

Galactic Stellar Populations in the Era of Large Surveys

ŽELJKO IVEZIĆ

*Department of Astronomy, University of Washington, Box 351580, Seattle, WA
98195*

TIMOTHY C. BEERS

*Department of Physics & Astronomy and JINA: Joint Institute for Nuclear
Astrophysics, Michigan State University, East Lansing, MI 48824*

MARIO JURÍĆ

*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge,
MA 02138*

Key Words methods: data analysis – stars: statistics – Galaxy: disk, stellar
content, structure, interstellar medium

Abstract Studies of stellar populations, collections of stars with common spatial, kinematic, chemical, or age distributions, have been reinvigorated during the last decade by the advent of large-area sky surveys such as SDSS, 2MASS, RAVE, and others. Samples measured by these

new surveys presently include up to several hundred million stars. These data enable analyses that, together with theoretical and modeling advances, are revolutionizing our understanding of the nature of the Milky Way, and galaxy formation and evolution in general. The formation of galaxies like the Milky Way was long thought to be a steady process leading to a smooth distribution of stars. However, the abundance of substructure in the multi-dimensional space of various observables, such as position, kinematics, and metallicity, is by now proven beyond doubt, and demonstrates the importance of mergers in the growth of galaxies. Unlike smooth models that involve simple components, the new data reviewed here clearly show many irregular structures, such as the Sagittarius dwarf tidal stream in the halo and the Monoceros stream closer to the Galactic plane. These recent developments have made it abundantly clear that the Milky Way is a complex and dynamical structure that is still being shaped by the merging of neighboring smaller galaxies. We also briefly discuss the next generation of wide-field surveys, such as SkyMapper, Pan-STARRS, Gaia and LSST, which will improve measurement precision manyfold, and comprise billions of individual stars. The ultimate goal, development of a coherent and detailed story of the assembly and evolutionary history of the Milky Way and other large spirals like it, now appears well within reach.

CONTENTS

INTRODUCTION	3
<i>The Big Picture: Structure Formation and Near-field Cosmology</i>	3
<i>Stellar Populations: Definition and Role</i>	5
<i>Observational Methods: Photometry, Spectroscopy, Astrometry</i>	6
THE ADVENT OF LARGE-AREA DIGITAL SURVEYS	9
<i>SDSS Imaging and Spectroscopic Surveys</i>	9
<i>SDSS-POSS Proper Motion Survey</i>	11
<i>2MASS Imaging Survey</i>	11
<i>RAVE Spectroscopic Survey</i>	12
WHAT DID WE LEARN DURING THE LAST DECADE?	12

<i>Separation of the Main Structural Components</i>	12
<i>The Milky Way Disk</i>	12
<i>The Milky Way Halo</i>	12
<i>Ultra-faint Satellites and Connection with the Local Group</i>	13
UNANSWERED QUESTIONS	13
<i>The Milky Way Bulge and Disk</i>	13
<i>The Milky Way Halo and Satellites</i>	13
<i>Modeling Ambiguities</i>	13
THE ROAD AHEAD	13
<i>Pan-STARRS, SkyMapper, and the Dark Energy Survey</i>	14
<i>WISE</i>	14
<i>Gaia</i>	15
<i>LSST</i>	16
<i>New Analysis Challenges for Massive Datasets</i>	18

1 INTRODUCTION

1.1 The Big Picture: Structure Formation and Near-field Cosmology

The current cosmological paradigm states that the Universe had its beginning in the Big Bang. Galaxies, the fundamental building blocks of the Universe, formed soon after the Big Bang. A major objective of modern astrophysics is to understand when and how galaxies formed, and how they have evolved since then. Our own galaxy, the Milky Way, provides a unique opportunity to study a galaxy in great detail by measuring and analyzing the properties of a large number of individual stars.

The formation of galaxies like the Milky Way was long thought to be a steady

process that created a smooth distribution of stars, with this standard view exemplified by the Bahcall & Soneira (1980) and Gilmore, Wyse & Kuijken (1989) models, and described in detail by e.g. Majewski (1993). In these smooth models, the Milky Way is usually modeled by three discrete components described by relatively simple analytic expressions: the thin disk, the thick disk, and the halo. Instead, recent discoveries of complex substructure in the distribution of the Milky Way's stars have deeply shaken this standard view. Unlike those smooth models that involve simple components, the new data indicate much more irregular structures, such as the Sgr dwarf tidal stream in the halo and the Monoceros stream closer to the Galactic plane. These recent developments, based on accurate large-area surveys, have made it abundantly clear that the Milky Way is a complex and dynamical structure that is still being shaped by the infall (merging) of neighboring smaller galaxies.

Numerical simulations suggest that this merger process plays a crucial role in setting the structure and motions of stars within galaxies, and is a generic feature of current cosmological models. Since the individual stars that make up the stellar populations in the Milky Way can be studied in great detail, their characterization provides *clues about galaxy merging process that cannot be extracted from observations of distant galaxies.*

The main purpose of this review is to summarize recent revolutionary progress in data availability and the paradigm shifts in our understanding of galaxy formation and evolution resulting from this progress. We begin with a brief overview of methodology and major sky surveys, and then describe the main results. We conclude by discussing the main unanswered questions and prospects for the future.

1.2 Stellar Populations: Definition and Role

In astronomy, the term *stellar populations* is often associated with *Populations I, II and III*. These stellar classes represent a sequence of decreasing metallicity and increasing age. Here, we will use the term “stellar population” to mean any collection of stars with common spatial, kinematic, chemical, luminosity, or age distributions. For example, a sample of red giant stars selected using appropriate observables and selection criteria is considered a population, although such a sample can include both Population I and Population II stars. Similarly, we will often consider populations of “disk” and “halo” stars, or samples selected from a narrow color range. In summary, any sample of stars that share some common property that is appropriate for mapping the Galaxy in the space of various observables is hereafter considered to be a “population”.

Most studies of the Milky Way can be described as investigations of the stellar distribution, or statistical behavior of various populations, in the seven-dimensional (7-D) phase space spanned by the three spatial coordinates, three velocity components, and metallicity. Depending on the quality, diversity and quantity of data, such studies typically concentrate on only a limited region of this 7-D space (e.g. the nearby solar neighborhood, pencil beam surveys, kinematically biased surveys), or consider only marginal distributions (e.g., number density of stars irrespective of their metallicity or kinematics, proper motion surveys without metallicity or radial velocity information). *The main reason for the substantial progress over the last decade is the ability of modern sky surveys to deliver the necessary data for determining phase-space coordinates of a star for unprecedented numbers of faint stars detected over a large sky area.*

The large number of stars enables detailed studies of various distribution, in-

cluding determination of the distributions’ *shape*, rather than only low-order statistics as done with small samples. Deviations from Gaussian shapes often encode more information about the history of galaxy assembly than their mean and dispersion. The large samples are especially important with considering multi-variate distributions (as opposed to one-dimensional marginal distributions), when the so-called “curse of dimensionality” prevents their accurate determination with small samples.

In addition to increasing the sample size, the ability to detect faint stars is crucial for extending the sample distance limit. With SDSS, it has become possible to detect even main sequence (dwarf) stars to a distance limit exceeding 10 kpc and thus to probe both disk and halo with the same dataset. The main advantage of main sequence stars over probes such as RR Lyrae stars, blue horizontal branch (BHB) stars and red giant stars for studying Galaxy is that they are much more numerous, and thus enable a much higher spatial resolution of the resulting phase-space maps (assuming a fixed number of stars per multi-dimensional pixel in phase space).

1.3 Observational Methods: Photometry, Spectroscopy, Astrometry

In order to determine coordinates of a star in the 7-D phase space, a variety of astronomical techniques must be used. As always, the most crucial quantity to measure is stellar distance. The largest sample of stars with trigonometric distances, obtained by the Hipparcos survey, is too shallow to complement deep surveys such as SDSS and 2MASS (for overview of these surveys see below). Until the all-sky Gaia survey measures trigonometric distances for about a billions stars

brighter than $r = 20$ (see the last section), various photometric methods need be employed to estimate distances to stars. A common aspect of these methods is that luminosity (i.e., absolute magnitude) of a star is determined from its color measurements, and then distance is determined from the difference between absolute and apparent magnitudes. For certain populations, for example RR Lyrae stars, a good estimate of their absolute magnitude is a simple constant; for other populations, such as main sequence stars, absolute magnitude depends on both effective temperature and metallicity, and sometimes on age (or surface gravity) too. A photometric parallax method for main sequence stars is summarized below.

The most accurate measurements of stellar metallicity are based on spectroscopic observations (see below for a method for estimating metallicity using photometric data). The spectroscopic measurements are especially important when studying the extremely low end of the metallicity distribution. In addition to chemical composition, spectroscopic observations enable radial velocity measurements. The two largest existing stellar spectroscopic surveys are SDSS and RAVE (see the next section).

To measure all three velocity vector components, precise astrometric observations are also needed. The projection of the velocity vector into tangent plane (i.e., perpendicular to radial velocity component) is measured using proper motion (astrometric shift per unit time), which can be combined with distance estimate to yield velocity. The proper motion measurements place an additional constraint that at least two astrometric epoch must be available.

Therefore, both multi-color imaging, multi-epoch astrometry, and spectroscopy are required for measuring coordinates of a star in the 7-D position-velocity-

metallicity phase space. It is the advent of massive and accurate imaging and spectroscopic surveys that enabled the major progress in the Milky Way mapping during the last decade.

1.3.1 PHOTOMETRIC PARALLAX METHOD FOR MAIN SEQUENCE STARS

In order to estimate distances to main sequence stars with an accuracy of 10-20% using photometric parallax relation, multi-band optical photometry accurate to several percent (i.e., to several hundredths of magnitude) is required. This stringent requirement comes from the steepness of the color-luminosity relation (derivative of absolute magnitude in the r band with respect to $r - i$ color reaches ~ 10 mag/mag at the blue end), and is the main reason why it was not possible to use this method with large sky surveys prior to SDSS.

Using globular cluster data, I08 derived a polynomial expression for the absolute magnitude of main sequence stars in the SDSS r band as a function of their $g - i$ color and metallicity (see their eqs. A2 and A7). The accuracy of the resulting magnitudes is in the range 0.1-0.2 mag (I08; SIJ08), and the method enables studies of the ~ 100 pc to ~ 10 kpc distance range when used with SDSS data.

1.3.2 PHOTOMETRIC METALLICITY METHOD FOR MAIN SEQUENCE STARS

Stellar metallicity, together with effective temperature and surface gravity, is one of the three main parameters that affect observed properties of stars. In addition to being an informative observable when deciphering the Milky Way history (e.g., Majewski 1993; Freeman & Bland-Hawthorn 2002a; Helmi 2008; Majewski 2010; and references therein), the knowledge of stellar metallicity is crucial for accurate estimates of distances using photometric parallax relation.

The most accurate measurements of stellar metallicity are based on spectro-

scopic observations (see below for a method for estimating metallicity using photometric data). Despite the recent progress in the availability of stellar spectra (approaching a million!), the number of stars detected in imaging surveys is vastly larger. In addition to generally providing better sky and depth coverage than spectroscopic surveys, imaging surveys obtain essentially complete flux-limited samples of stars. The simple selection criteria used for the photometric surveys are advantageous when studying Galactic structure, compared to the complex targeting criteria that are typically used for spectroscopic samples.

As first suggested by Schwarzschild, Searle, & Howard (1955), the depletion of metals in a stellar atmosphere has a detectable effect on the emergent flux, in particular in the blue region where the density of metallicity absorption lines is highest (Beers & Christlieb 2005, and references therein). Recent analysis of SDSS data demonstrated that for blue F and G main sequence stars, a reasonable metallicity estimate can be derived from using the $u - g$ color (I08, B10). The expression A1 from B10, applicable to stars with $0.2 < g - r < 0.6$, was calibrated using $\sim 100,000$ stars with spectroscopic metallicity, and has errors in the range 0.2-0.3 dex when used with SDSS data (for stars in the range $-2 < [Fe/H] < 0$).

2 THE ADVENT OF LARGE-AREA DIGITAL SURVEYS

2.1 SDSS Imaging and Spectroscopic Surveys

The SDSS is a digital photometric and spectroscopic survey which covered about one quarter of the Celestial Sphere in the North Galactic cap, and produced a smaller area ($\sim 250 \text{ deg}^2$) but much deeper survey in the Southern Galactic hemisphere (Adelman-McCarthy et al. 2006 and references therein). SDSS data are publicly available from <http://www.sdss.org>. The Data Release 7 has a sky

coverage of about 12,000 deg, and includes photometric measurements for 357 million unique objects (approximately half are stars) and spectra for 660,000 stars.

The flux densities of objects detected in imaging data are measured almost simultaneously in five bands (u , g , r , i , and z) with effective wavelengths of 3540 Å, 4760 Å, 6280 Å, 7690 Å, and 9250 Å. The completeness of SDSS catalogs for point sources is $\sim 99.3\%$ at the bright end and drops to 95% at magnitudes of 22.1, 22.4, 22.1, 21.2, and 20.3 in u , g , r , i and z , respectively. Astrometric positions are accurate to about 0.1 arcsec per coordinate for sources brighter than $r \sim 20.5^m$ (Pier et al. 2003), and the morphological information from the images allows robust point source-galaxy separation to $r \sim 21.5^m$ (Lupton et al. 2001). The SDSS photometric accuracy is 0.02 mag (root-mean-square, at the bright end), with well controlled tails of the error distribution (Ivezić et al. 2003). The absolute zero point calibration of the SDSS photometry is accurate to within ~ 0.02 mag (Ivezić et al. 2004).

For comparison, the best large-area optical sky survey prior to SDSS, the photographic Palomar Observatory Sky Survey, had only two photometric bands and several times larger photometric errors (Sesar et al. 2006).

In addition to its imaging survey data, SDSS has obtained well over half a million stellar spectra. These spectra have wavelength coverage 3800–9200 Å and spectral resolution of ~ 2000 , with a signal-to-noise ratio per pixel of 5 at $g=20$. The spectra are automatically processed: for each object a spectral type (galaxy vs. star etc.), radial velocity and its error are determined by matching the measured spectrum to a well-calibrated set of templates. Random errors for the radial velocity measurements are a function of spectral type, but are usually

$< 5 \text{ km s}^{-1}$ for stars brighter than $g \sim 18$, rising to $\sim 25 \text{ km s}^{-1}$ for stars with $g = 20$ (Pourbaix et al. 2005).

2.2 SDSS-POSS Proper Motion Survey

The time difference of about half a century between the first Palomar Observatory Sky Survey (POSS) and SDSS imaging observations provides an excellent baseline to measure proper motions for tens of millions of stars to faint brightness levels. Munn et al. (2004) addressed the problem of large systematic astrometric errors in POSS catalogs by recalibrating the USNO-B catalog using the positions of galaxies measured by SDSS. As a result of this calibration, the SDSS-POSS proper motion measurements are now available for about 50 million unresolved sources, mostly stars. This catalog also includes about 70,000 spectroscopically confirmed SDSS quasars that were used to robustly estimate the proper motion errors. The random errors increase from $\sim 3 \text{ mas yr}^{-1}$ at the bright end to $\sim 6 \text{ mas yr}^{-1}$ at $g = 20$ (the sample completeness limit), with systematics typically an order of magnitude smaller and without discernible spatial variation. Even for stars at 1 kpc, the implied tangential velocity errors are as small as 10-20 km/s, and well matched to the SDSS radial velocity accuracy. This catalog represents an improvement of more than a factor of 100 over previously available data sets *both in size and accuracy*.

2.3 2MASS Imaging Survey

Two Micron All Sky Survey (2MASS, Skrutskie et al. 2006) is an all-sky near-IR survey with limiting (Vega-based, 10σ) magnitudes of $J=15.8$, $H=15.1$, and $K=14.3$. The 2MASS point source catalog contains positional and photometric

information for 471 million sources. Majewski et al. (2003) have demonstrated that M giant candidates color-selected from 2MASS database are extremely powerful probe for finding and tracing halo substructure out to 100 kpc.

2.4 RAVE Spectroscopic Survey

3 WHAT DID WE LEARN DURING THE LAST DECADE?

Explain that we are biased towards SDSS, and that we don't consider the bulge (give references to relevant work).

Discussion is organized as i) spatial distribution, ii) metallicity distribution, iii) kinematic distributions, iv) smooth background and departures from it

3.1 Separation of the Main Structural Components

Start with halo vs. disk, and thin vs. thick disk separation

3.2 The Milky Way Disk

Open with a description of leading models and why it is hard to find out what thick disk is...

Describe v_{Phi} vs FeH issues, and finalize with results from Lee et al.: figs 2, 5 and 7:

Discuss various models, as discussed by Lee et al. and other papers

3.3 The Milky Way Halo

Studies of the Galactic halo can help constrain the formation history of the Milky Way and the galaxy formation process in general. For example, within the framework of hierarchical galaxy formation (Freeman & Bland-Hawthorn 2002b), the

spheroidal component of the luminous matter should reveal substructures such as tidal tails and streams (Bullock, Kravtsov & Weinberg 2001; Harding et al. 2001; Helmi & White 1999; Johnston, Hernquist & Bolte 1996). The number of these substructures, due to mergers and accretion over the Galaxy’s lifetime, may provide a crucial test for proposed solutions to the “missing satellite” problem (Bullock, Kravtsov & Weinberg 2000). Substructures are expected to be ubiquitous in the outer halo (galactocentric distance $> 15 - 20$ kpc), where the dynamical timescales are sufficiently long for them to remain spatially coherent (Johnston, Hernquist & Bolte 1996; Mayer et al. 2002), and indeed many have been discovered (e.g., Belokurov et al. 2007a,b, 2006, 2007c, Grillmair 2009, Grillmair & Dionatos 2006, Ivezić et al. 2000, Jurić et al. 2008, Newberg et al. 2007, 2002, Vivas & Zinn 2006, Yanny et al. 2000).

Most distant halo: various luminous tracers, such as main sequence turn-off stars, RR Lyrae variables, or red giants are used to detect halo substructures, and of them, RR Lyrae stars have proven to be especially useful.

Streams (Grillmair!)

Sesar et al. figs:

3.4 Ultra-faint Satellites and Connection with the Local Group

The Cambridge group results, Wilman’s work

4 UNANSWERED QUESTIONS

4.1 The Milky Way Bulge and Disk

4.2 The Milky Way Halo and Satellites

4.3 Modeling Ambiguities

5 THE ROAD AHEAD

Pasted...

The results discussed 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 ultimately the Large Synoptic Survey Telescope (Ivezić et al. 2008a). These surveys will extend the faint limit of the current surveys, such as SDSS, by up to 5 magnitudes. In addition, upcoming Gaia mission (Perryman et al. 2001, Wilkinson et al. 2005) will provide superb astrometric and photometric measurement accuracy for sources with $r < 20$ that will enable unprecedented science programs, and WISE mission will extend the probed wavelength range to $22 \mu m$

5.1 Pan-STARRS, SkyMapper, and the Dark Energy Survey

Summarize PS, SM and DES...

5.2 WISE

Pasted...

NASA's Wide-field Infrared Survey Explorer (WISE; ?) mapped the sky at 3.4, 4.6, 12, and $22 \mu m$ in 2010 with an angular of 6 – 12 arcsec. WISE achieved

5σ point source sensitivities better than 0.08, 0.11, 1 and 6 mJy (corresponding to AB magnitudes of 19.1, 18.8, 16.4 and 14.5) in unconfused regions on the ecliptic in the four bands (for comparison, WISE represents an improvement over the IRAS survey's $12\ \mu\text{m}$ band sensitivity by about a factor of 1000). The astrometric precision for high signal-to-noise sources is better than 150 mas. The survey sensitivity improves toward the ecliptic poles due to denser coverage and lower zodiacal background. Saturation affects photometry for sources brighter than approximately 8.0, 6.7, 3.8 and -0.4 mag (Vega) at 3.4, 4.6, 12 and $22\ \mu\text{m}$, respectively.

The WISE Preliminary Release¹ includes data from the first 105 days of WISE survey observations. Primary release data products include an Atlas of 10,464 calibrated, coadded Image Sets, a Source Catalog containing positional and photometric information for over 257 million objects detected on the WISE images, and an Explanatory Supplement that provides a user's guide to the WISE mission and format, content, characteristics and cautionary notes for the Release products.

From the viewpoint of stellar population studies, WISE...

5.3 Gaia

Gaia is an ESA Cornerstone mission set for launch in 2012. Building on experience from HIPPARCOS, it will survey the sky to a magnitude limit of $r \sim 20$ (approximately, see the next section) and obtain astrometric and three-band photometric measurements for about 1 billion sources, as well as radial velocity and chemical composition measurements (using 847-874 nm wavelength range) for

¹<http://wise2.ipac.caltech.edu/docs/release/prelim/preview.html>

150 million stars with $r < 18$. The final data product, the Gaia Catalogue, is expected to be published by 2020.

The Gaia’s payload will include two telescopes sharing a common focal plane, with two $1.7^\circ \times 0.6^\circ$ viewing fields separated by a highly stable angle of 106.5° . The focal plane includes a mosaic of 106 CCDs, with a total pixel count close to one billion. Due to spacecrafts’ rotation and precession, the whole sky will be scanned in TDI (drift scanning) mode about 70 times on average during 5 years of operations. Gaia will produce broad-band G magnitudes with sensitivity in the wavelength range 330-1020 nm (FWHM points at ~ 400 nm and ~ 850 nm). The spectral energy distribution of each source will be sampled by a spectrophotometric instrument providing low resolution spectra in the blue (BP , effective wavelength ~ 520 nm) and the red (RP , effective wavelength ~ 800 nm). In addition, the RVS instrument (radial velocity spectrograph) will disperse the light in the range 847–874 nm, for which it will include a dedicated filter. Therefore, there are four passbands associated with the Gaia instruments: G , G_{BP} , G_{RP} and G_{RVS} .

5.4 LSST

The Large Synoptic Survey Telescope (LSST) is the most ambitious currently planned ground-based optical survey, with a unique survey capability in the faint time domain. The LSST design is driven by four main science themes: probing dark energy and dark matter, taking an inventory of the Solar System, exploring the transient optical sky, and mapping the Milky Way. LSST will be a large, wide-field ground-based system designed to obtain multiple images covering the sky that is visible from Cerro Pachón in Northern Chile. The current baseline

design, with an 8.4m (6.7m effective) primary mirror, a 9.6 deg² field of view, and a 3.2 Gigapixel camera, will allow about 10,000 square degrees of sky to be covered using pairs of 15-second exposures twice per night every three nights on average, with typical 5σ depth for point sources of $r \sim 24.5$ (AB). The system is designed to yield high image quality as well as superb astrometric and photometric accuracy. The total survey area will include 30,000 deg² with $\delta < +34.5^\circ$, and will be imaged multiple times in six bands, *ugrizy*, covering the wavelength range 320–1050 nm. The project is scheduled to begin the regular survey operations before the end of this decade. About 90% of the observing time will be devoted to a deep-wide-fast survey mode which will uniformly observe a 18,000 deg² region about 1000 times (summed over all six bands) during the anticipated 10 years of operations, and yield a coadded map to $r \sim 27.5$. These data will result in databases including 10 billion galaxies and a similar number of stars, and will serve the majority of the primary science programs. The remaining 10% of the observing time will be allocated to special projects such as a Very Deep and Fast time domain survey.

LSST will obtain proper motion measurements of comparable accuracy to those of Gaia at their faint limit, and smoothly extend the error vs. magnitude curve deeper by 5 mag (for details see Eyer et al., in preparation). With its *u*-band data, LSST will enable studies of metallicity and kinematics using the *same sample* of stars out to a distance of ~ 100 kpc (~ 200 million F/G main sequence stars brighter than $g = 23.5$, for a discussion see I08).

LSST will produce a massive and exquisitely accurate photometric and astrometric dataset for about 10 billion Milky Way stars. The coverage of the Galactic plane will yield data for numerous star-forming regions, and the *y* band data will

penetrate through the interstellar dust layer. Photometric metallicity measurements will be available for about 200 million main-sequence F/G stars which will sample the halo to distances of 100 kpc (?). No other existing or planned survey will provide such a massive and powerful dataset to study the outer halo (including Gaia which is flux limited at $r = 20$, and Pan-STARRS which will not have the u band). The LSST in its standard surveying mode will be able to detect RR Lyrae and classical novae out to 400 kpc, and hence explore the extent and structure of the halo out to half the distance to M31. All together, the LSST will enable studies of the stellar distribution beyond the presumed edge of the Galactic halo, of their metallicity distribution throughout most of the halo, and of their kinematics beyond the thick disk/halo boundary (?).

In the context of Gaia, the LSST can be thought of as its deep complement. A comparison of LSST and Gaia performance is given in Figure 18. Gaia will provide an all-sky catalog with unsurpassed trigonometric parallax, proper motion and photometric measurements to $r \sim 20$ for about 10^9 stars. LSST will extend this map to $r \sim 27$ over half of the sky, detecting about 10^{10} stars. Because of Gaia's superb astrometric and photometric quality, and LSST's significantly deeper reach, the two surveys are highly complementary: Gaia will map the Milky Way's disk with unprecedented detail, and LSST will extend this map all the way to the halo edge (Eyer et al., in prep).

5.5 New Analysis Challenges for Massive Datasets

Talk about Massive Data Mining, Knowledge Discovery, and Data Visualization.

References

1. Belokurov V, Evans NW, Bell EF, Irwin MJ, Hewett PC, et al. 2007a. *Astrophysical Journal Letters* 657:L89–L92
2. Belokurov V, Evans NW, Irwin MJ, Lynden-Bell D, Yanny B, et al. 2007b. *Astrophysical Journal* 658:337–344
3. Belokurov V, Zucker DB, Evans NW, Gilmore G, Vidrih S, et al. 2006. *Astrophysical Journal Letters* 642:L137–L140
4. Belokurov V, Zucker DB, Evans NW, Kleyana JT, Koposov S, et al. 2007c. *Astrophysical Journal* 654:897–906
5. Bond NA, Ivezić Ž, Sesar B, Jurić M, Munn JA, et al. 2010. *Astrophysical Journal* 716:1–29
6. Bullock JS, Kravtsov AV, Weinberg DH. 2000. *Astrophysical Journal* 539:517–521
7. Bullock JS, Kravtsov AV, Weinberg DH. 2001. *Astrophysical Journal* 548:33–46
8. 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>
9. Freeman K, Bland-Hawthorn J. 2002a. *Annu. Rev. Astro. Astrophys.* 40:487–537
10. Freeman K, Bland-Hawthorn J. 2002b. *Annu. Rev. Astro. Astrophys.* 40:487–537
11. Grillmair CJ. 2009. *Astrophysical Journal* 693:1118–1127
12. Grillmair CJ, Dionatos O. 2006. *Astrophysical Journal Letters* 643:L17–L20
13. Harding P, Morrison HL, Olszewski EW, Arabadjis J, Mateo M, et al. 2001. *Astronomical Journal* 122:1397–1419

14. Helmi A. 2008. *AAPR?* 15:145–188
15. Helmi A, White SDM. 1999. *Monthly Notices of the R.A.S.* 307:495–517
16. Ivezić Ž, Goldston J, Finlator K, Knapp GR, Yanny B, et al. 2000. *Astronomical Journal* 120:963–977
17. Ivezić Z, et al. 2008a. *ArXiv e-prints*
18. Ivezić Ž, et al. 2008b. *Astrophysical Journal* 684:287–325
19. Johnston KV, Hernquist L, Bolte M. 1996. *Astrophysical Journal* 465:278–+
20. Jurić M, et al. 2008. *Astrophysical Journal* 673:864–914
21. Kaiser N, et al. 2002. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, eds. JA Tyson, S Wolff, vol. 4836 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*
22. Law DR, Johnston KV, Majewski SR. 2005. *Astrophysical Journal* 619:807–823
23. Majewski SR. 1993. *Annu. Rev. Astro. Astrophys.* 31:575–638
24. Majewski SR. 2010. In *IAU Symposium*, ed. G. Bruzual & S. Charlot, vol. 262 of *IAU Symposium*
25. Mayer L, Moore B, Quinn T, Governato F, Stadel J. 2002. *Monthly Notices of the R.A.S.* 336:119–130
26. Munn JA, et al. 2004. *Astronomical Journal* 127:3034–3042
27. Newberg HJ, Yanny B, Cole N, Beers TC, Re Fiorentin P, et al. 2007. *Astrophysical Journal* 668:221–235
28. Newberg HJ, Yanny B, Rockosi C, Grebel EK, Rix HW, et al. 2002. *Astrophysical Journal* 569:245–274
29. Perryman MAC, et al. 2001. *Astronomy & Astrophysics* 369:339–363

30. Schlafman KC, et al. 2009. *ArXiv e-prints*
31. Sesar B, Ivezić Ž, Grammer SH, Morgan DP, Becker AC, et al. 2010.
Astrophysical Journal 708:717–741
32. Vivas AK, Zinn R. 2006. *Astronomical Journal* 132:714–728
33. Wilkinson MI, et al. 2005. *Monthly Notices of the R.A.S.* 359:1306–1335
34. Yanny B, Newberg HJ, Kent S, Laurent-Muehleisen SA, Pier JR, et al. 2000.
Astrophysical Journal 540:825–841

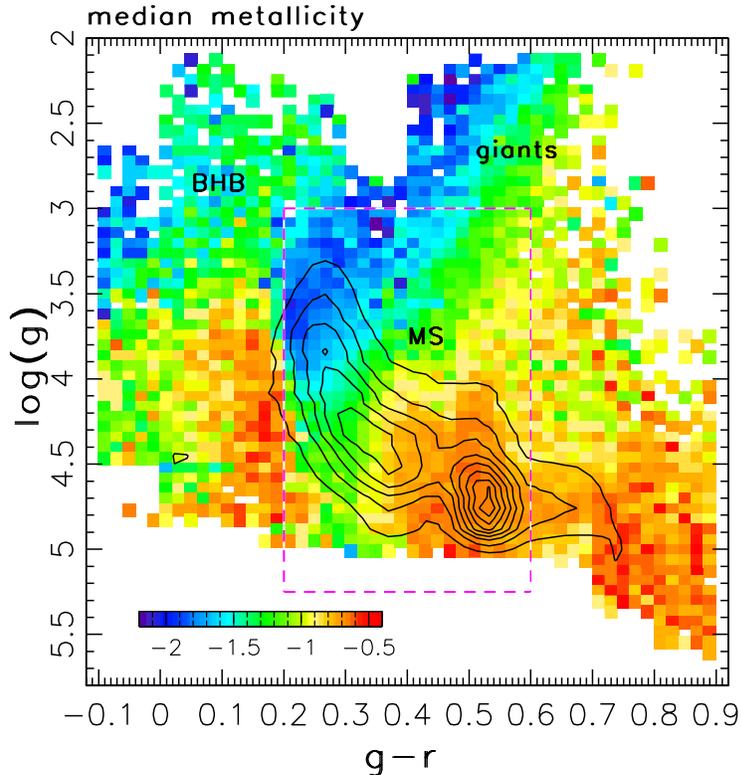


Figure 1: The stellar content of SDSS spectroscopic surveys (Figure 1 from Ivezić et al. (2008b)). Linearly spaced contours showing the distribution of $\sim 110,000$ stars with $g < 19.5$ and $0.1 < g - r < 0.9$ (corresponding to effective temperatures in the range 4500–8200 K) in the $\log(g)$ vs. $g - r$ plane. The multimodal distribution is a result of the SDSS target selection algorithm. The color scheme shows the median metallicity in all 0.02 mag by 0.06 dex large pixels that contain at least 10 stars. The fraction of stars with $\log(g) < 3$ (giants) is 4%, and they are mostly found in two color regions: $-0.1 < g - r < 0.2$ (BHB stars) and $0.4 < g - r < 0.65$ (red giants). They are dominated by low-metallicity stars ($[Fe/H] < -1$). The dashed lines outline the main-sequence (MS) region where photometric metallicity method can be applied.

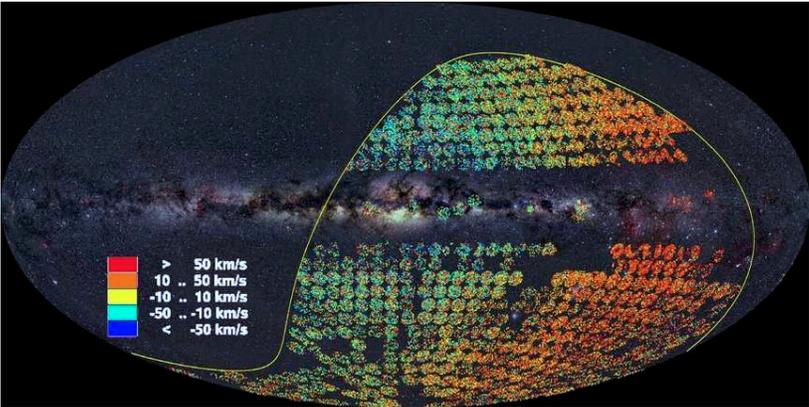


Figure 2: Add this to show RAVE progress?

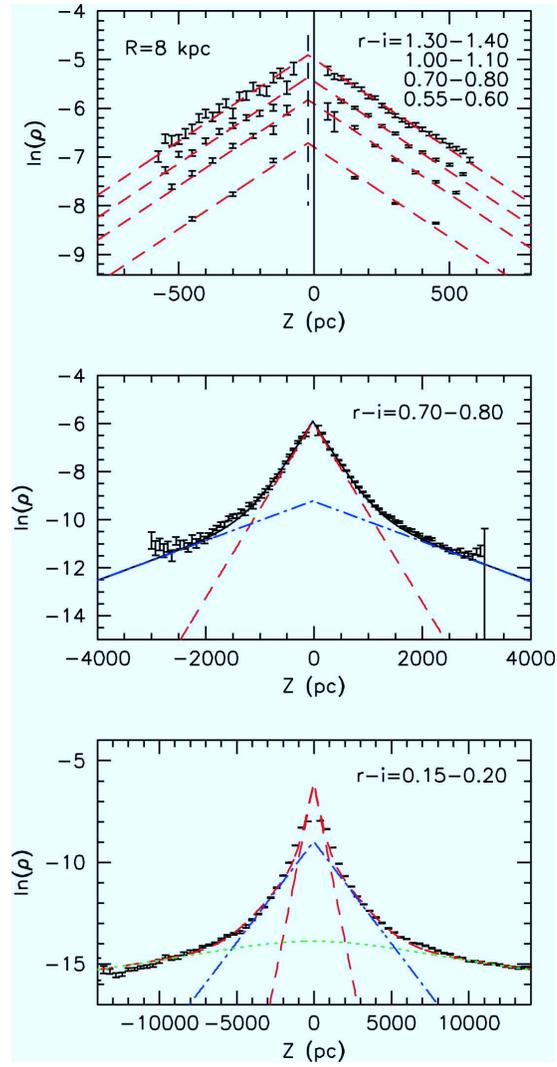


Figure 3: Figure 15 from Jurić et al. (2008).

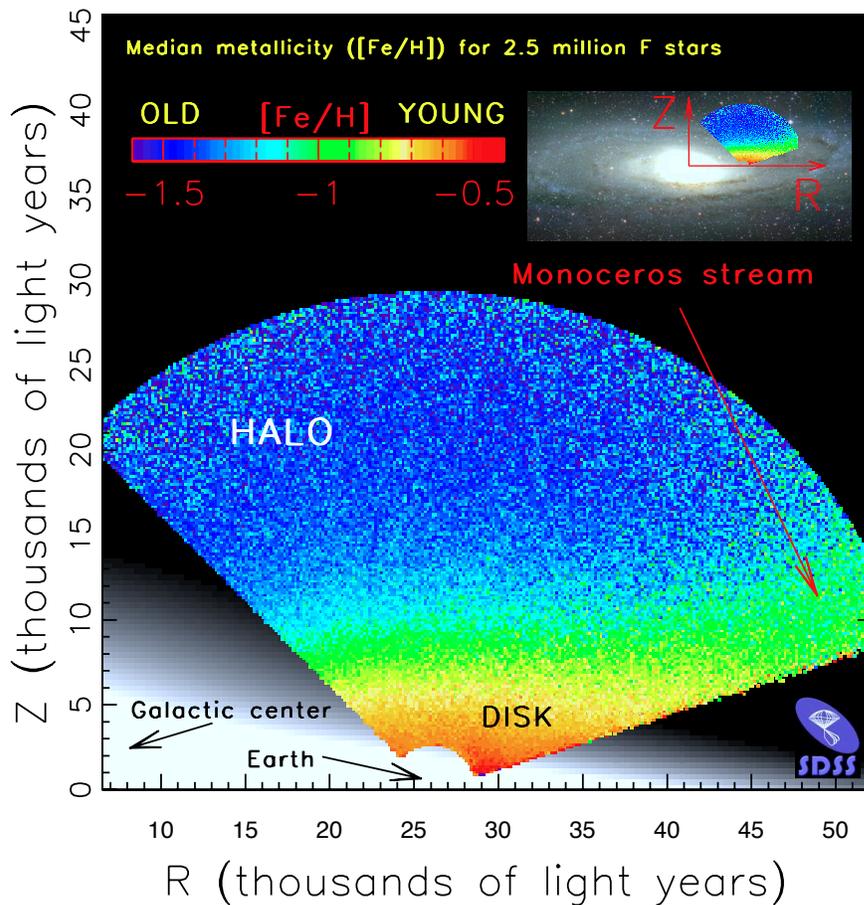


Figure 4: Dependence of the median photometric metallicity for ~ 2.5 million stars from SDSS with $14.5 < r < 20$ and $0.2 < g - r < 0.4$, and photometric distance in the 0.89 kpc range, in cylindrical Galactic coordinates R and $|Z|$. There are $\sim 40,000$ pixels (50 pc by 50 pc) contained in this map. Note that the gradient of the median metallicity is essentially parallel to the $|Z|$ axis, except in the Monoceros stream region, as marked.

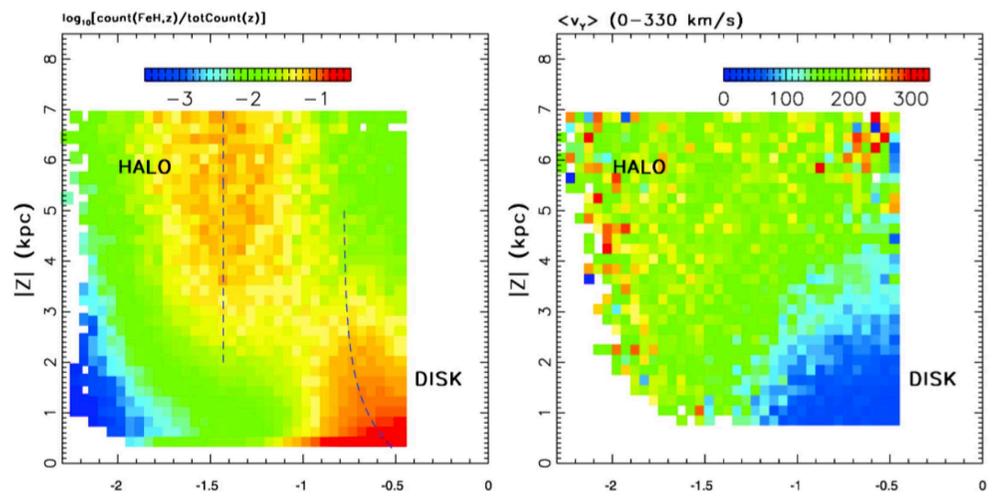


Figure 5: Figure 9 from Ivezić et al. (2008b). The left panel shows the conditional metallicity probability distribution for $\sim 300,000$ stars from a cylinder perpendicular to the Galactic plane, centered on the Sun, and with a radius of 1 kpc. The right panel shows the median heliocentric rotational velocity component (the value of $\sim 220 \text{ km s}^{-1}$ corresponds to no rotation) as a function of metallicity and distance from the Galactic plane for the stars from the left panel that also satisfy $b > 80^\circ$.

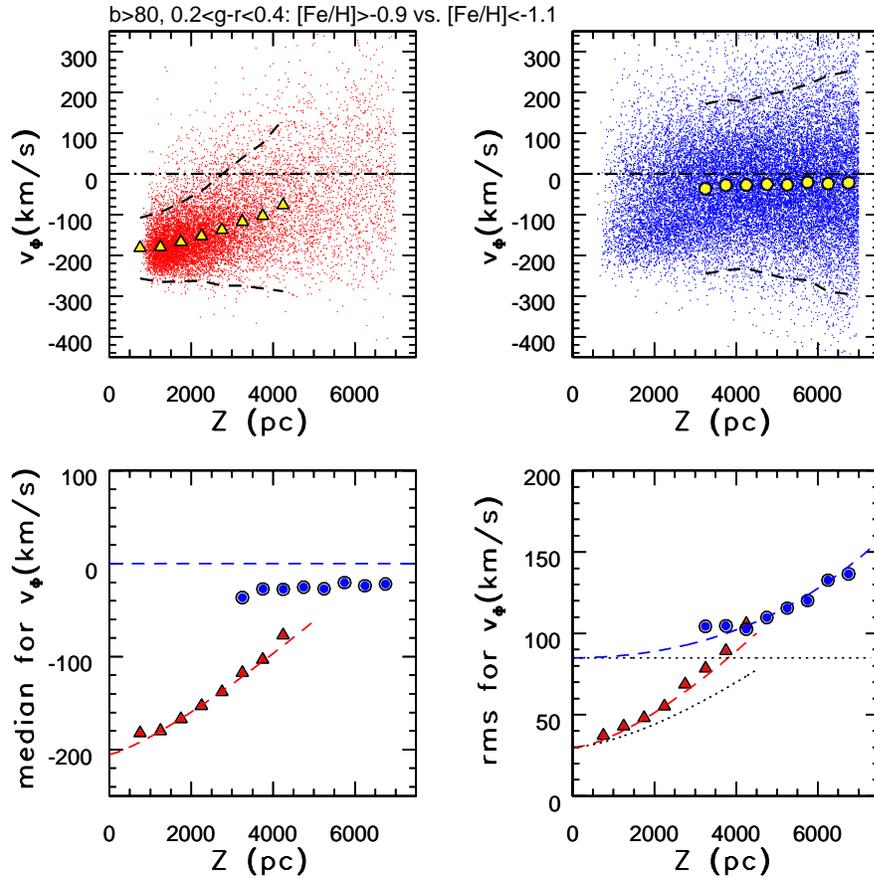


Figure 6: Figure 5 from Bond et al. (2010). A comparison of rotational velocity (see their eq. 8), v_ϕ , on distance from the Galactic plane, Z , for 14,000 high-metallicity ($[Fe/H] > 0.9$; top-left panel) and 23,000 low-metallicity ($[Fe/H] < 1.1$; top right) stars with $b > 80^\circ$. In the top two panels, individual stars are plotted as small dots, and the medians in bins of Z are plotted as large symbols. The 2σ envelope around the medians is shown by dashed lines. The bottom two panels compare the medians (left) and dispersions (right) for the two subsamples shown in the top panels, and the dashed lines in the bottom two panels show predictions of a kinematic model. The dotted lines in the bottom-right panel

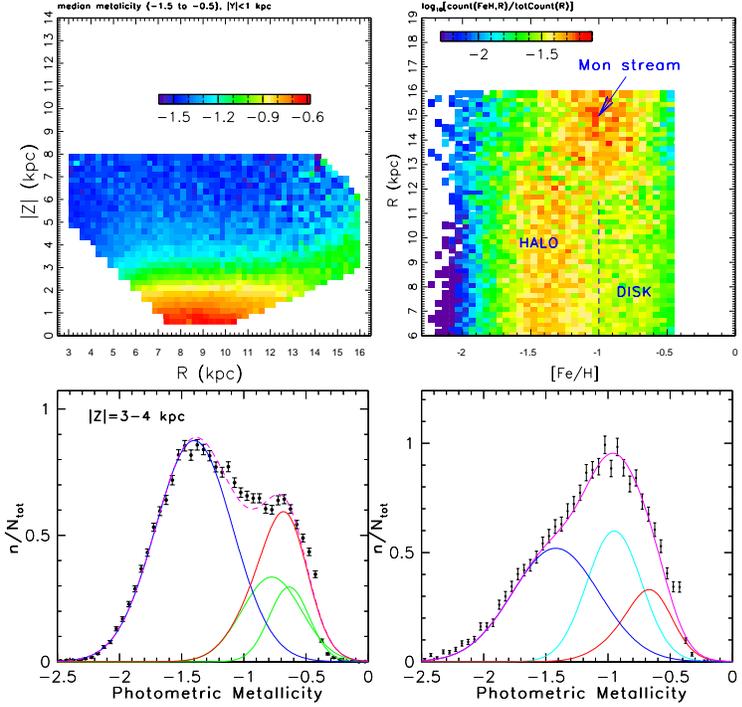


Figure 7: Figure 18 from Ivezić et al. (2008b). Top left panel: Dependence of the median photometric metallicity for ~ 1.04 million stars with $14.5 < r < 20$, $0.2 < g - r < 0.4$, and $|Y| < 1$ kpc, in cylindrical Galactic coordinates R and $|Z|$. This Y range is selected to include the Monoceros stream, which represents an overdensity by a factor of ~ 1.52 in a region around $R \sim 15$ kpc and $|Z| \sim 3 - 4$ kpc. As discernible from the map, this region has a larger median metallicity than expected for this $|Z|$ range based on extrapolation from smaller R . Top right panel: Conditional metallicity probability distribution for a subsample of $\sim 111,000$ stars with $3 < |Z|/\text{kpc} < 4$. The strong overdensity at $R > 12$ kpc is the Monoceros stream. The bottom panels show the metallicity distribution (symbols with error bars) for a subsample of $\sim 40,000$ stars with $6 < R/\text{kpc} < 9$ (left) and for $\sim 12,000$ stars with $13 < R/\text{kpc} < 16$ (right). The lines represent empirical fits discussed Ivezić et al. (2008b). The cyan line in the bottom right panel is a 0.22 dex wide Gaussian centered on $[Fe/H] = -0.95$. It accounts for 33% of stars in the sample that presumably belong to the Monoceros stream.

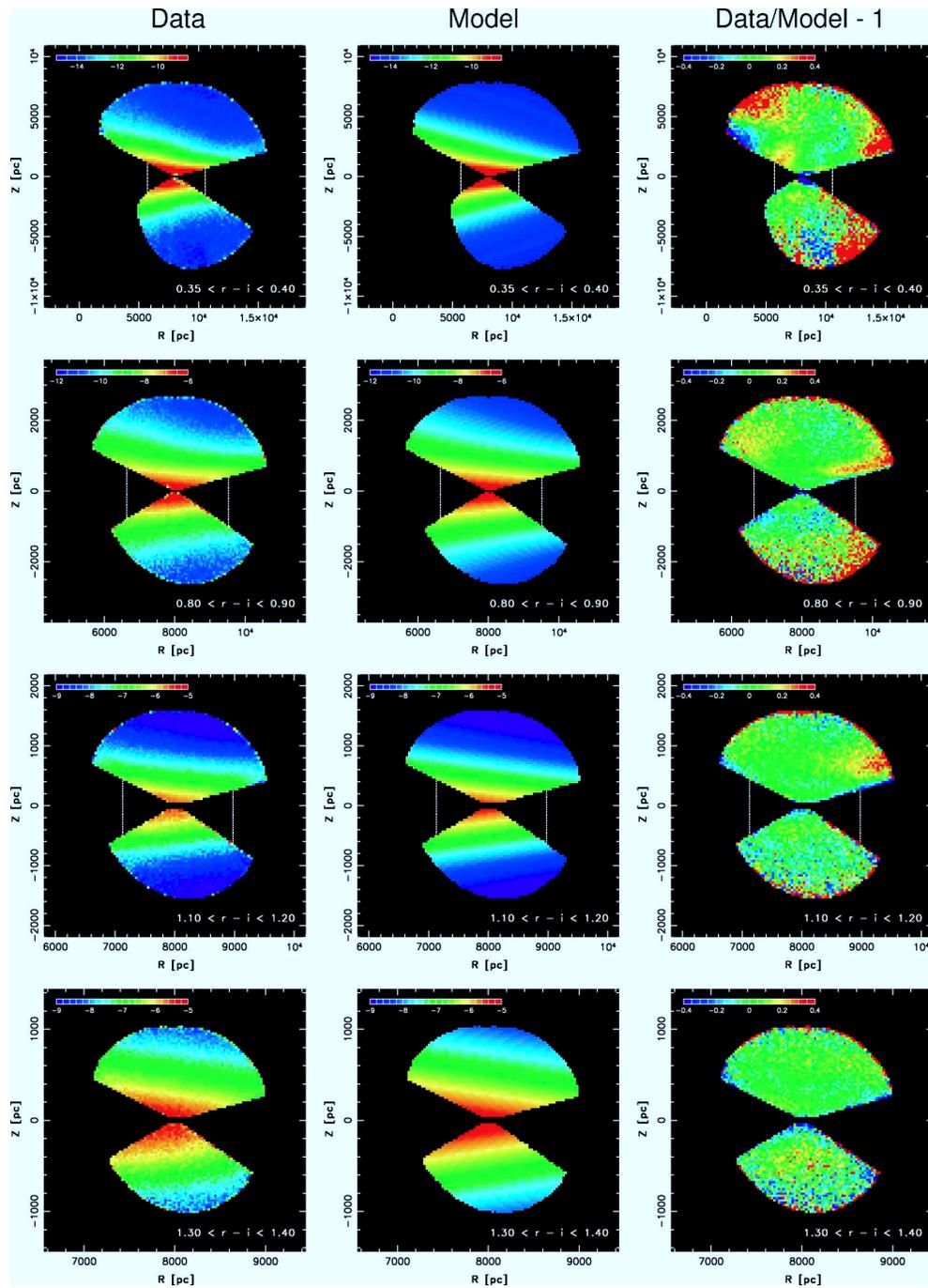


Figure 8: Figure 15 from Jurić et al. (2008).

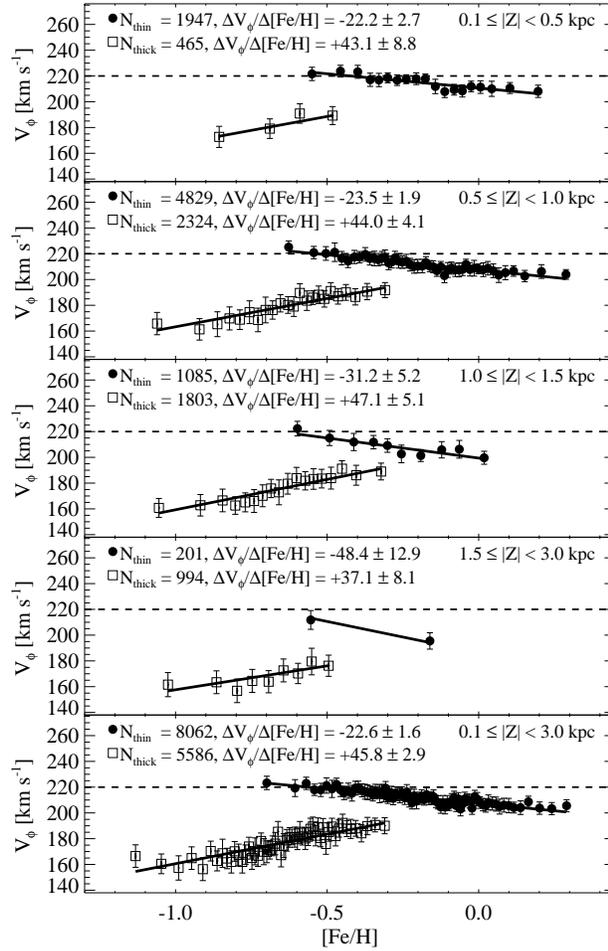


Figure 11: Lee fig. 7

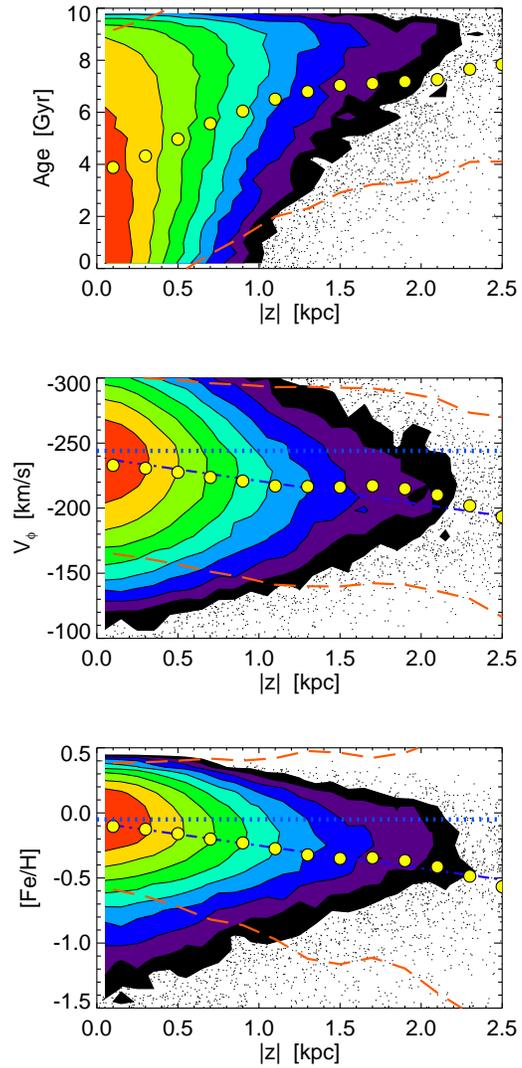


Figure 12: Figure 7 from ?). It shows that radial migration models produce distributions in qualitative agreement with observations.

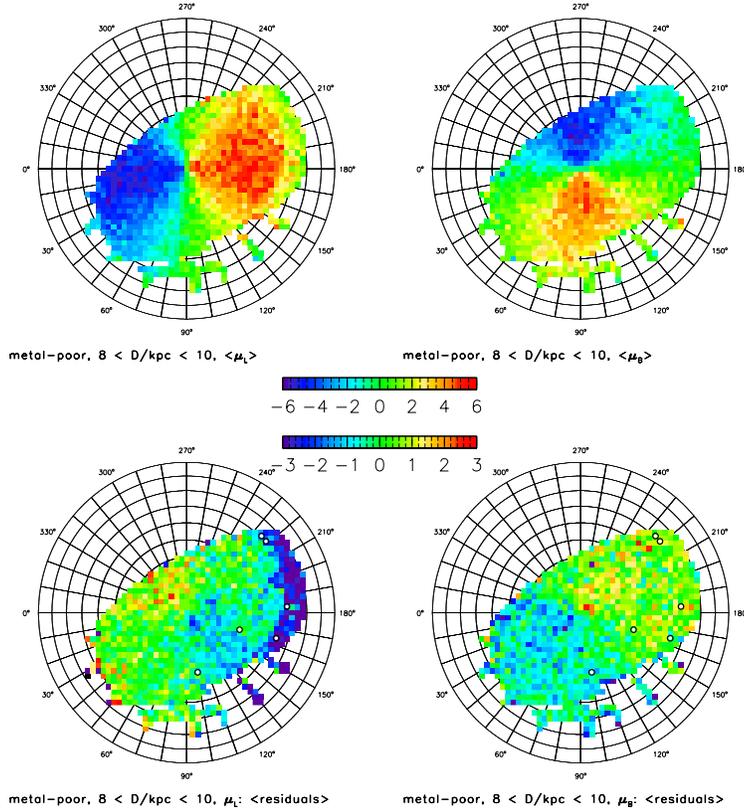


Figure 13: Figure 21 from Bond et al. (2010). Distribution of the median longitudinal proper motion in a Lambert projection of the North Galactic cap for low-metallicity (spectroscopic $[Fe/H] < 1.1$), blue ($0.2 < gr < 0.4$) stars, with distances in the range 8-10 kpc. The top two panels show the median longitudinal (left) and latitudinal (right) proper motions, and the two bottom panels show the median difference between the observed and model-predicted values. The maps are color-coded according to the legends in the middle (mas yr^{-1}); note that the bottom scale has a harder stretch to emphasize structure in the residual maps). In the bottom panels, the white symbols show the positions of the six northern cold substructures identified by Schlaufman et al. (2009).

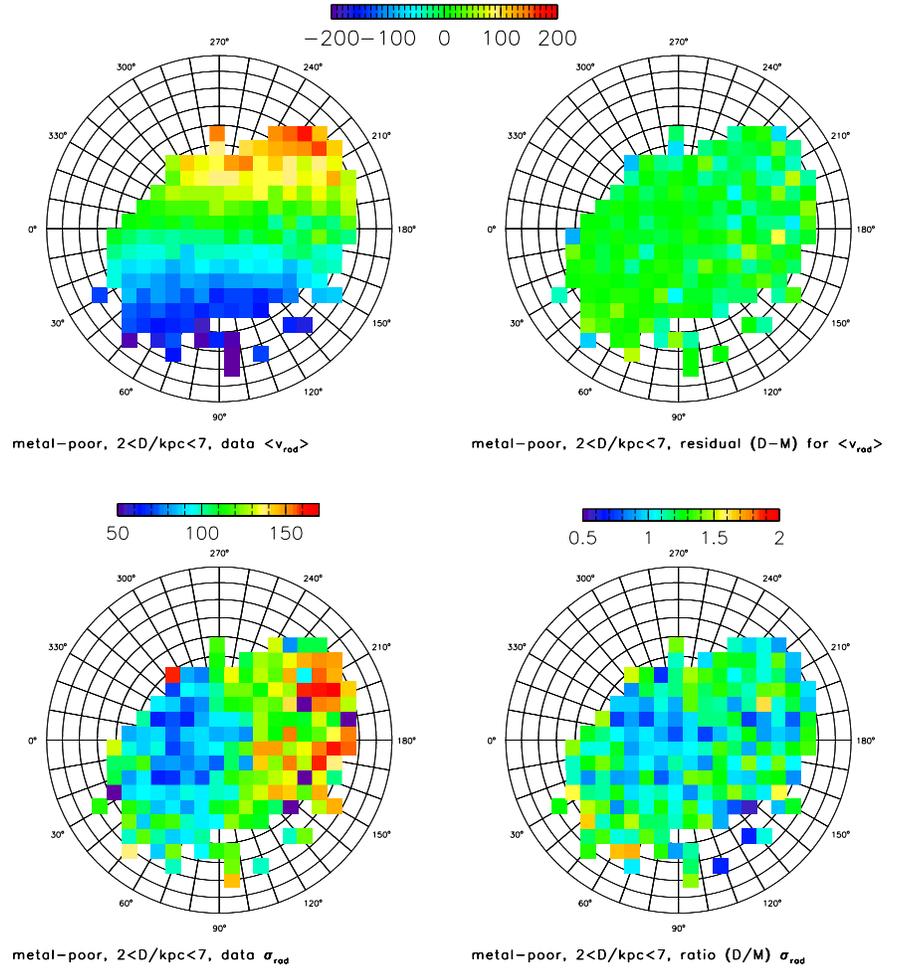


Figure 14: Figure 16 from Bond et al. (2010). Comparison of medians and dispersions for the measured and modeled radial velocities of 20,000 blue ($0.2 < gr < 0.4$) halo stars (spectroscopic $[Fe/H] < 1.1$) at distances, $D = 27$ kpc, and $b > 20^\circ$. The top-left panel shows the median measured radial velocity in each pixel, color-coded according to the legend shown at the top (units are km s^{-1}). The top-right panel shows the difference between this map and an analogous map based on model-generated values of radial velocity, using the same scale as in the top-left panel. The bottom-left panel shows the dispersion of measured radial velocities, color-coded according to the legend above it. The bottom-right panel shows the ratio of this map and an analogous map based on model-generated values of radial velocity, color-coded according to the legend above it.

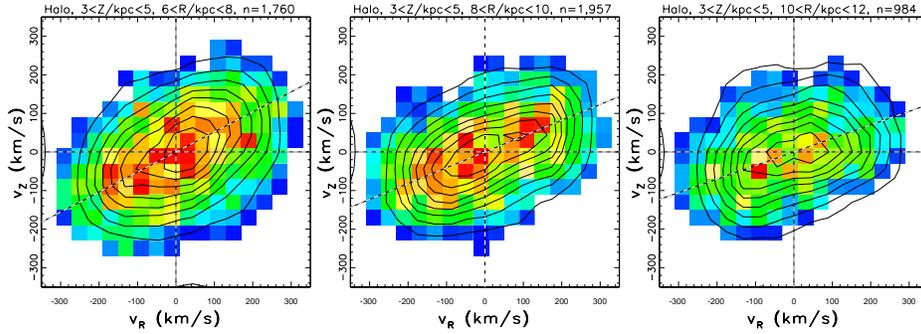


Figure 15: Figure 13 from Bond et al. (2010). The two-dimensional v_Z vs. v_R projections of the velocity distribution for three subsamples of candidate halo stars selected using spectroscopic metallicity ($3 < [Fe/H] < 1.1$), with $3 < Z/\text{kpc} < 5$, and $6 < R/\text{kpc} < 8$ (left), $8 < R/\text{kpc} < 10$ (middle), and $10 < R/\text{kpc} < 12$ (right). The distributions are shown using linearly spaced contours, and with a color-coded map showing smoothed counts in pixels (low to high from blue to red). The measurement errors are typically 60 km s^{-1} , and the dashed lines show the median direction toward the Galactic center. Note the strong evidence for a velocity-ellipsoid tilt, and the variation of the tilt with R so that the ellipsoid always points towards the Galactic center.

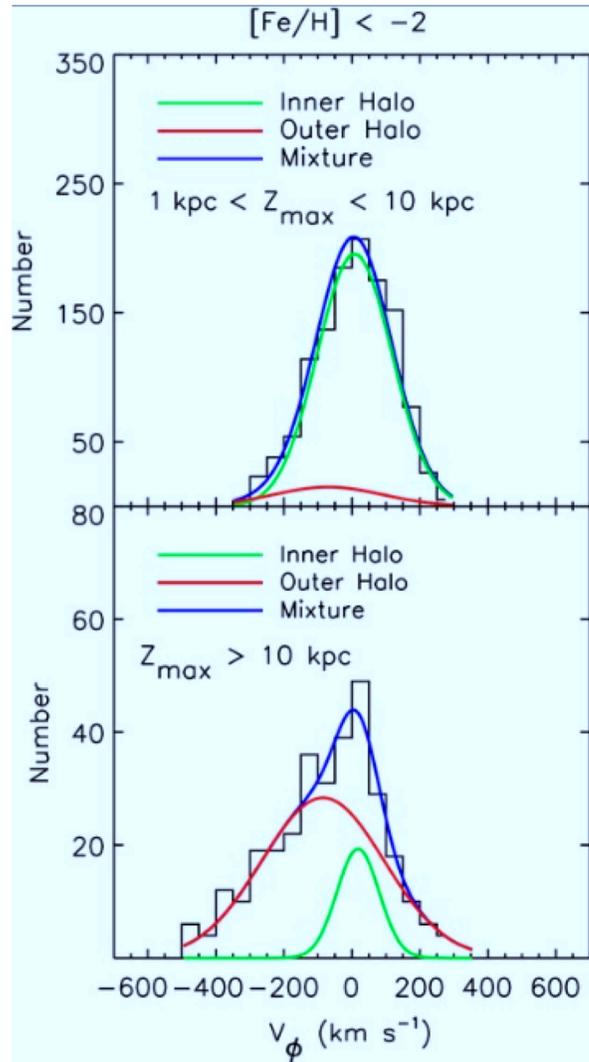


Figure 16: Figure 10 from Carollo et al.

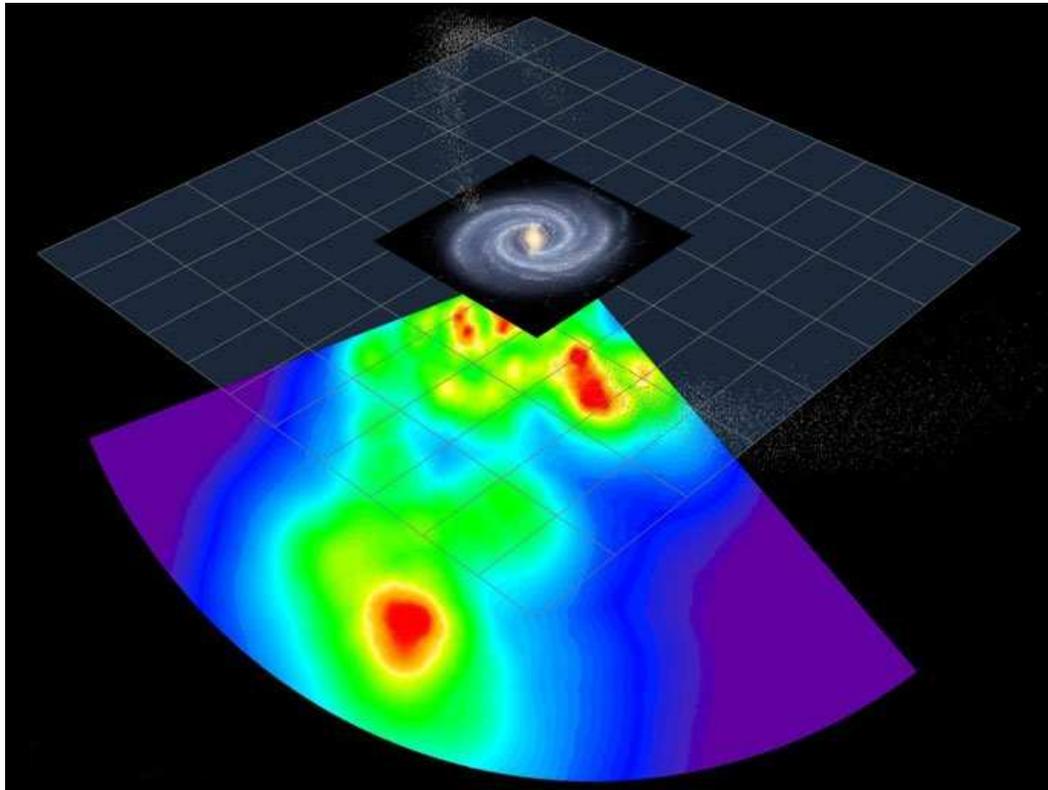


Figure 17: Figure 12 from Sesar et al. (2010). The stripe 82 plane and artists concept of the disk plane. The white dots show the Sagittarius dSph and its tidal streams, as modeled by the Law, Johnston & Majewski (2005) “spherical” model.

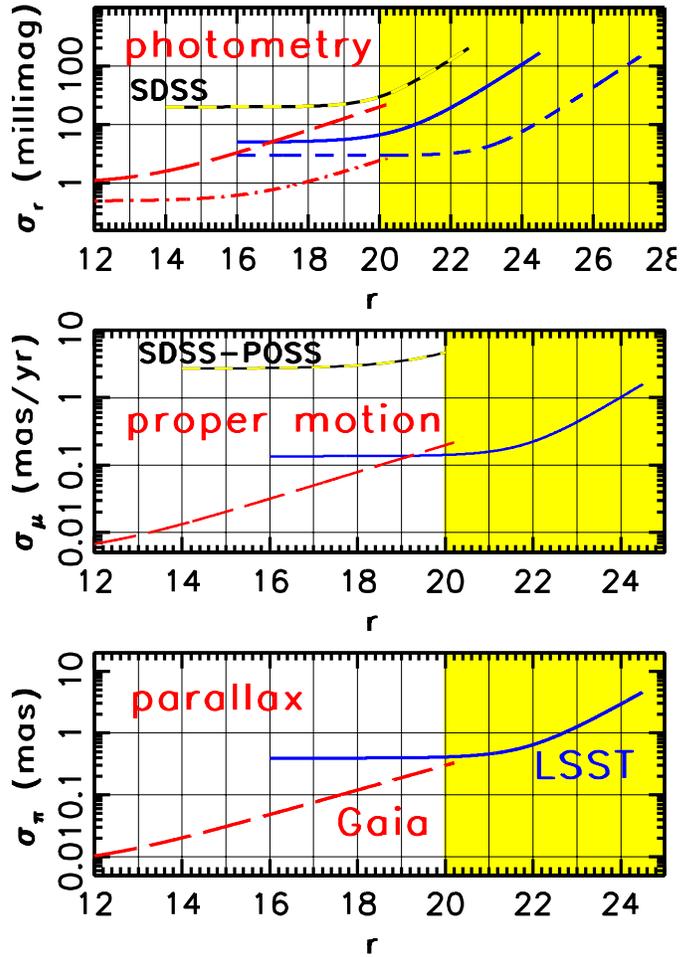


Figure 18: A comparison of photometric, proper motion and parallax errors for SDSS, Gaia and LSST, as a function of apparent magnitude r , for a G2V star (Eyer et al, in prep). In the top panel, the curve marked “SDSS” corresponds to a single SDSS observation. The red curves correspond to Gaia; the long-dashed curve shows a single *transit* accuracy, and the dot-dashed curve the end of mission accuracy (assuming 70 transits). The blue curves correspond to LSST; the solid curve shows a single *visit* accuracy, and the short-dashed curve shows accuracy for co-added data (assuming 230 visits in the r band). The curve marked “SDSS-POSS” in the middle panel shows accuracy delivered by the proper motion catalog of Munn et al. (2004).