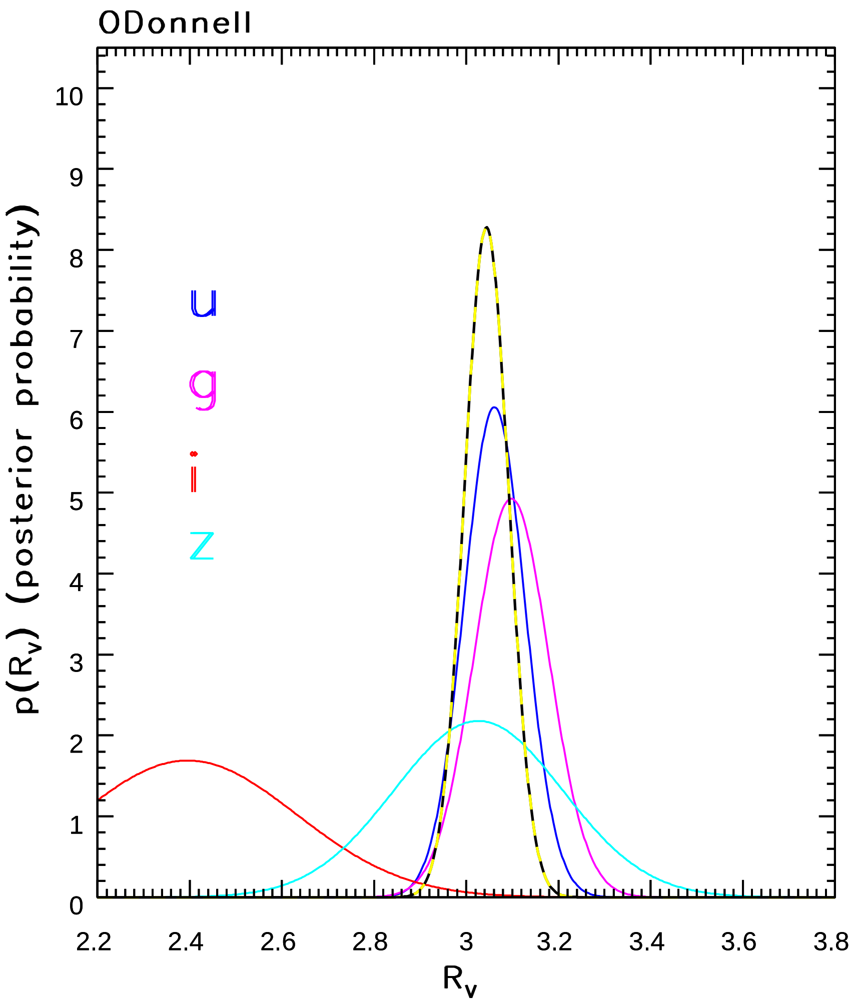


Fig. 6.— Analogous to Figure 5, except that O'Donnell (1994) dust reddening law is used. Note that the predicted extinction in the i band is inconsistent with constraints from other bands.

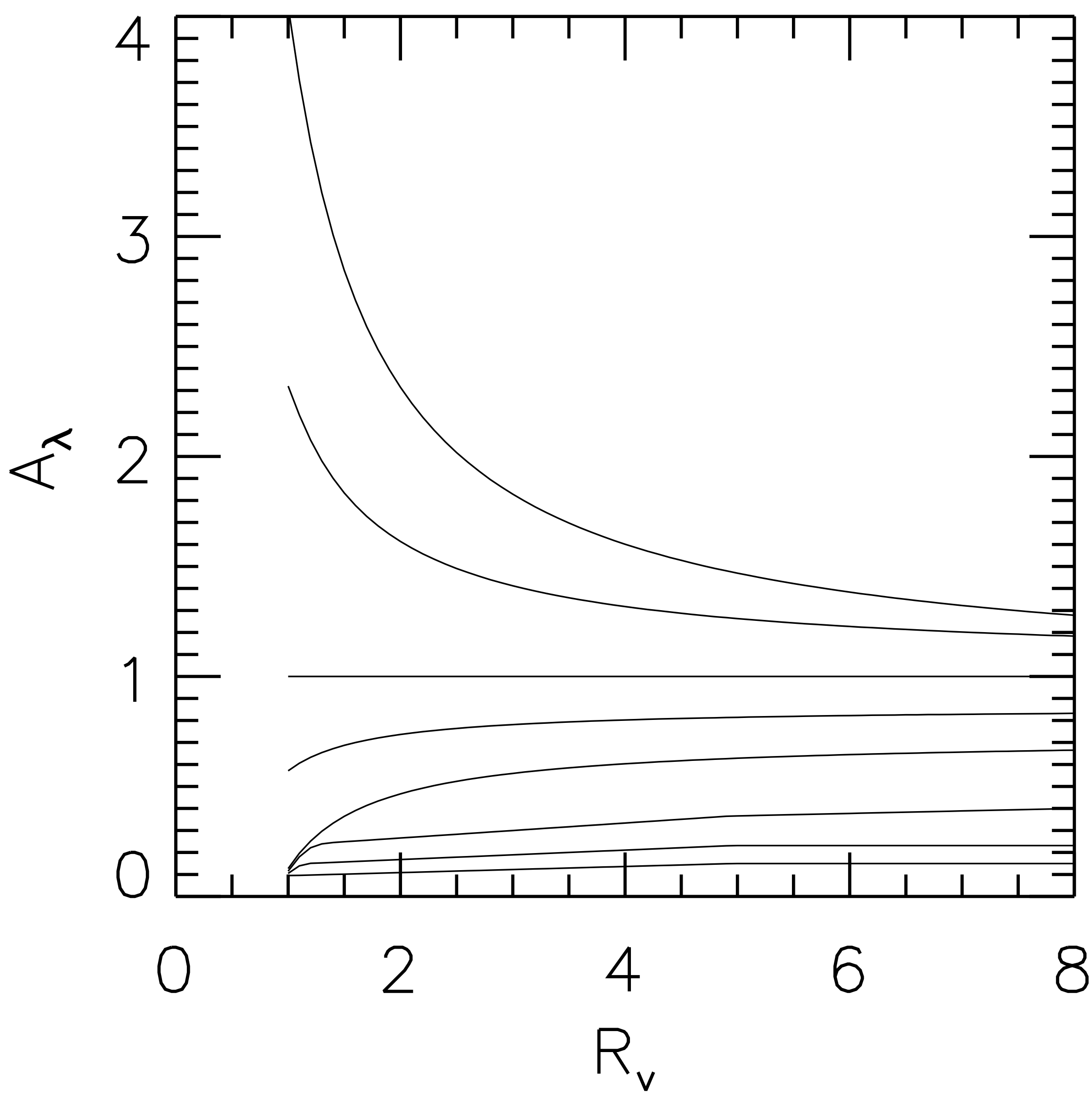


Fig. 7.— The adopted A_λ/A_r ratio, shown as a function of R_V, for $\lambda = (ugrizJHK)$, top to bottom. The curves are computed for an F star using the CCM (**?**) dust reddening law.

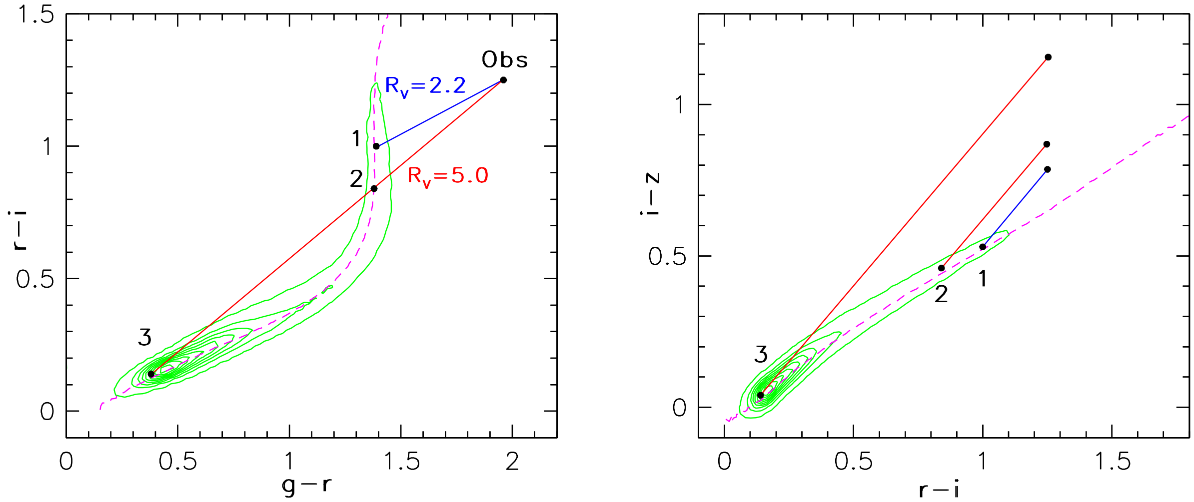


Fig. 8.— An illustration of the constraints on intrinsic stellar colors, extinction in the r band, A_r, and the ratio of total to selective extinction, R_V. In both diagrams, the linearly-spaced contours show the main stellar locus as observed at high galactic latitudes. The dashed lines mark the median stellar locus from Covey et al. (2007). In the left panel, the dot marked "Obs" represents a hypothetical observation. Depending on the 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 which breaks the degeneracy between the intrinsic stellar color and A_r. The degeneracy is broken because the reddening vectors in the right panel are nearly parallel despite very different R_V values.

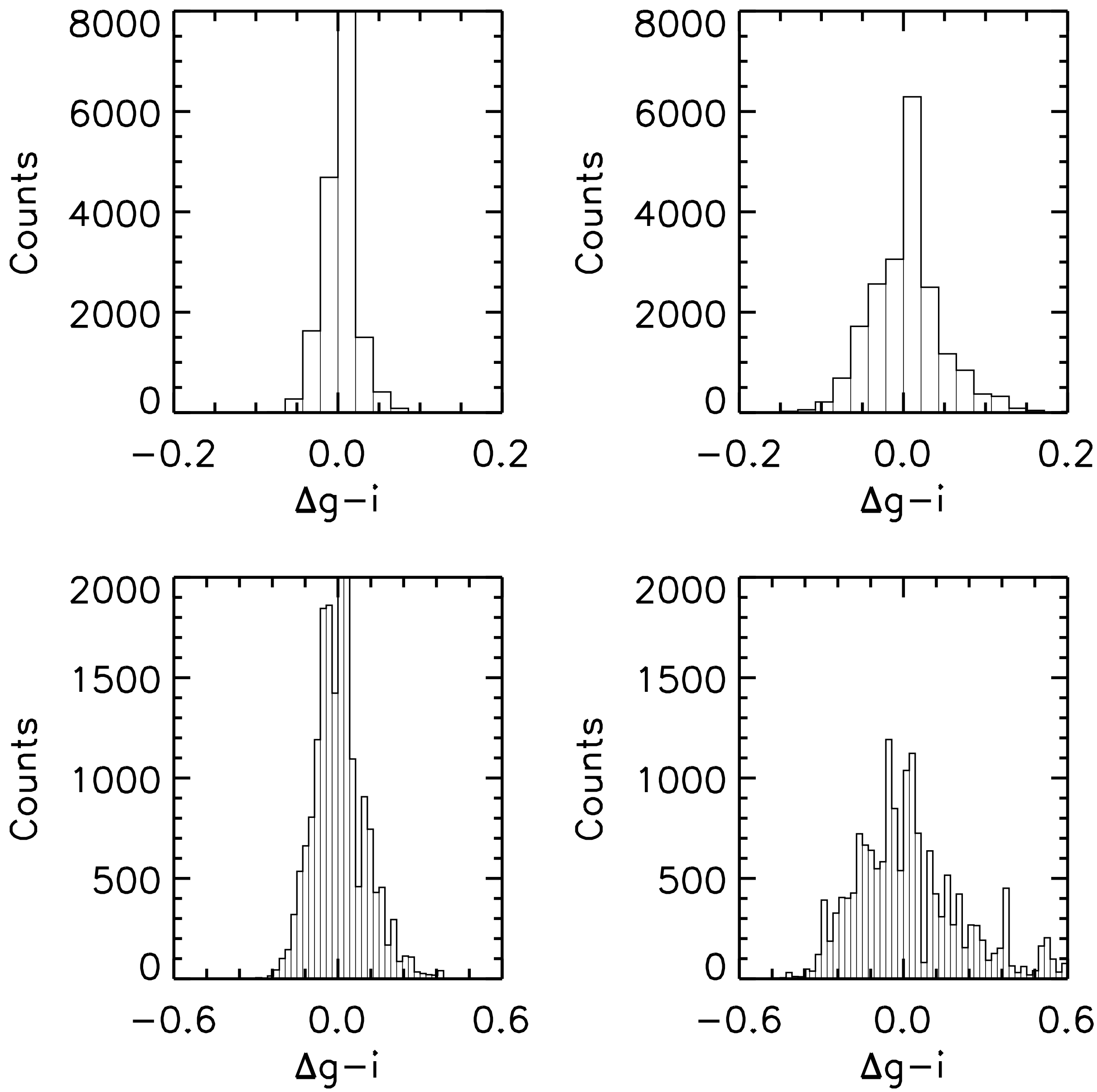


Fig. 9.— A Monte Carlo study of best-fit stellar model errors (parametrized by the $g-i$ color) as a function of photometric errors, for a fiducial star with $g-i$=1.95 and A_r=1.5. The photometric errors are generated from gaussian distributions with widths equal to 0.01 mag (top left), 0.02 mag (top right), 0.04 mag (bottom left) and 0.08 mag (bottom right). The errors in the best-fit $g-i$ are about twice as large as assumed photometric errors.

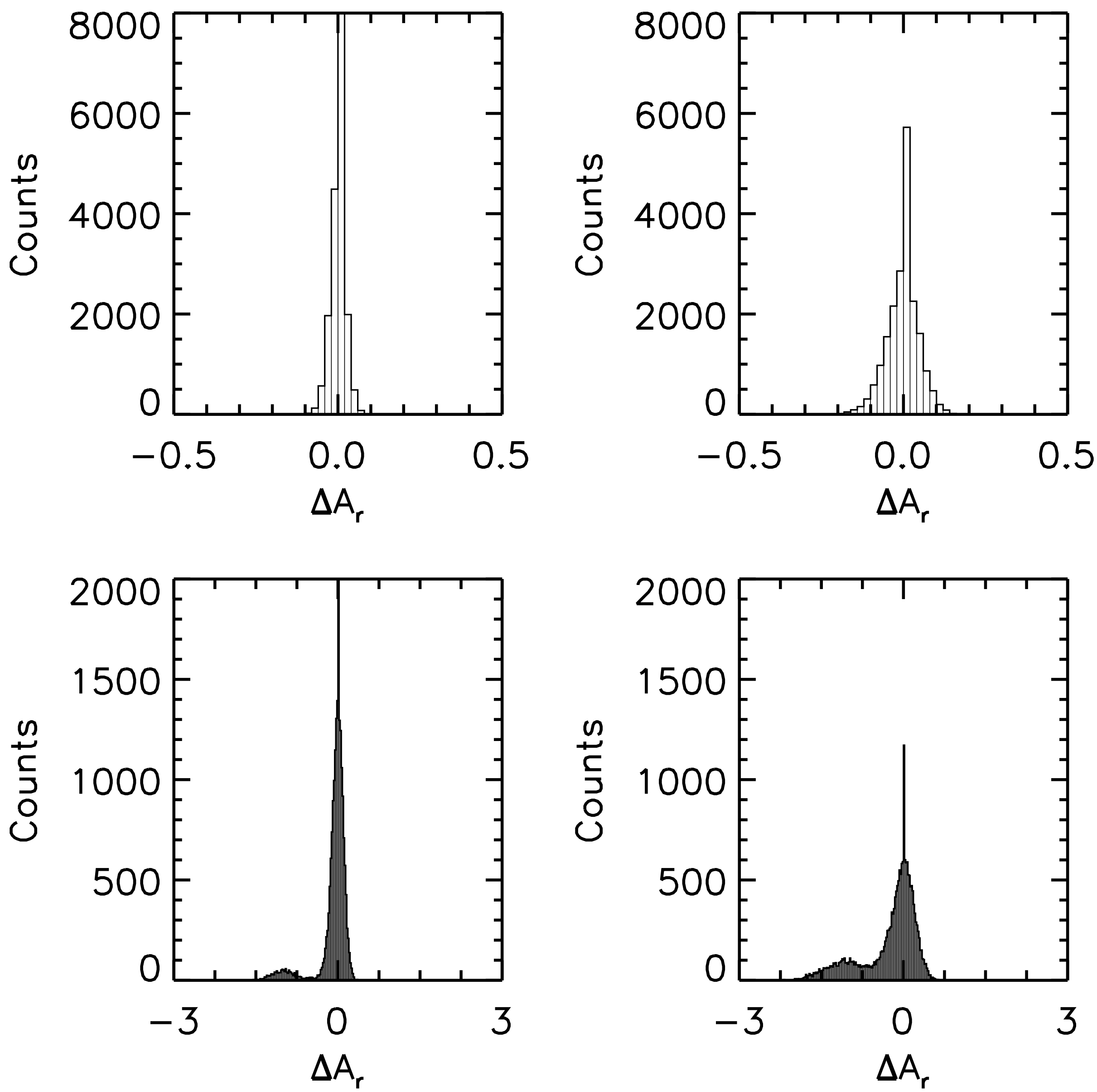


Fig. 10.— Analogous to Fig. 9, except that the errors in the best-fit A_r are shown. Note that for large photometric errors (the bottom two panels), the A_r error distribution becomes bimodal.

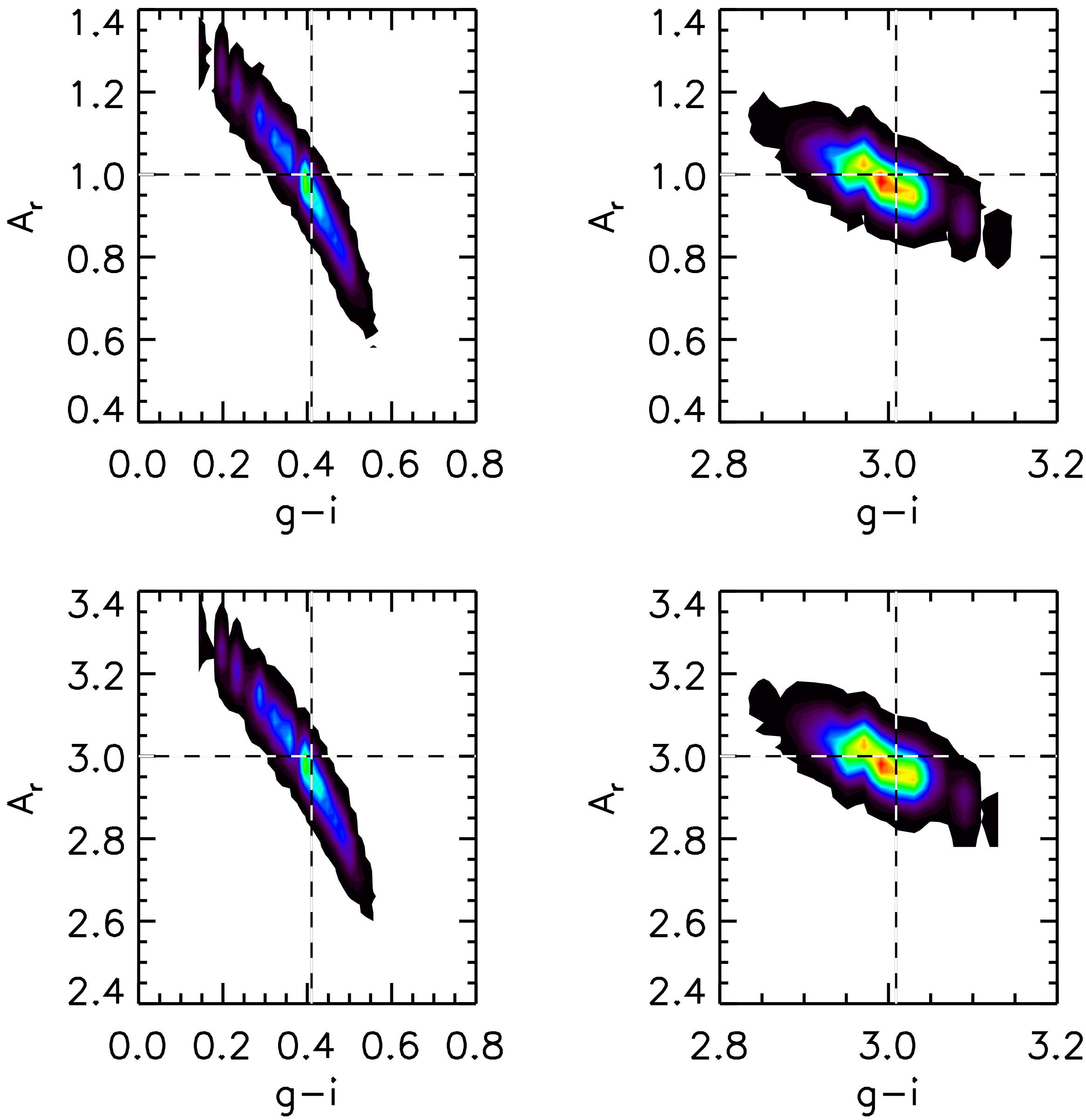


Fig. 11.— Analysis of the covariance in the best-fit values for A_r and $g-i$ using a simulated dataset. The panels show the distributions of the best-fit values for A_r and $g-i$ for two different fiducial stars (left column: a blue star with true $g-i$=0.4; right column: a red star with true $g-i$=3.0), 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, and does not strongly depend on assumed A_r.

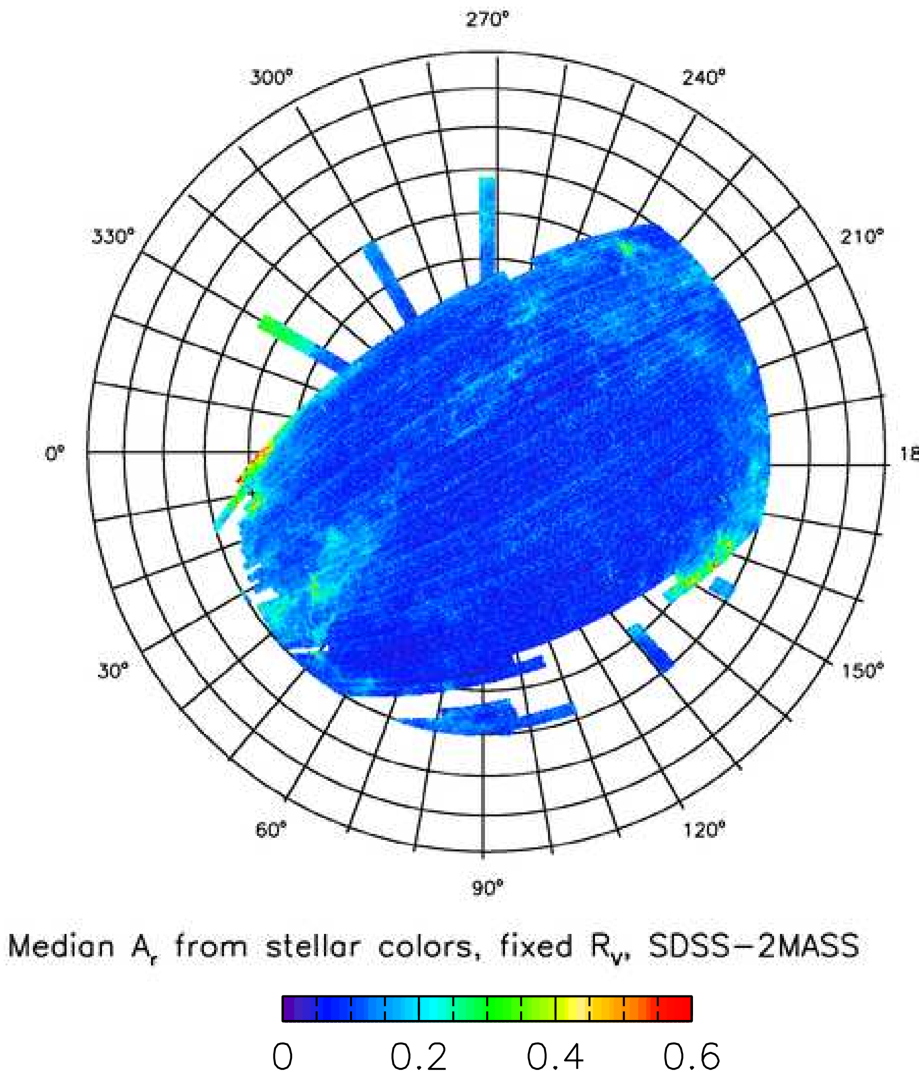


Fig. 12.— The color-coded maps shows the median best-fit A_r based on SDSS and 2MASS data for 12 million stars with $15 < r < 20$ from the northern galactic cap ($|b| > 30°$). The values are linearly color-coded according to the legend. The fitting was done with a fixed extinction curve, and only the stellar SEDs and the amount of dust (A_r) are varied when fitting the observed colors.

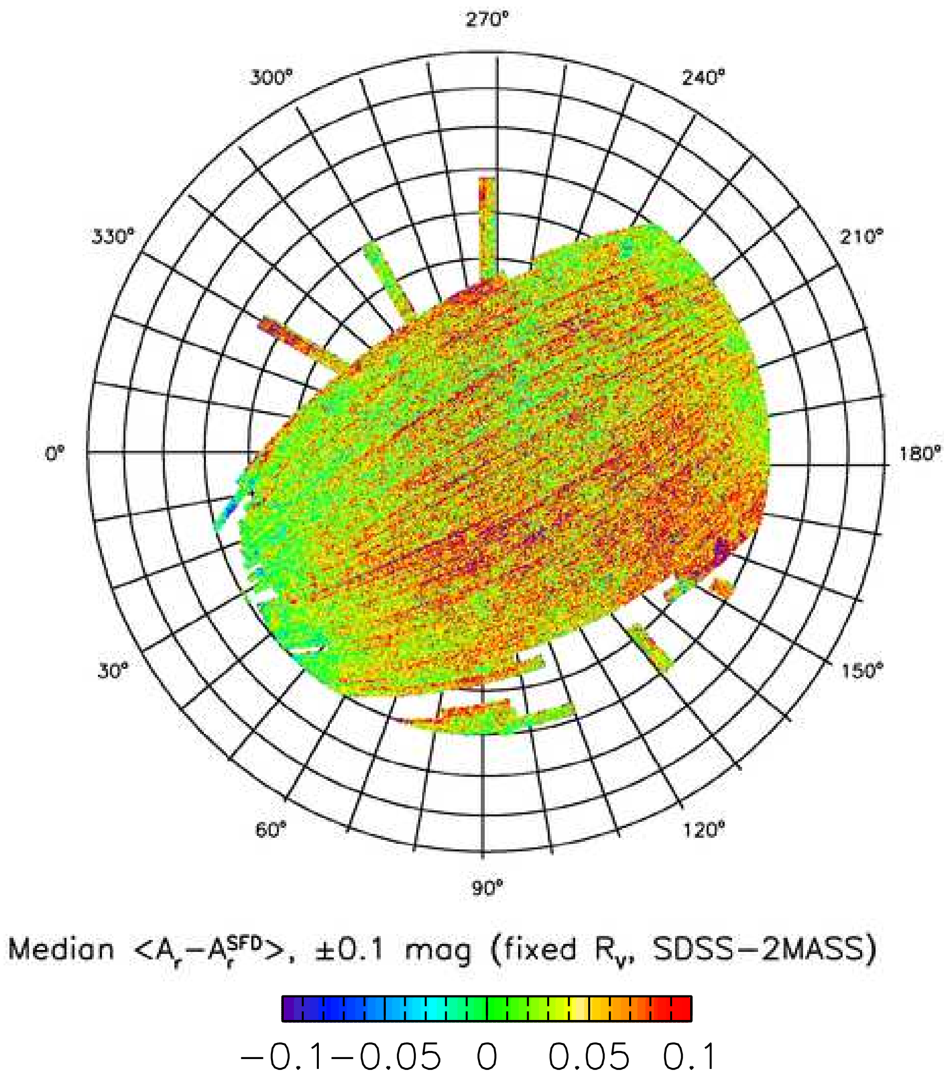


Fig. 13.— The color-coded maps shows the median difference between the best-fit A_r shown in Figure 13 and the values given by the SFD map at the position of each star. The values are linearly color-coded according to the legend. Note the striping that is reminiscent of the SDSS scanning pattern. These coherent residuals imply problems with the transfer of SDSS photometric zeropoints across the sky.

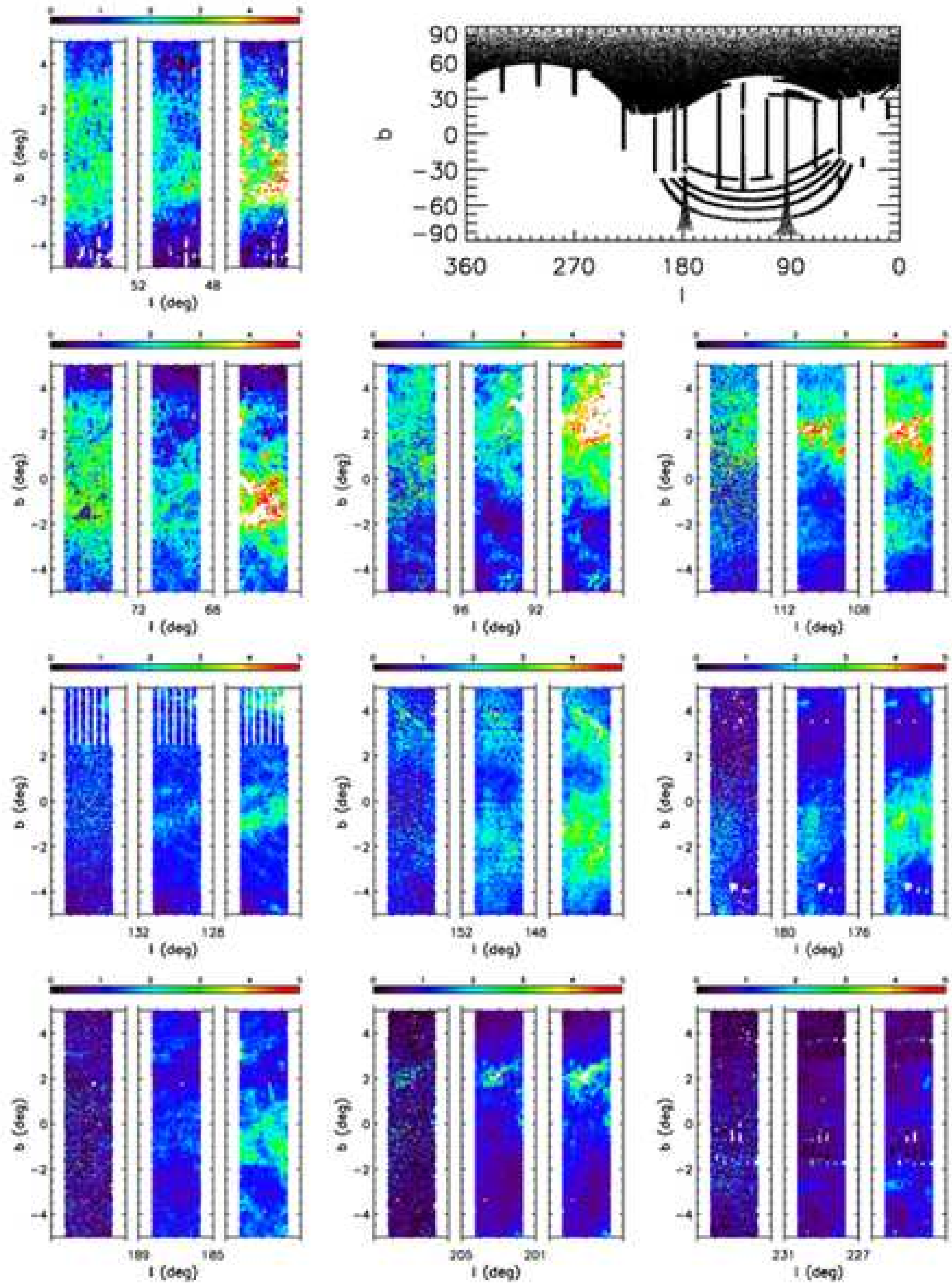


Fig. 14.— The color-coded maps show the best-fit A_r based on SDSS data for the ten analyzed SEGUE stripes. A fixed $R_V = 3.1$ is assumed. The legend above each panel shows the color scale, and each 4x4 arcmin^2 pixel shows the median A_r. For each stripe, three distance ranges are shown: 0.3-0.6 kpc (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. Each stripe is limited to the range of galactic latitude, $|b| < 5°$. The top right panel shows the sky coverage of the full analyzed dataset.

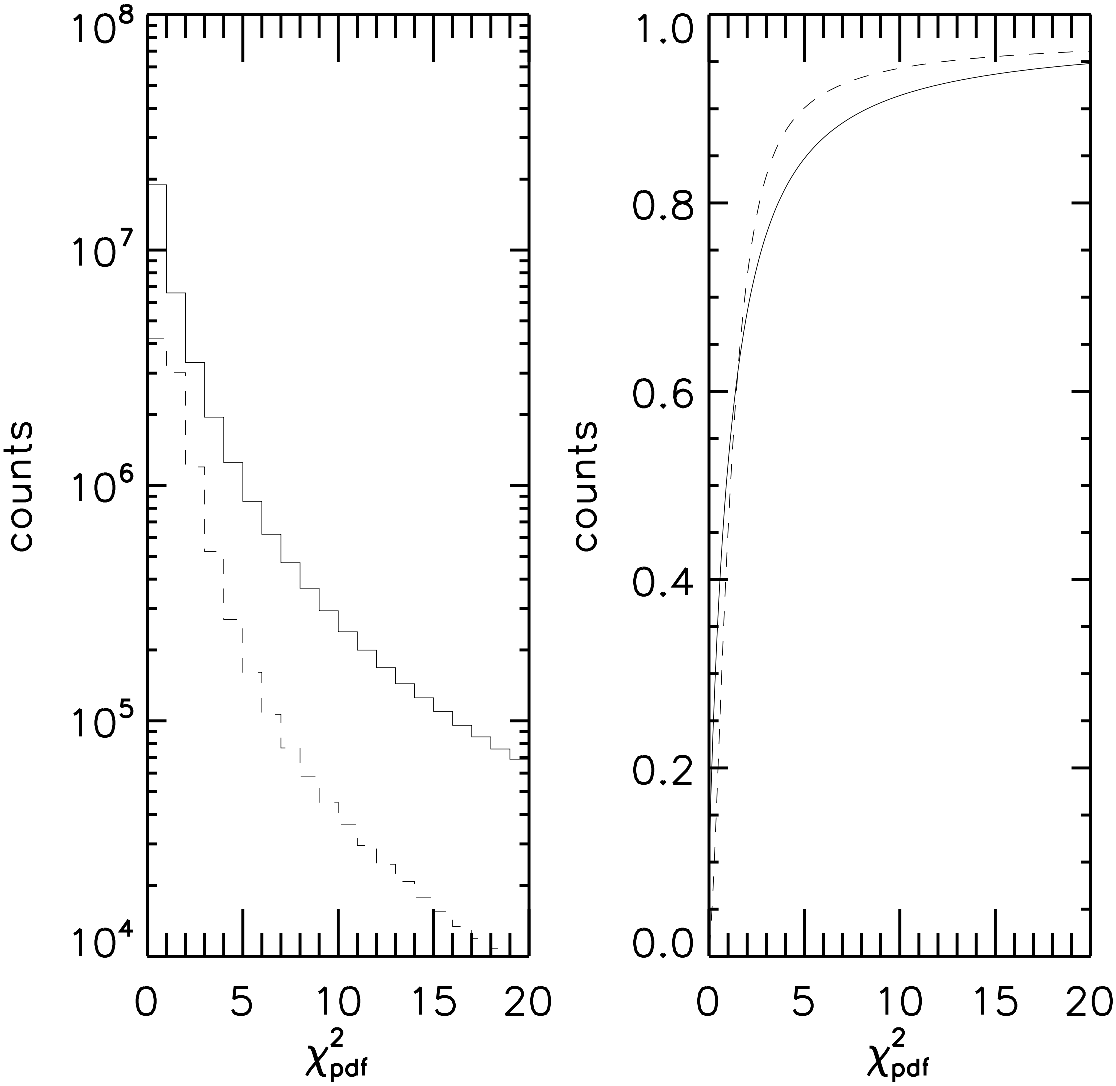


Fig. 15.— The distribution of the best-fit $\chi^2_{\rm pdf}$ for SDSS-only fits (solid line) and SDSS-2MASS fits (dashed line) are plotted in the left panel. The right panel displays the cumulative $\chi^2_{\rm pdf}$ for the SDSS-only and SDSS-2MASS fits.

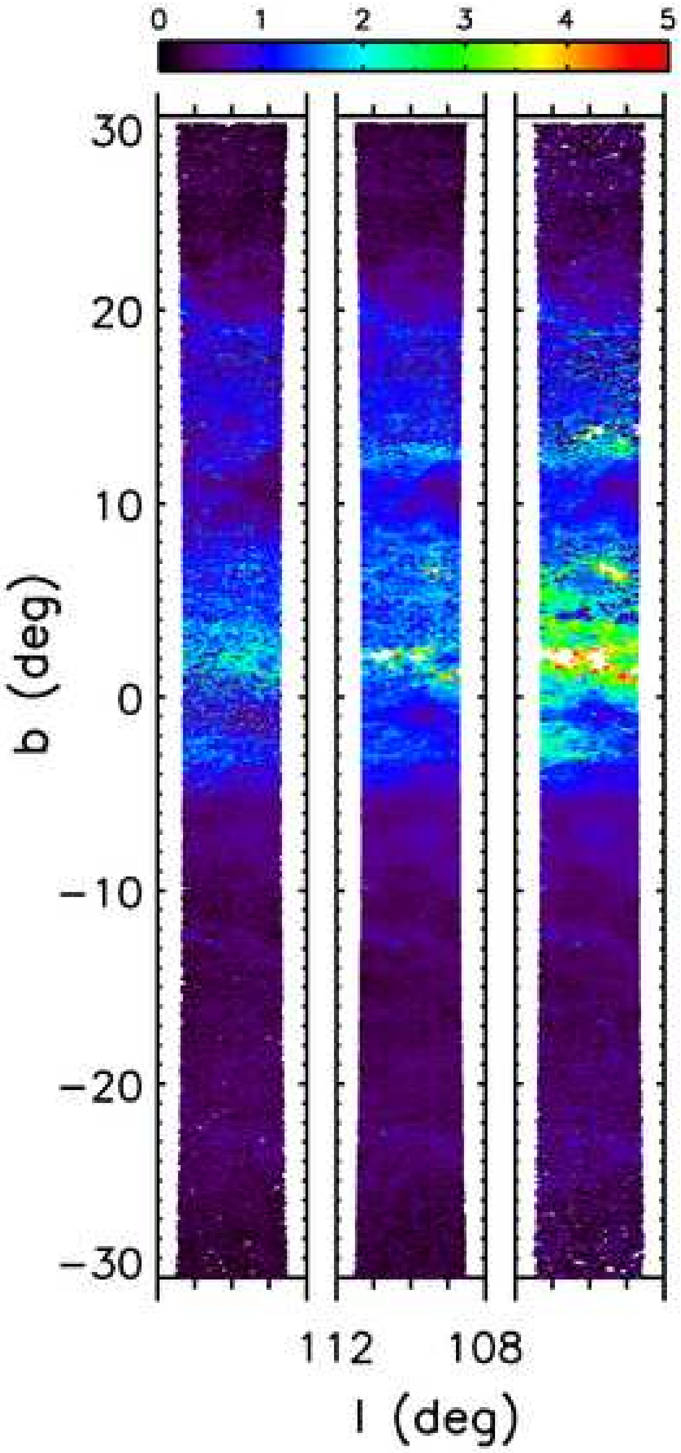


Fig. 16.— Analogous to Fig. 14, except that only the $l \sim 110°$ slice is shown. Note that A_r increases with distance.

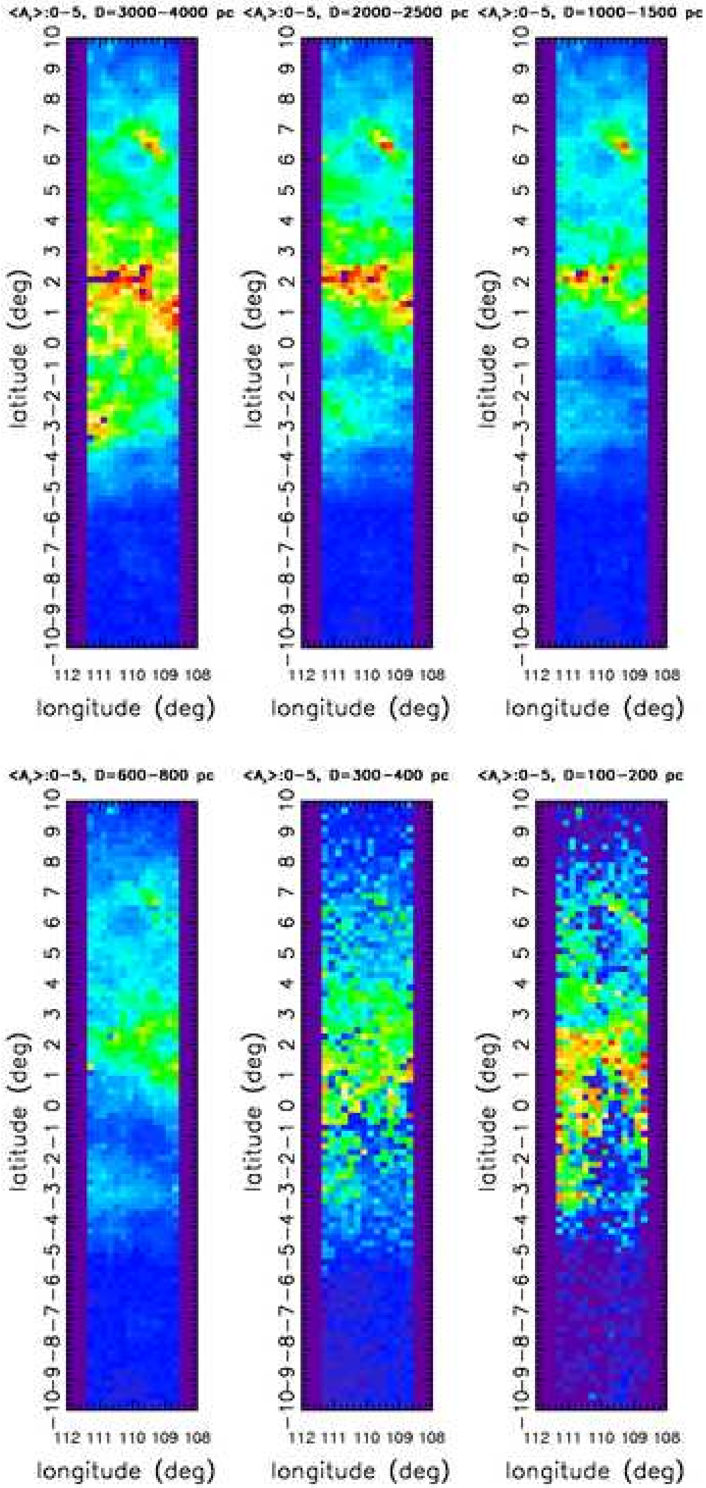


Fig. 17.— **ZI: redo this figure with public data, and for D slices 100-200, 200-300, 300-400, 2000-2500 and 3000-4000.** Analogous to Figure 16, except for different distance slices. The high extinction values in the closest slice (bottom right panel) are most likely due to distant giants with underestimated distances.

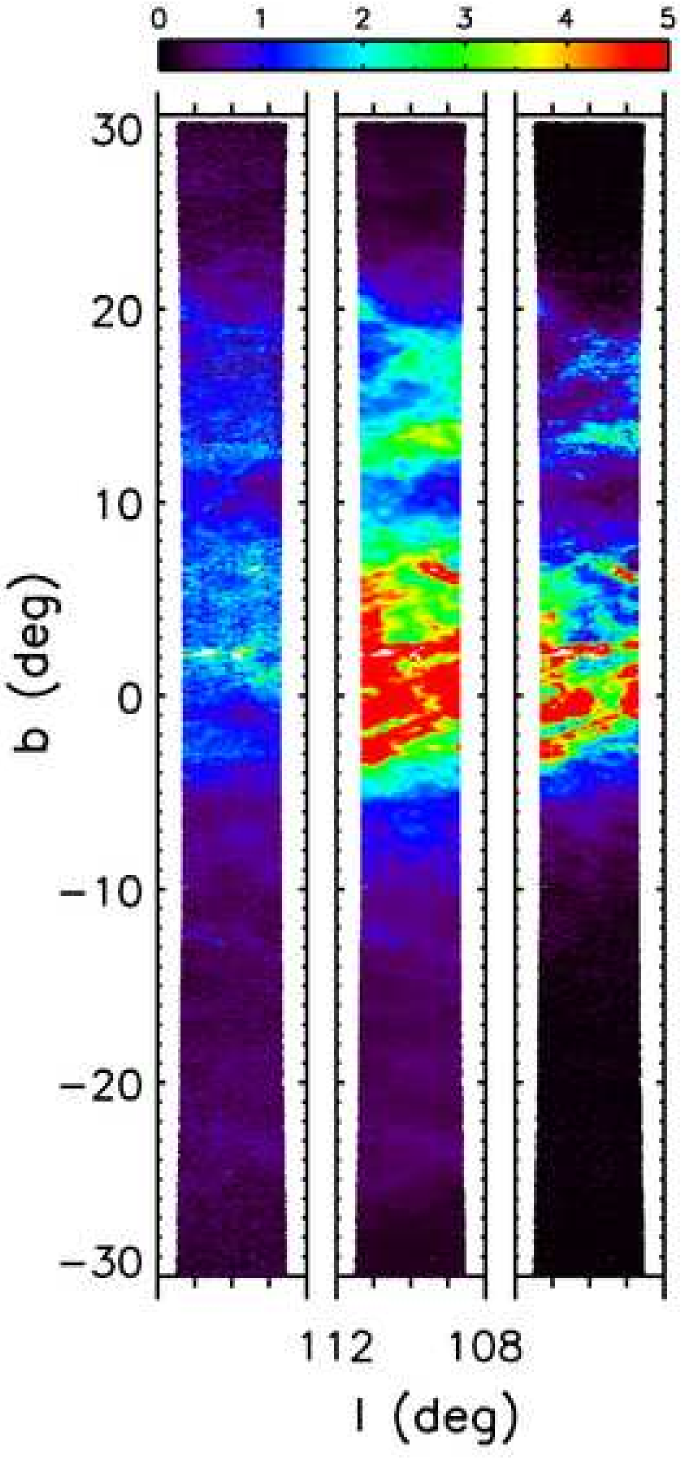


Fig. 18.— Analysis of the differences between best-fit A_r values (left panel, based on SDSS data) and the SFD values (middle panel). Their difference is shown in the right panel. The legend on top shows the color scale, and each pixel shows the median value for stars with good fits ($\chi^2_{\rm pdf} < 3$), and distance in the range 0.8-1.2 kpc. If the SFD maps are correct, then the structure seen in the right panel must be more distant than $\sim$1 kpc.

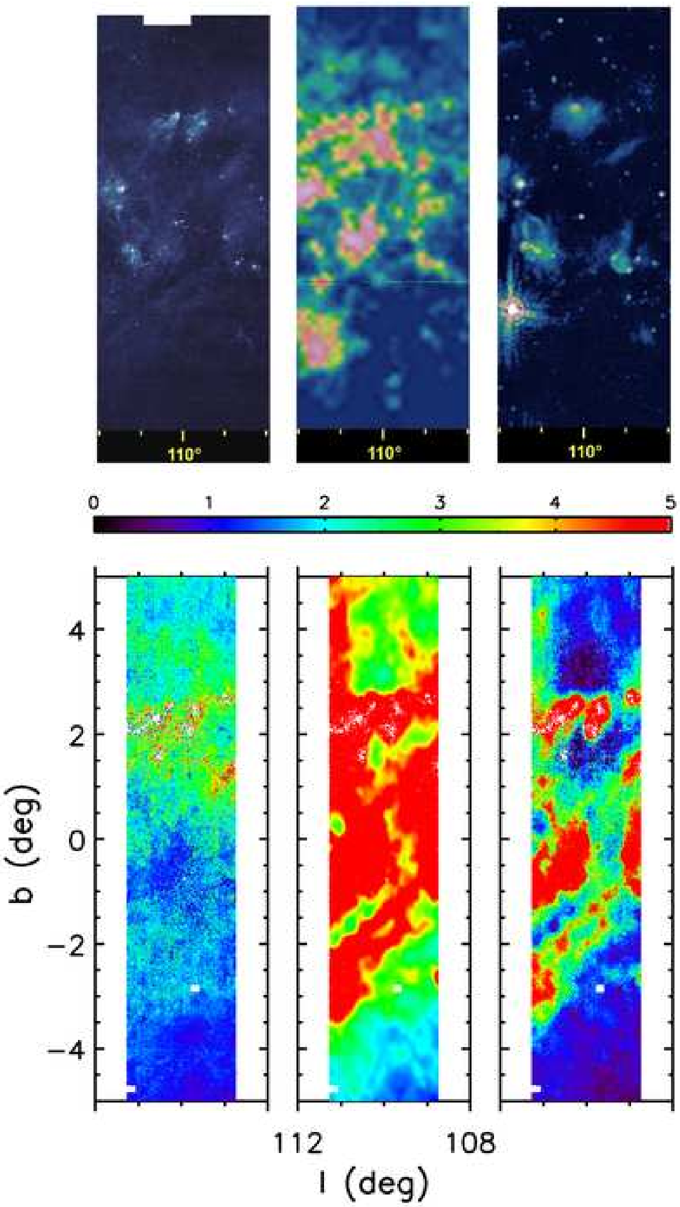


Fig. 19.— The bottom three panels show the $|b| < 5°$ subregion of the panels shown in Fig. 18. The top three panels show the mid-IR (left), CO (middle) and radio continuum (right) maps on approximately the same scale (XXX add references). Assuming that the SFD map is not grossly incorrect, the dust extinction determined here implies that most of the molecular cloud structures seen in the top middle panel must be more distant than $\sim$1 kpc.

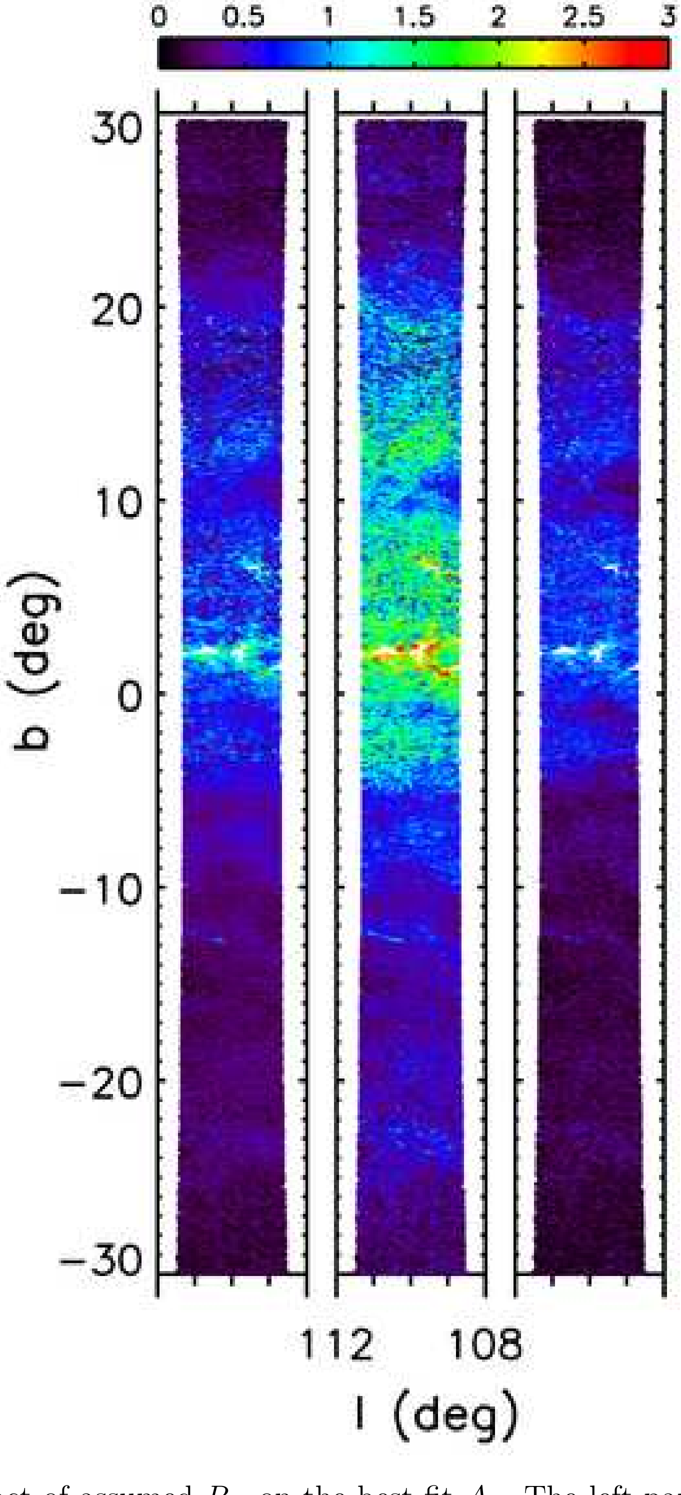


Fig. 20.— The impact of assumed R_V on the best-fit A_r. The left panel shows the median best-fit A_r for $R_V = 2$, and the middle panel for $R_V = 4$ (only SDSS data and stars with $\chi^2_{\rm pdf} < 3$ are used). The difference between the two maps is shown in the right panel.

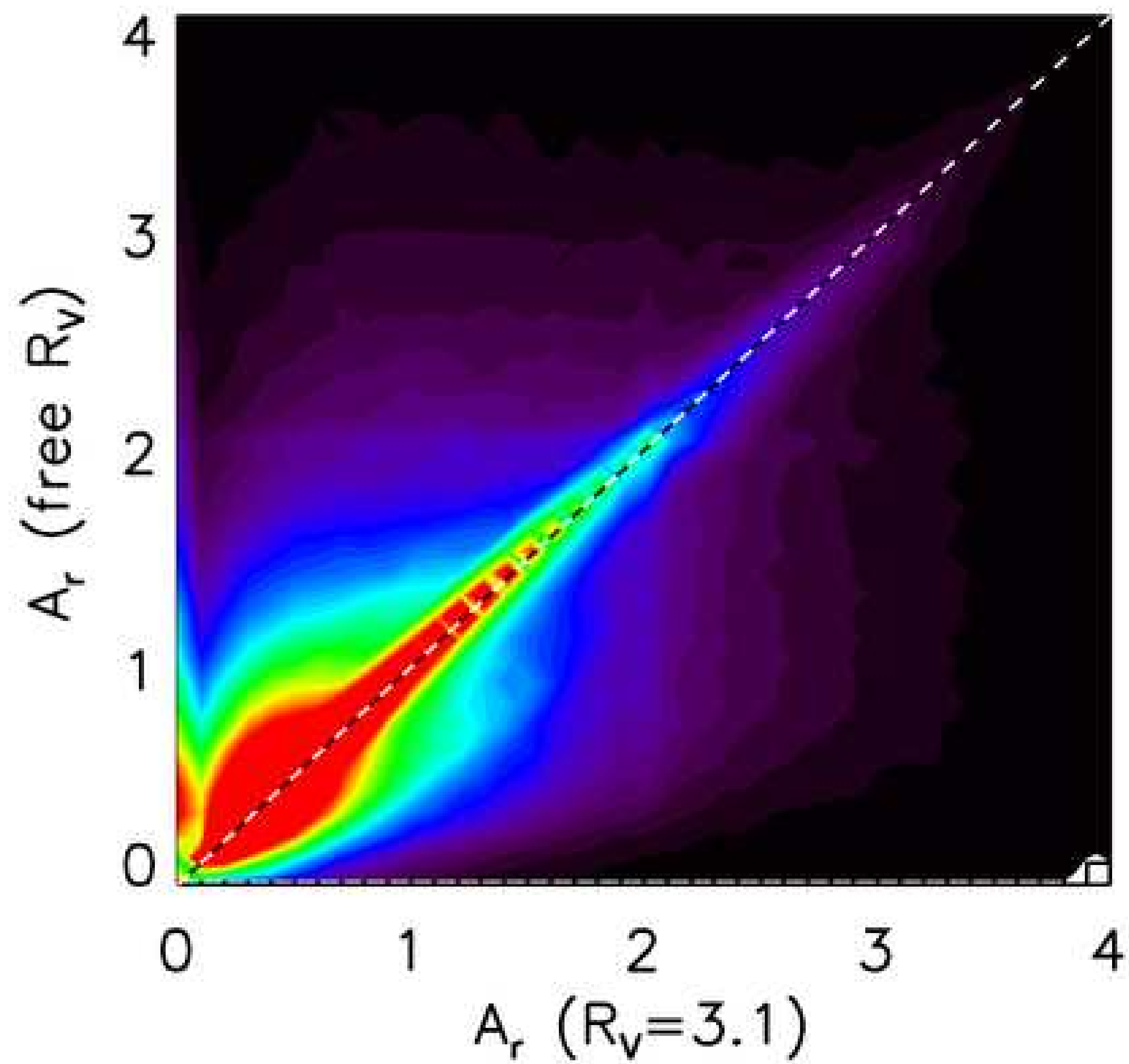


Fig. 21.— A comparison of the best-fit A_r values for two different treatments of R_V, for stars in the $l = 110$ SEGUE strip, and using only SDSS data. The x axis shows the A_r values obtained for a fixed $R_V = 3.1$ and the y axis for a variable R_V with a best-fit R_V of $1.5 < R_V < 5.0$. The number density of stars increases from black to blue to red.

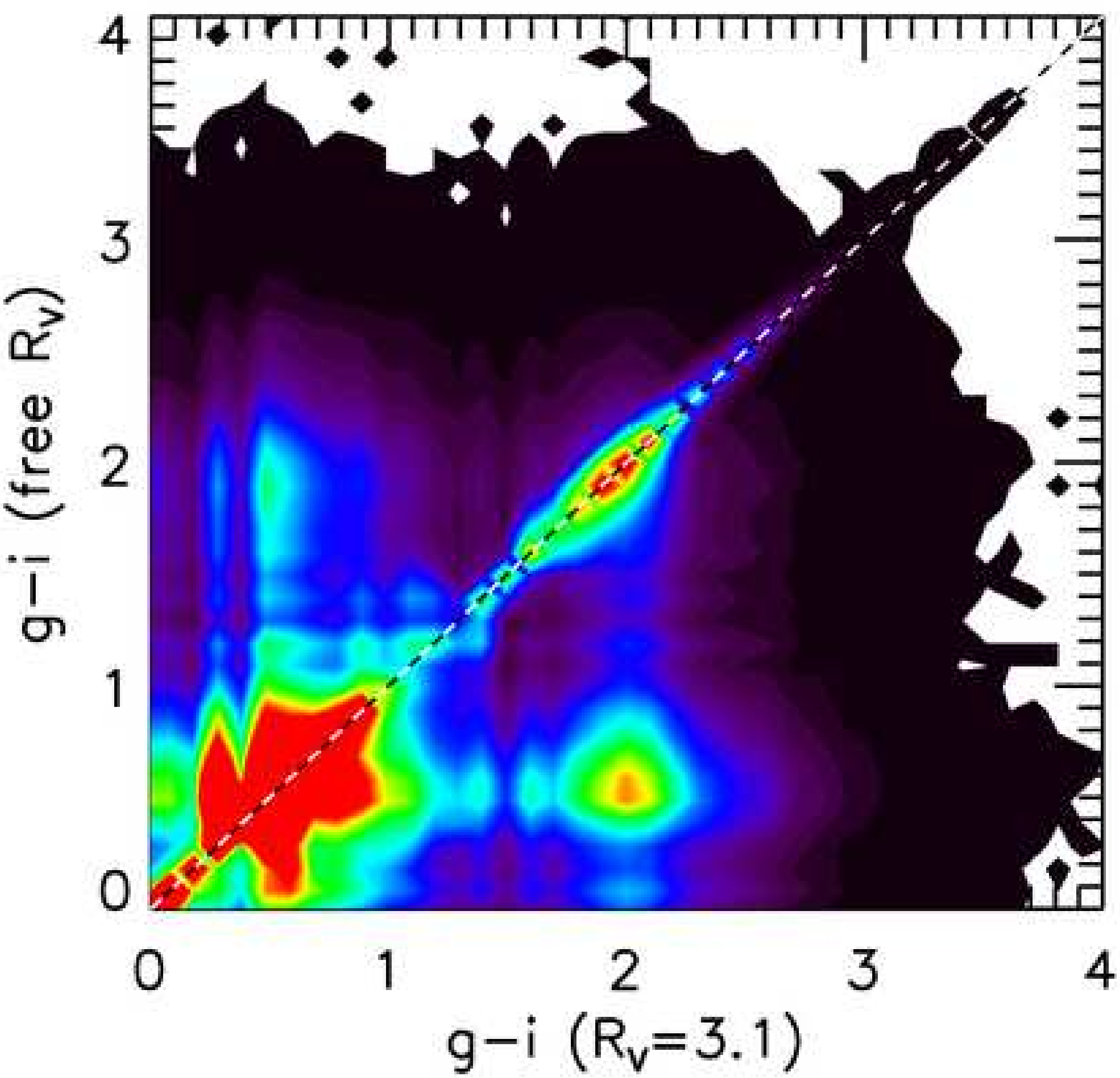


Fig. 22.— Analogous to Fig. 21, except that the best-fit $g - i$ values are compared.

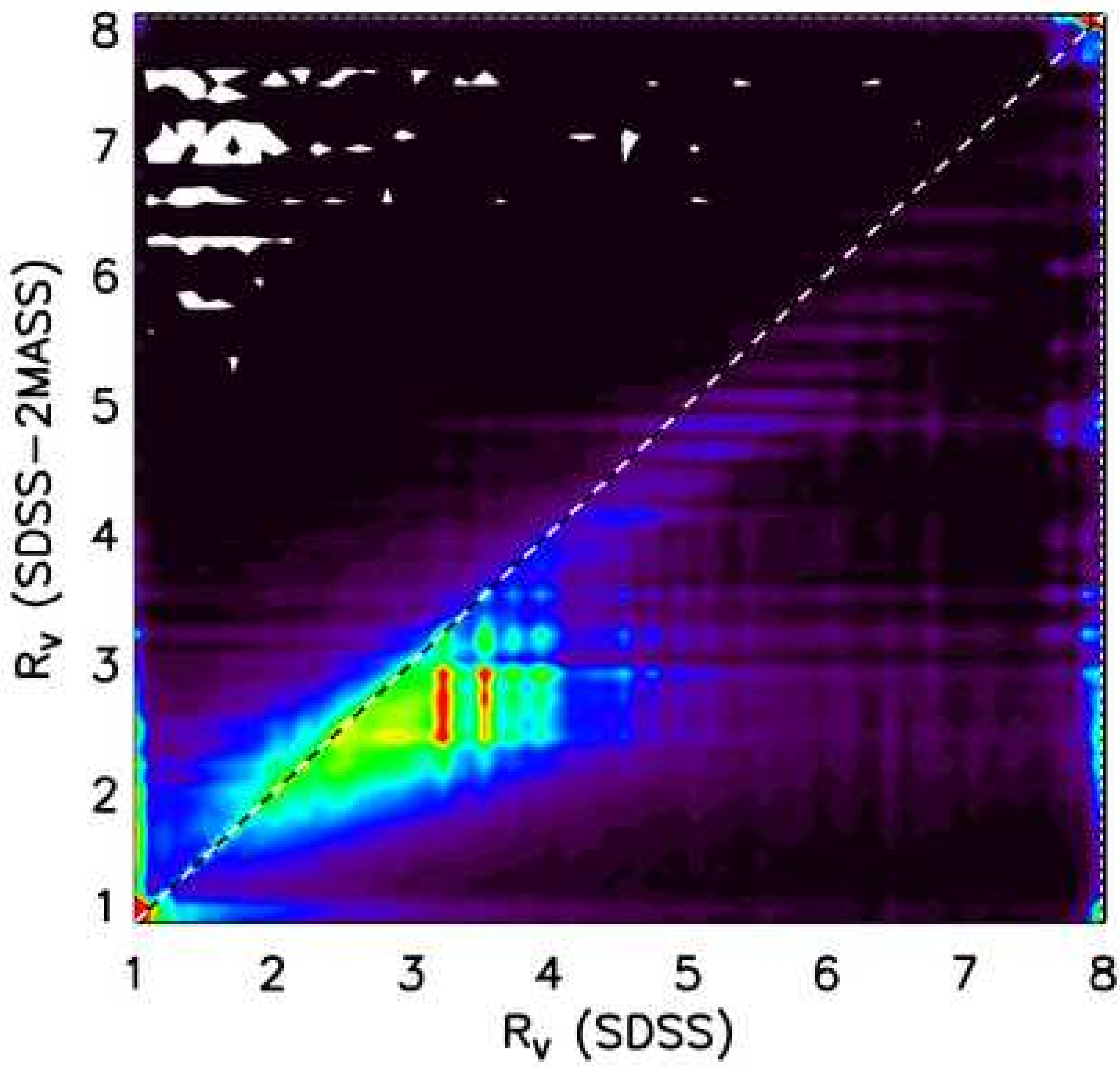


Fig. 23.— Analogous to Fig. 21, except that the best-fit R_V values are compared with SDSS fits on the x axis and SDSS-2MASS fits on the y axis.

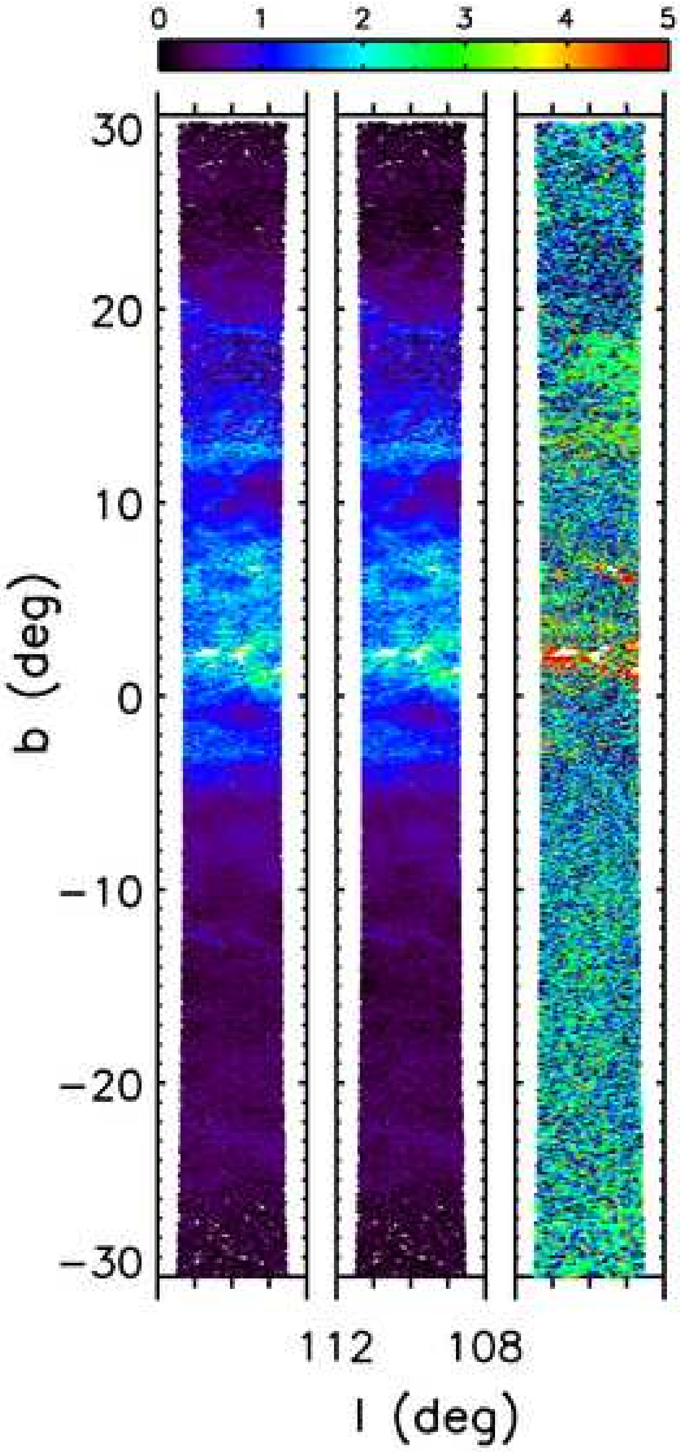


Fig. 24.— Analysis of the differences in A_r (right panel) between fits based only on SDSS data (left panel) and fits based on both SDSS and 2MASS data (middle panel). The legend on top shows the color-code for the left and middle panels, and each pixel shows the median A_r for stars with $\chi^2_{\rm pdf} < 3$. 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 (A_r(SDSS-2MASS)-A_r(SDSS)).

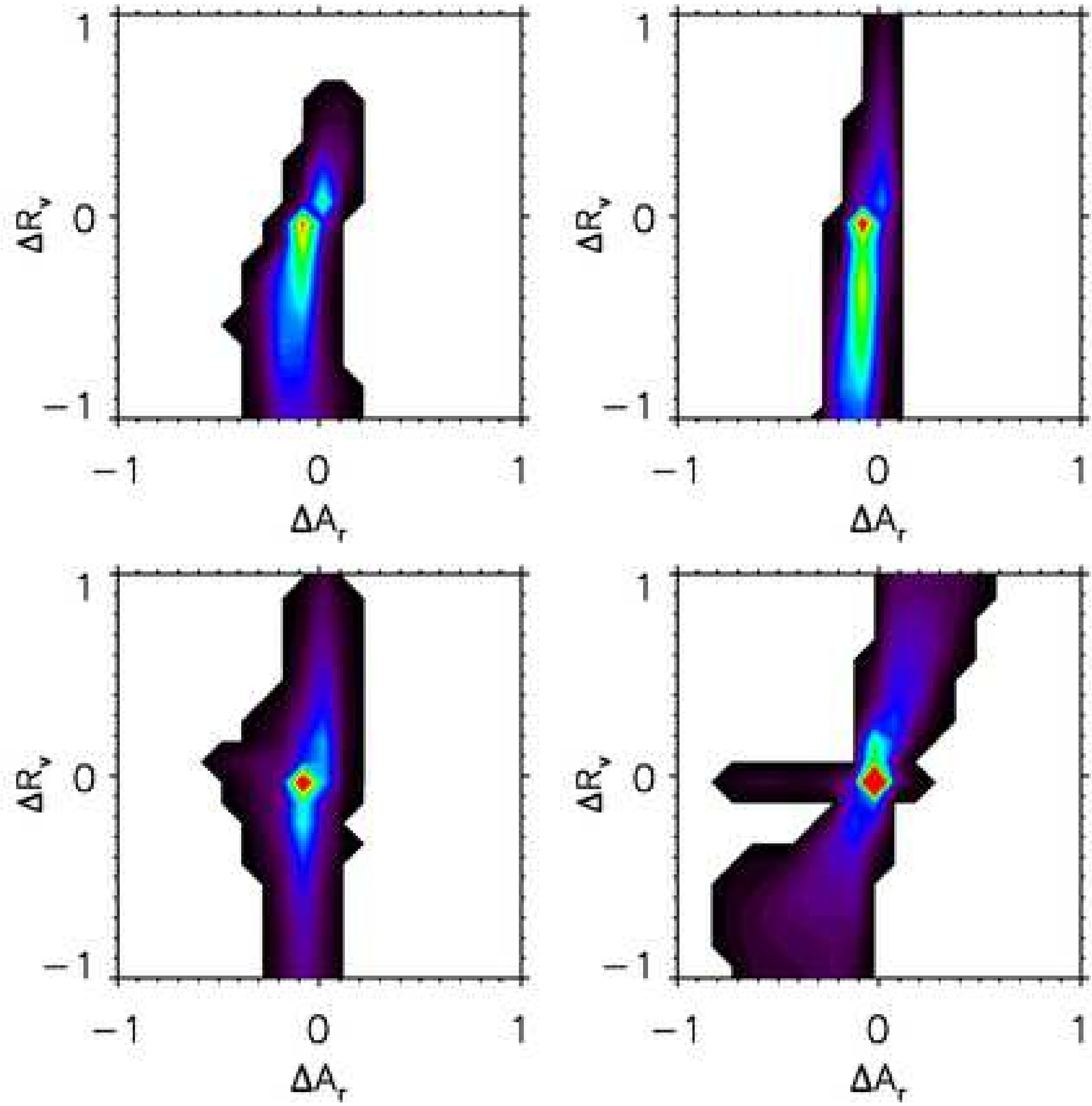


Fig. 25.— A comparison of the best-fit A_r and R_V values for fits based on only SDSS and both SDSS and 2MASS data. The contour plots show the distribution of the values of ΔA_r = A_r(SDSS-2MASS)-A_r(SDSS), and analogously for R_V. The four panels correspond to four quartiles of the best-fit $g-i$ color, increasing $g-i$ from lowest to highest clockwise from the upper left (g-i<0.55, 0.55<g-i<0.85, 0.85<g-i<1.75, and 1.75<g-i).

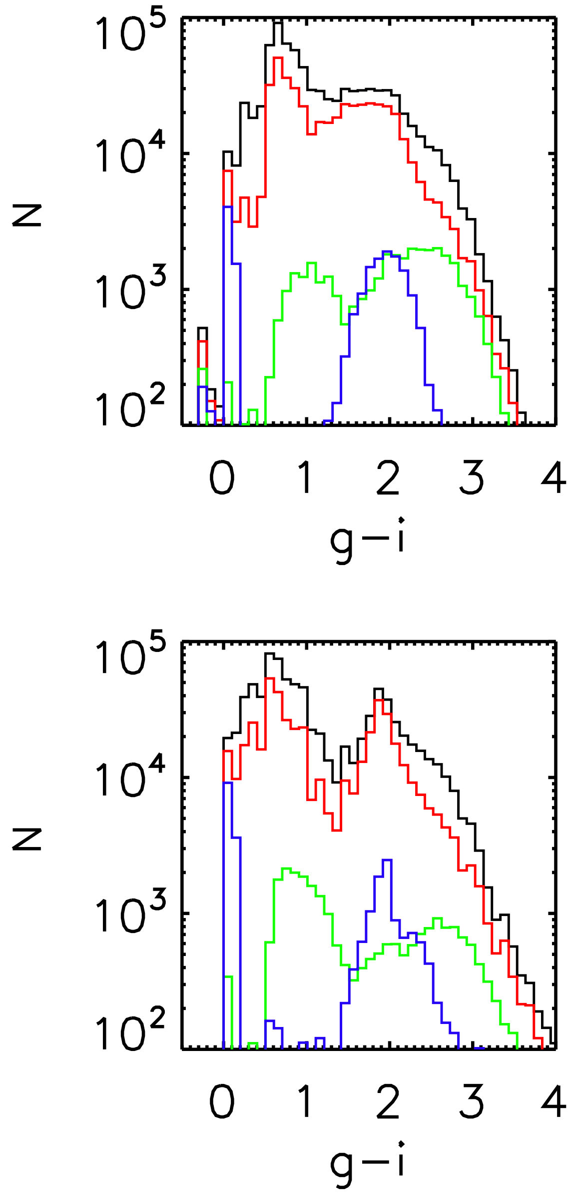
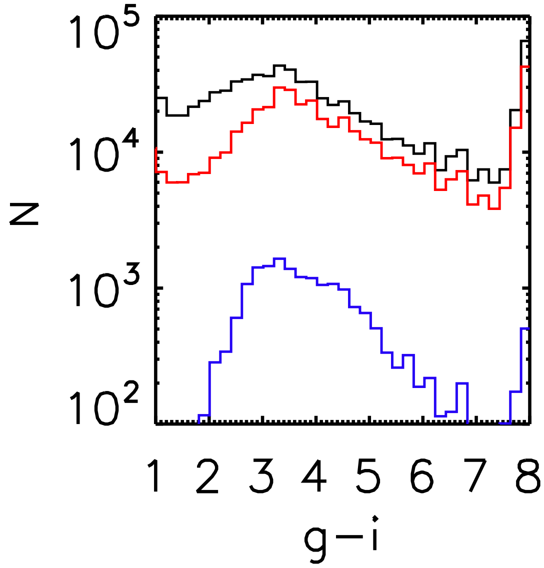


Fig. 26.— A comparison of the best-fit $g - i$ distribution for several subsamples of stars, for fits based only on SDSS data in the top panel, and for SDSS-2MASS fits in the bottom panel. Only stars with best-fit $\chi^2_{\rm pdf} < 3$ are used. The black line shows the full sample, the red line shows stars with $-5^\circ < b < 20^\circ$, the green line shows stars with $R_V = 1.0$, and the blue line shows stars with $A_r > 1.8$ and (main sequence) distance estimate below 200 pc.



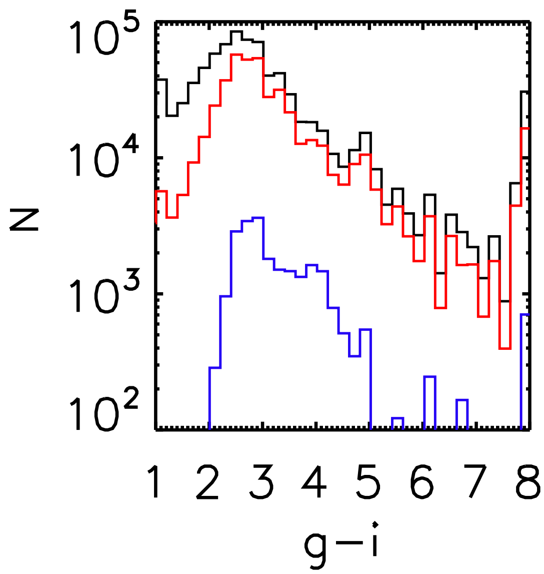


Fig. 27.— Analagous to Fig. 26, except that the distribution of best-fit R_V values are compared. The black line shows the full sample, the red line shows stars with $-5° < b < 20°$, and the blue line shows stars with $A_r > 1.8$ and (main sequence) distance estimate below 200 pc.

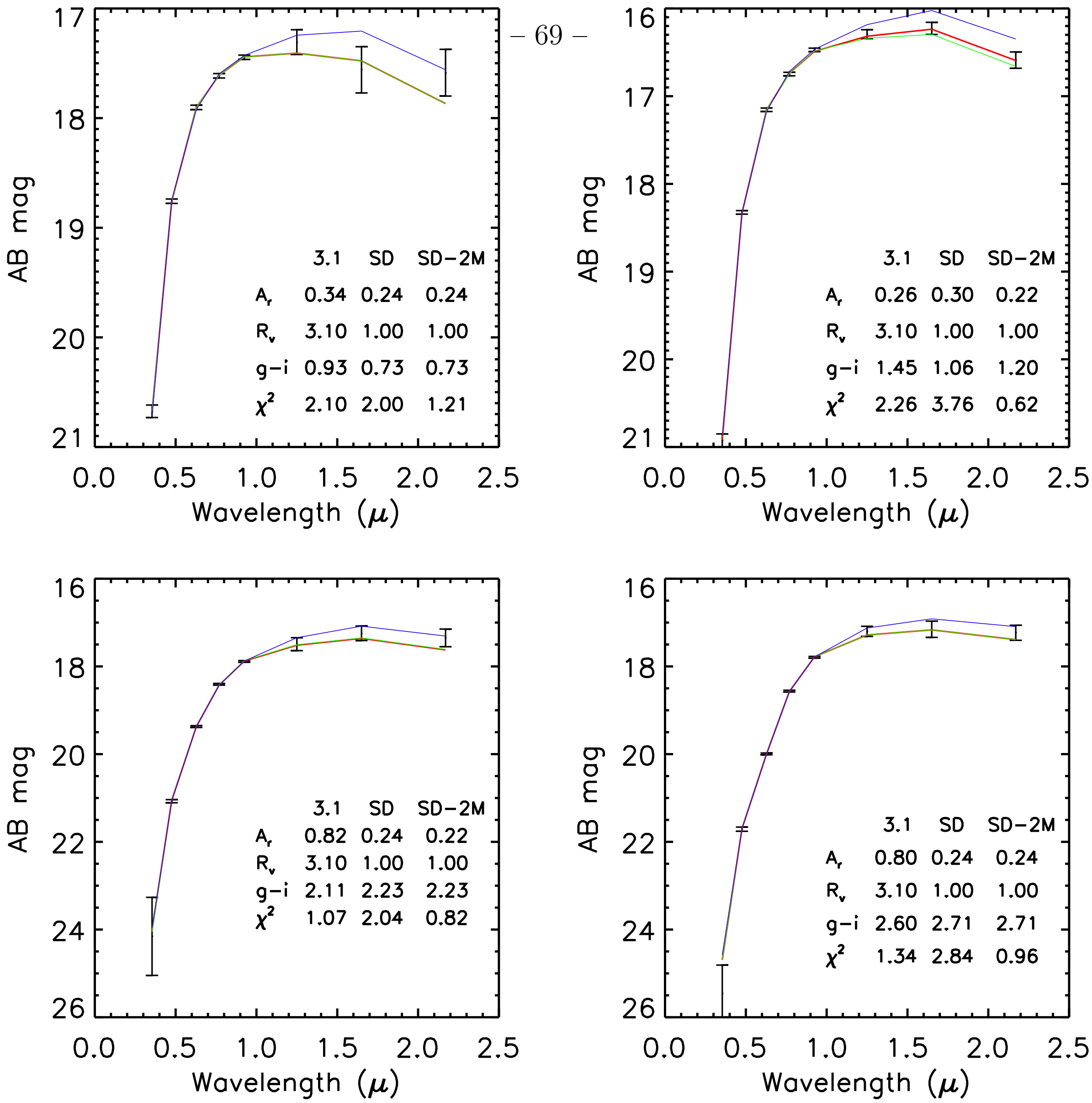


Fig. 28.— - A comparison of three different types of best-fit SEDs: using only SDSS data with fixed $R_V = 3.1$ (blue line), using only SDSS data with free R_V (red line) and using both SDSS and 2MASS data with free R_V (green line). The best-fit parameters are shown in each panel, and the panels correspond to four fiducial stars with different measured colors and a "runaway" best-fit value of $R_V = 1$ (the lowest possible value allowed in fitting). In the case of fixed $R_V = 3.1$ fit constrained only by SDSS data, the models overpredict 2MASS fluxes. When R_V is allowed to drift to the minimum value, both SDSS and SDSS-2MASS cases provide good fits. This agreement argues that such anomalous extinction curve result is not due to an unresolved companion.

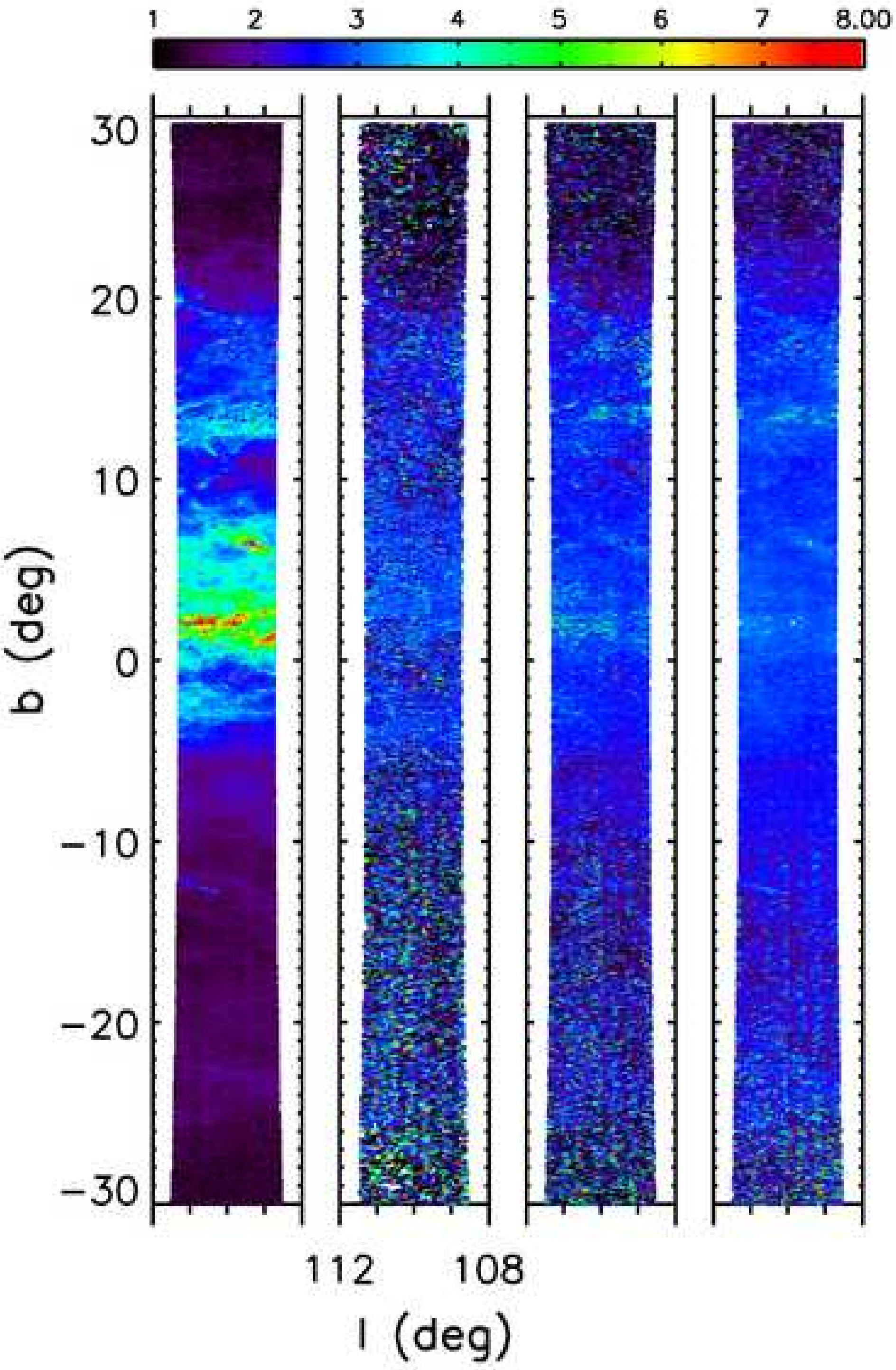


Fig. 29.— An illustration of the three-dimensional distribution of the best-fit R_V. The first panel (left) is added for reference and displays the median A_r for stars with χ^2 <3 using SDSS-2MASS data for the SEGUE $l = 110$ strip (using display scale 1–5). The next three panels display the median R_V values in distance slices 0.1-0.5 kpc (left), 0.5-1 kpc (middle), and $>$ 1kpc (right). The legend on top shows the color scale, and the pixel size is 4×4 arcmin^2.

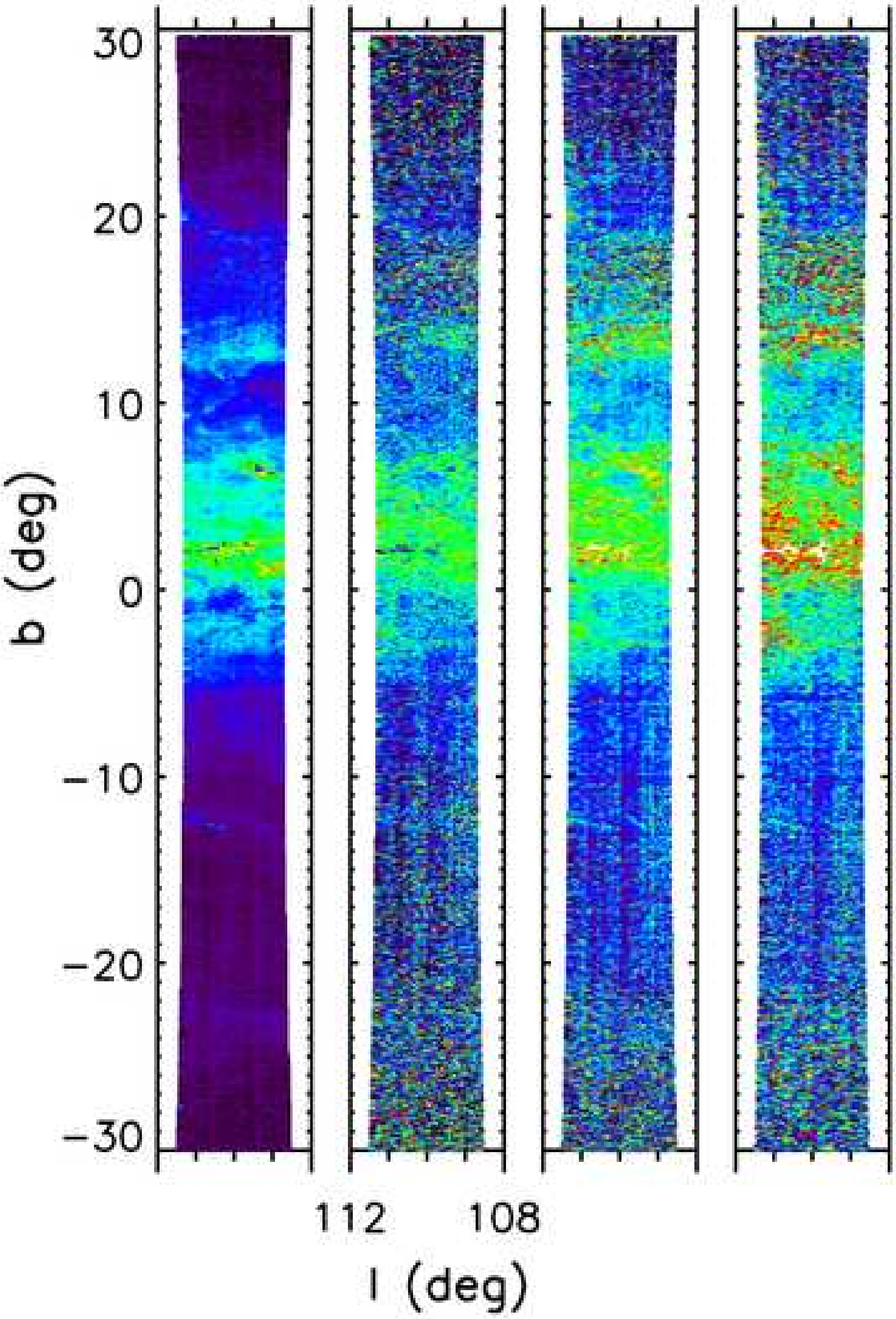


Fig. 30.— Analogous to Figure 29, except that this map shows values based on only SDSS data and has extended distance range: 0.5-1 kpc (left), 1-2 kpc (middle), and 2-3 kpc (right).

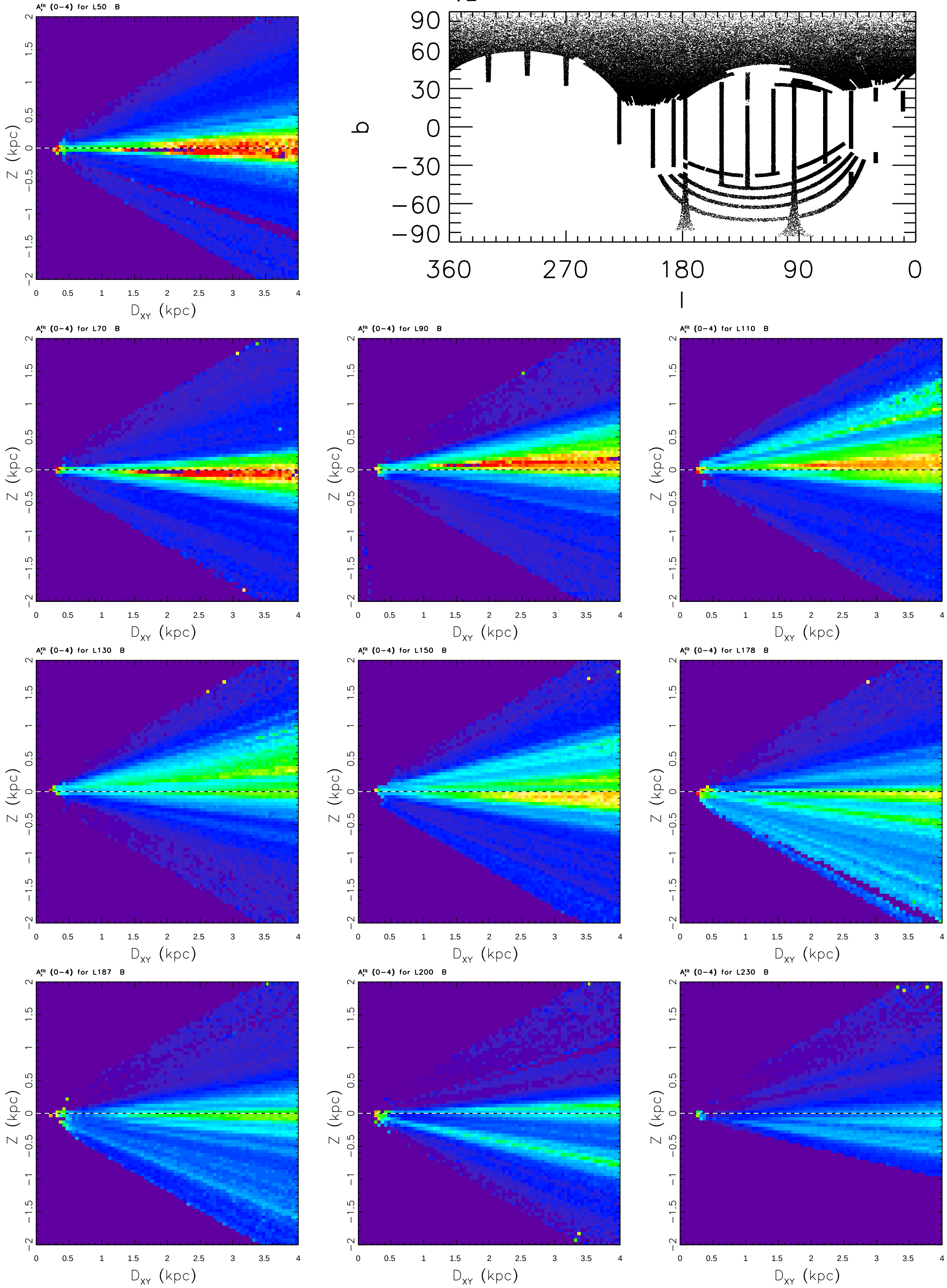


Fig. 31.— The best-fit A_r as a function of distance from the Galactic plane, Z, and distance along the plane, D_{xy}. Each pixel shows the median A_r between the observer and that point (i.e. this is **not** a cross-section of 3D dust map!). The top right panel shows the sky coverage of the analyzed data. Note that the extinction generally increases with distance for most lines

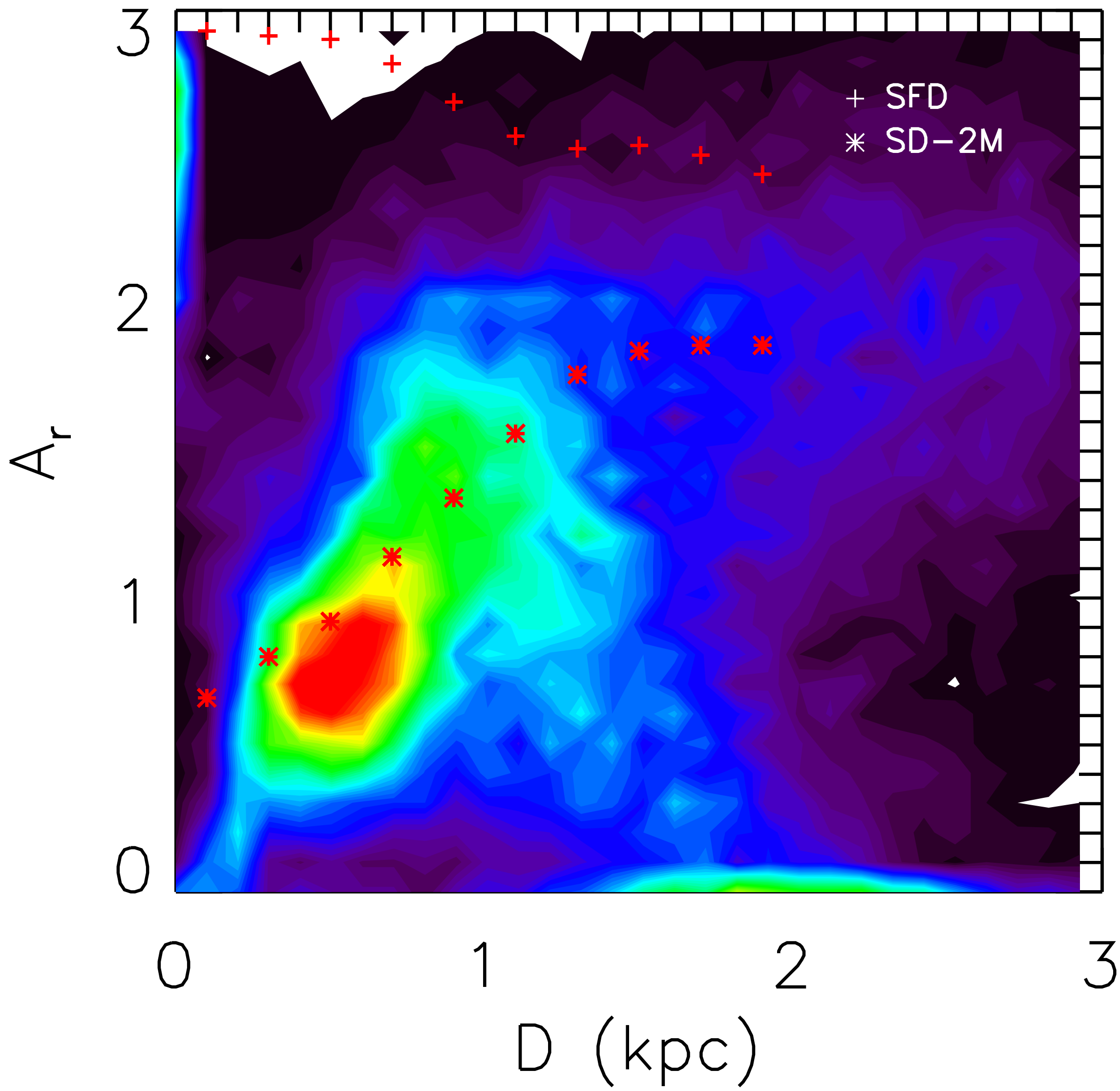


Fig. 32.— The increase of the best-fit extinction along the line of sight, A_r, as a function of best-fit distance to individual stars. The crosses represent the median SFD A_r value and the asterisks are the best-fit A_r values using SDSS-2MASS data in distance bins of 200pc.

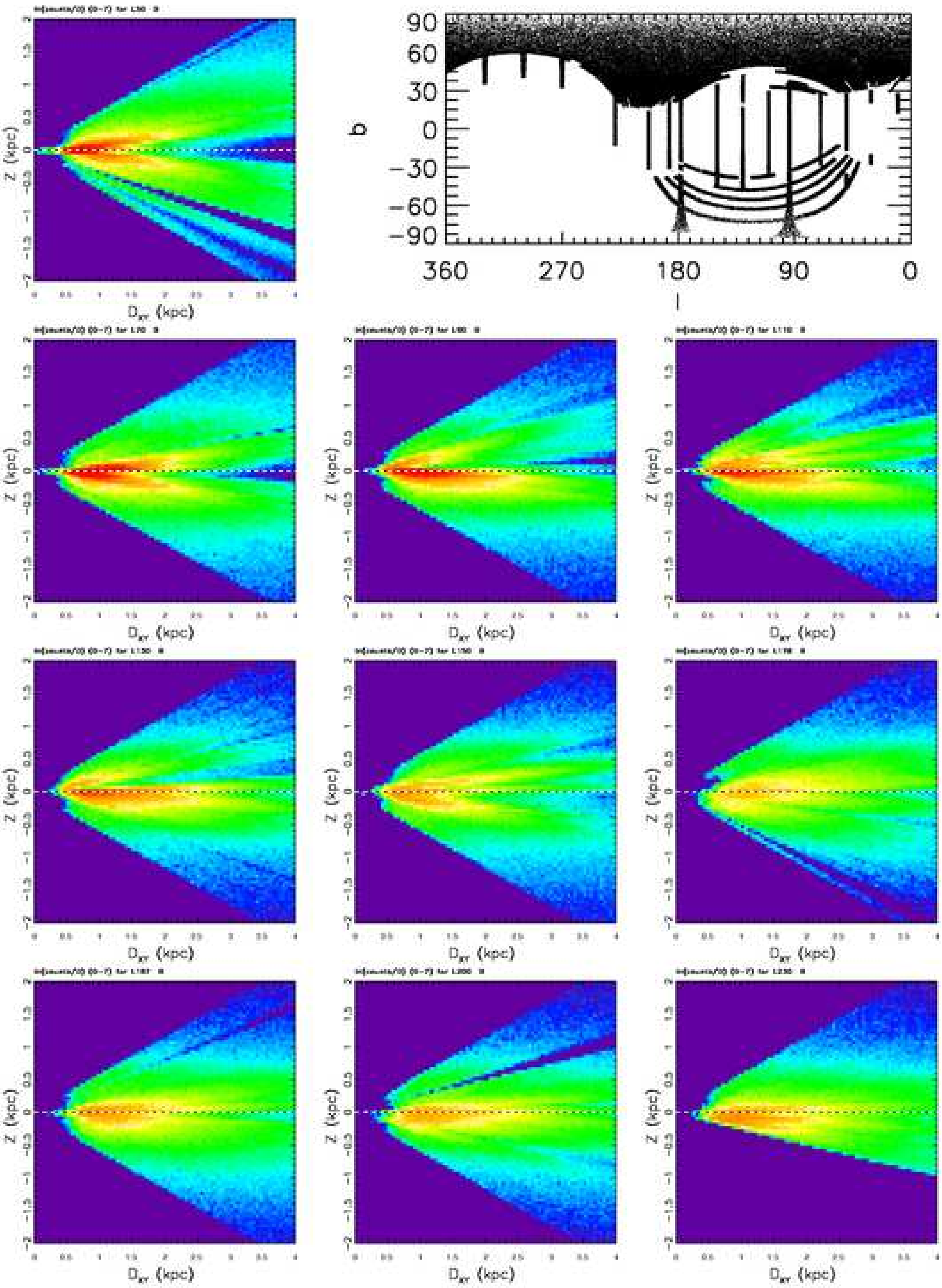


Fig. 33.— Similar to Figure 31, except that the local volume number density of stars is shown (arbitrarily normalized). The top right panel shows the sky coverage of the analyzed data. The fall-off of the stellar volume number density at distance beyond 1-2 kpc is due to the sample faint limit, and does not reflect the disk structure.