*R*esearch in
*A*stronomy and
*A*strophysics

Mapping the Milky Way with LAMOST I: method and overview

Chao Liu[1], Yan Xu[1], Jun-Chen Wan[1,2], Hai-Feng Wang[1,2], Jeffrey L. Carlin[3], Li-Cai Deng[1],
Heidi Jo Newberg[4], Zi-Huang Cao[1], Yong-Hui Hou[5], Yue-Fei Wang[5] and Yong Zhang[5]

[1] Key Lab of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China; *liuchao@nao.cas.cn*
[2] University of the Chinese Academy of Sciences, Beijing 100049, China
[3] LSST, 933 North Cherry Avenue, Tucson, AZ 85721, USA
[4] Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[5] Nanjing Institute of Astronomical Optics & Technology, National Astronomical Observatories, Chinese Academy of Sciences, Nanjing 210042, China

Abstract We present a statistical method to derive the stellar density profiles of the Milky Way from spectroscopic survey data, taking into account selection effects. We assume the selection function, which can be altered during observations and data reductions, of the spectroscopic survey is based on photometric colors and magnitude. Then the underlying selection function for a line-of-sight can be recovered well by comparing the distribution of the spectroscopic stars in a color-magnitude plane with that of the photometric dataset. Subsequently, the stellar density profile along a line-of-sight can be derived from the spectroscopically measured stellar density profile multiplied by the selection function. The method is validated using *Galaxia* mock data with two different selection functions. We demonstrate that the derived stellar density profiles reconstruct the true ones well not only for the full set of targets, but also for sub-populations selected from the full dataset. Finally, the method is applied to map the density profiles for the Galactic disk and halo, using the LAMOST RGB stars. The Galactic disk extends to about $R = 19$ kpc, where the disk still contributes about 10% to the total stellar surface density. Beyond this radius, the disk smoothly transitions to the halo without any truncation, bending or breaking. Moreover, no over-density corresponding to the Monoceros ring is found in the Galactic anti-center direction. The disk shows moderate north–south asymmetry at radii larger than 12 kpc. On the other hand, the R–Z tomographic map directly shows that the stellar halo is substantially oblate within a Galactocentric radius of 20 kpc and gradually becomes nearly spherical beyond 30 kpc.

Key words: methods: statistical — Galaxy: structure — Galaxy: disk — Galaxy: halo — surveys: LAMOST

1 INTRODUCTION

The stellar density profile is important for unveiling the nature of the Galaxy. Either photometric survey data, which are supposed to be complete to some extent, or spectroscopic survey data, which can provide more accurate stellar parameters but obviously do not completely cover the sky due to lower efficiency of sampling, can be used to count the stars in three-dimensional (3D) space.

Although star count modeling has been studied for a long time, not until recently has it been able to pro-

vide reliable measurements of the Galactic disks and halo based on modern surveys (e.g., Chen et al. 2001; Jurić et al. 2008; Watkins et al. 2009; Bovy et al. 2012; Xue et al. 2015; also see the review by Bland-Hawthorn & Gerhard 2016). Jurić et al. (2008) fitted the star counts of various spectral types of stars from SDSS photometric survey data with two exponential disks and a power-law stellar halo. They found that the scale length of the thin disk is 2.6 kpc, substantially smaller than that of the thick disk, which is 3.6 kpc. However, Bovy et al. (2012) showed that the scale lengths change from ~ 4.5 kpc for

mono-abundance populations with small scale heights to $\sim 2\,$kpc for those with large scale heights. Comparison between the two results is non-trivial, since the former work identified the thick disk by geometry, while the latter one defined the thick disk based on chemical abundances.

It is worthwhile to point out that although the disk profile is often empirically simplified as an exponential or hyperbolic secant form, it is composed of substantially asymmetric structures such as the bar and spiral arms. Furthermore, López-Corredoira et al. (2002) unveiled that the disk is significantly flared and warped in the outskirts using red clump giant stars selected from the 2MASS catalog (Skrutskie et al. 2006). Moreover, recent studies also demonstrated vertical asymmetric structures in the outer disk (Xu et al. 2015) as well as in the solar neighborhood (Widrow et al. 2012).

Regarding the stellar halo, Watkins et al. (2009) and Deason et al. (2011) used RR Lyrae and Blue Horizontal Branch stars, both of which are accurate standard candles, to measure the shape of the stellar halo, respectively. They both found the halo density profile follows a broken power-law with constant axis ratios. Later, Xue et al. (2015) suggested that, instead of a broken power-law with constant axis ratio, a single power-law with a radially variable axis ratio can also fit the density profile of the halo well. Notably, Xu et al. (2006, 2007) found that the stellar halo is tri-axial from star counts using photometric data.

The advantages of using multi-band photometric data in studies of star counts are that (1) the sampling for the photometric stars is approximately complete within a limiting magnitude; (2) multiple colors can be used to select specific stellar populations and also determine photometric parallax for the stars; and (3) the large number of stellar samples can improve the statistical performance. However, there are a few disadvantages of photometric data, including: (1) it is hard to accurately select stellar populations based on chemical abundances or ages; (2) the dwarf/giant separation is difficult in most broad-band photometric data; (3) the distance estimation based on photometric data in principle suffers from larger uncertainty and potential systematic bias; and (4) binarity may affect the results of the star counts since unresolved binaries usually show slightly different colors than single stars.

In general, spectroscopic data can provide relatively precise stellar parameter estimates, which lead to better distance determinations. Moreover, the stellar parameters derived from spectroscopic data are helpful in better identifying a stellar population. Therefore, star counts

based on spectroscopic data should provide a valuable complement to photometric studies. This approach can play an even more important role in understanding the evolution of the Galaxy via the structures of different stellar populations. However, because it is difficult to sample large regions of sky spectroscopically to a completeness that can easily be achieved by a photometric survey, the stellar density measurements from a spectroscopic survey highly depend on correction of the selection function explicitly or implicitly induced in the survey.

Techniques to correct for the selection function have been discussed in many works. Liu & van de Ven (2012) corrected the selection effects for SEGUE G-dwarf stars by comparing the stellar density in color-magnitude-distance grids between spectroscopic and photometric survey data. A similar method has been used in Zhang et al. (2013) for the measurement of local dark matter density. Bovy et al. (2012) proposed a Bayesian technique to model the stellar density profiles with analytic forms accounting for the selection effects. This technique was later applied to APOGEE survey data (Bovy et al. 2016) and SEGUE K-giant stars (Xue et al. 2015). It is noted that Xia et al. (2016) also used a similar technique with LAMOST survey data.

Recently, the LAMOST survey (Cui et al. 2012; Zhao et al. 2012; Deng et al. 2012) has collected more than 5 million low resolution stellar spectra in the Data Release 3 (DR3) catalog. These stars cover much of the stellar halo to a Galactocentric radius of 80–100 kpc (Liu et al. 2014a) and cover the Galactic anti-center region to as far as 50 kpc. This provides a remarkably large number of distant stars with which to measure the shapes of the outer disk and halo of the Milky Way. In this work, we revise the approach of stellar density measurement from the spectroscopic survey data originally proposed by Liu & van de Ven (2012) and re-organize it in terms of Bayesian statistics. The post-observation selection function is taken into account during the determination of stellar density. Meanwhile, we do not presume any analytic forms for the density profiles of either the disks or the halo. This allows us to better detect the potentially more complicated shapes of the structures, e.g., asymmetric features in the disk or the possibly radial variation of the axis ratio of the halo, in a non-parametric manner.

The paper is organized as follows. In Section 2, we propose a method to derive the stellar density profile along a given line-of-sight with consideration of the selection function. In Section 3, we validate the method using mock data with various selection functions. In Section 4, we apply it to the LAMOST red giant branch

(RGB) stars and demonstrate the derived stellar density map for the disk and halo components. In Section 5, we discuss the smoothing effect of the LAMOST plate due to its large field-of-view and how to deal with the selection function slightly tuned by observations. Finally, we draw brief conclusions in the last section.

2 STELLAR DENSITY PROFILE ALONG A LINE-OF-SIGHT

Given a line-of-sight with Galactic coordinates (l, b), we assume that the photometric data are always complete within the limiting magnitude[1]. We also assume that the selection of the spectroscopic targets is only associated with the color-magnitude diagram. The selection function can be related to color(s) and magnitude(s) in a complicated way. In principle, observations and data reduction may lose some data and thus the final selection function is slightly tuned. However, such a change would not induce substantial selection bias in stellar metallicity, age or kinematics, which are critical for distinguishing various stellar populations. Therefore, the color-magnitude based selection function slightly tuned by the observations and data processing would not induce systematics in stellar populations. It is convenient to assume that this function is continuous so that it is approximately constant in a sufficiently small region around color index c and magnitude m.

2.1 Generality

The probability density distribution (PDF) $p_{\rm ph}(D|c, m, l, b)$ represents the probability of finding a star at distance D at given c, m, l and b. The probability of finding a star in a small volume $(D, D + \Delta D)$ is then written as

$$
\begin{aligned}
\Pr([D, D + \Delta D]|c, m, l, b) = \\
\int_{D}^{D+\Delta D} p_{\rm ph}(D|c, m, l, b)dD \\
\doteq p_{\rm ph}(D|c, m, l, b)\Delta D.
\end{aligned} \tag{1}
$$

On the other hand, $\Pr([D, D + \Delta D]|c, m, l, b)$ is also equivalent to the fraction of stars, which can be calculated from the underlying stellar density profile, located within the small volume. Thus, it can also be written as

$$
\begin{aligned}
\Pr([D, D + \Delta D]|c, m, l, b) = \nu_{\rm ph}(D|c, m, l, b) \\
\frac{\Omega D^2 \Delta D}{\int_0^\infty \nu_{\rm ph}(D|c, m, l, b)\Omega D^2 dD},
\end{aligned} \tag{2}
$$

---

[1] The limiting magnitude of a photometric survey is usually defined by the magnitude at signal-to-noise of 5 or 10 σ.

where $\nu_{\rm ph}$ is the volume stellar density measured from the photometric stars (i.e., the complete dataset), $\Omega D^2 \Delta D$ is the volume element between D and $D + \Delta D$, and Ω is the solid angle associated with the line-of-sight. Combining Equations (1) with (2), we have

$$
\begin{aligned}
p_{\rm ph}(D|c, m, l, b)\Delta D = \\
\frac{\nu_{\rm ph}(D|c, m, l, b)\Omega D^2 \Delta D}{\int_0^\infty \nu_{\rm ph}(D|c, m, l, b)\Omega D^2 dD}.
\end{aligned} \tag{3}
$$

Similarly, for the spectroscopic data, Equation (3) becomes

$$
\begin{aligned}
p_{\rm sp}(D|c, m, l, b)\Delta D = \\
\frac{\nu_{\rm sp}(D|c, m, l, b)\Omega D^2 \Delta D}{\int_0^\infty \nu_{\rm sp}(D|c, m, l, b)\Omega D^2 dD},
\end{aligned} \tag{4}
$$

where $\nu_{\rm sp}$ represents the stellar density measured from counting the spectroscopic survey stars only. Within the small region around c and m, the selection function is assumed to be flat with either the color index or magnitude. Thus, it does not change the probability of finding a star in both the photometric and spectroscopic samples, i.e.,

$$
p_{\rm ph}(D|c, m, l, b) = p_{\rm sp}(D|c, m, l, b). \tag{5}
$$

Then, Equations (3) and (4) can be combined together via Equation (5)

$$
\nu_{\rm ph}(D|c, m, l, b) = \nu_{\rm sp}(D|c, m, l, b)S^{-1}(c, m, l, b), \tag{6}
$$

where

$$
S(c, m, l, b) = \frac{\int_0^\infty \nu_{\rm sp}(D|c, m, l, b)\Omega D^2 dD}{\int_0^\infty \nu_{\rm ph}(D|c, m, l, b)\Omega D^2 dD} \tag{7}
$$

is the selection function at c and m along (l, b).

Then the stellar density profile for all photometric stars along (l, b) can be derived by integrating over c and m

$$
\begin{aligned}
\nu_{\rm ph}(D|l, b) = \\
\iint \nu_{\rm sp}(D|c, m, l, b)S^{-1}(c, m, l, b)dcdm.
\end{aligned} \tag{8}
$$

2.2 Consideration of a Stellar Population

Now consider a stellar sub-population C selected from spectroscopic data based on, for example, the metallicity, luminosity or age. Following Bayes' theorem, the probability of finding a star at D given C, c, m, l and b is written as

$$
\begin{aligned}
p_{\rm ph}(D|C, c, m, l, b) = \\
\frac{p_{\rm ph}(C|D, c, m, l, b)p_{\rm ph}(D|c, m, l, b)}{p_{\rm ph}(C|c, m, l, b)}.
\end{aligned} \tag{9}
$$

If there is no special selection function for C members at given c and m along (l, b), then we would expect that $p_{ph}(C|D, c, m, l, b)$, i.e. the probability about C at given D, c, m, l and b, should be the same as that for the spectroscopic data, $p_{sp}(C|D, c, m, l, b)$. Subsequently, by integrating over D, we obtain that $p_{ph}(C|c, m, l, b) = p_{sp}(C|c, m, l, b)$. Finally, applying Equation (5) to this expression, all the terms of p_{ph} on the right hand side of Equation (9) can be replaced with p_{sp}. Applying Bayes' theorem again to the right hand side, we obtain

$$p_{ph}(D|C, c, m, l, b) = p_{sp}(D|C, c, m, l, b). \quad (10)$$

Therefore, similar to Equation (6), we infer that

$$\nu_{ph}(D|C, c, m, l, b) =$$
$$\nu_{sp}(D|C, c, m, l, b)S^{-1}(C, c, m, l, b), \quad (11)$$

where

$$S(C, c, m, l, b) = \frac{\int_0^\infty \nu_{sp}(D|C, c, m, l, b)\Omega D^2 dD}{\int_0^\infty \nu_{ph}(D|C, c, m, l, b)\Omega D^2 dD}. \quad (12)$$

Consequently,

$$\nu_{ph}(D|C, l, b) = \iint \nu_{sp}(D|C, c, m, l, b)S^{-1}$$
$$(C, c, m, l, b)dcdm. \quad (13)$$

2.3 Estimation of ν_{sp}

In Equations (8) and (13), ν_{ph}, as the density profile for the complete dataset, should be determined. We can estimate it by measuring ν_{sp} from the spectroscopic data. Usually, along a given line-of-sight, spectroscopic observations can only target a few hundred to a few thousand stars. This implies that ν_{sp} has to be established from a relatively small dataset. In order to deal with such a situation, we derive ν_{sp} by using kernel density estimation (KDE). The KDE method can account for both uncertainties in the distance and the relatively small number of stars in a given small region around (c, m). At (l, b), the PDF of D is contributed by all the stars around c and m in terms of

$$p_{sp}(D|c, m, l, b) = \frac{1}{n_{sp}(c, m, l, b)} \sum_i^{n_{sp}(c,m,l,b)} p_i(D), \quad (14)$$

where $p_i(D)$ is the PDF of the distance estimate for the ith star and $n_{sp}(c, m, l, b)$ is the number of spectroscopic stars at a given c, m, l and b. Bringing Equation (14) into Equation (4), we obtain

$$\nu_{sp}(D|c, m, l, b) = \frac{1}{\Omega D^2} \sum_i^{n_{sp}(c,m,l,b)} p_i(D). \quad (15)$$

Note that $\int \nu_{sp}(c, m, l, b)\Omega D^2 dD = n_{sp}(c, m, l, b)$, so they are then canceled in Equation (15).

In principle, a KDE determined ν is a continuous function and thus can return a value at any D. However, if there are very few stars nearby D or the given D is beyond the farthest star, the derived ν is not well constrained by observations and thus may no longer be reliable. Therefore, in practice, the stellar density value is sensible only at the positions with sufficient star samples. For convenience, we only adopt ν for the distance at which the spectroscopic stars are located.

It is worth noting that the stellar density derived for a given (l, b) is averaged over the solid angle of the field-of-view, Ω. Therefore, the derived density at the distance where a star is located does not equal the stellar density for the exact 3D spatial position of the star, but is only an approximation of the latter.

2.4 Estimation of S

Theoretically, S is defined by Equations (7) and (12). Practically, it is not possible to calculate S directly from ν_{ph} and ν_{sp}. Notice that the integration of ν is actually equivalent to the total number of stars at given c, m, l and b. Therefore, S can be evaluated from

$$S(c, m, l, b) = \frac{n_{sp}(c, m, l, b)}{n_{ph}(c, m, l, b)}. \quad (16)$$

For the case of sub-population C, $p_{ph}(C|c, m, l, b) = p_{sp}(C|c, m, l, b)$, and we then have

$$\frac{n_{ph}(C, c, m, l, b)}{n_{ph}(c, m, l, b)} = \frac{n_{sp}(C, c, m, l, b)}{n_{sp}(c, m, l, b)}. \quad (17)$$

Hence, we infer that $S(C, c, m, l, b) = S(c, m, l, b)$.

In order to improve the precision of the computation in Equations (8) and (13), we calculate S^{-1} instead of S in the rest of the paper.

3 VALIDATIONS WITH *GALAXIA* MOCK DATA

Before applying it to real data, we validate the method using a mock star catalog. We select a 20-square degree area toward the north Galactic pole from the *Galaxia* simulation (Sharma et al. 2011) and obtain 10 987 mock stars with K magnitude brighter than 15 mag. The *Galaxia* mock catalog contains distances, stellar parameters and ages, which can be used to test whether the derived density profiles are precise for some specially selected stellar populations.

We test the method with two different selection functions, which are discussed separately in the following subsections.

3.1 Tests with Selection Function T1

Before the tests, we construct the selection function to mimic the real spectroscopic survey. The selection of mock spectroscopic stars is based on the infrared (IR) bands J and K. First, we do not induce any selection in the color index $J - K$. Second, the selection function for K is divided into two parts at $K = 13$ mag. At $K < 13$ mag, we arbitrarily select stars, while at $K > 13$ mag, we apply a flat selection function to keep the mock spectroscopic stars evenly distributed along K. This mimics the strategy that the targeting is more biased to bright rather than faint sources. We denote such a selection function as **T1**.

We randomly draw 50 groups of mock spectroscopic samples according to **T1** and derive the density profile for each group of samples. As a sample, the mock spectroscopic stars from one of the 50 groups are highlighted with red circles in Figure 1, in which the black dots indicate the $J - K$ vs. K distribution for the complete dataset. The right panel shows the normalized distributions of K for the mock spectroscopic stars (red line) and the complete dataset (black line). It can be seen that the red line shows a trend that is similar to the black one at $K < 13$ and then becomes flat at $K > 13$. This reflects the definition of selection function **T1**. The normalized distributions of $J - K$ in the bottom panel demonstrate that even though we do not apply any selection function to the color index, the normalized distribution of $J - K$ for the mock spectroscopic stars (red line) is not exactly same as that for the full mock dataset (black line). This is a natural result of the selection function in K. There are more red stars in the regime $K > 13$ than in $K < 13$. The selection strategy for $K > 13$ reduces the sampling rate of fainter and redder stars in the mock spectroscopic stars. This leads to a slightly different distribution in $J - K$ with the complete dataset.

Figure 2 shows how the selection function S is calculated for the same group of mock spectroscopic stars as in Figure 1. In the left panel, color represents the number density of stars for the complete sample, i.e., $n_{\rm ph}(J - K, K, l, b)$. The size of each bin is $\Delta(J - K) = 0.1$ and $\Delta K = 0.25$. The plot exhibits a high density at $J - K \sim 0.8$ and $K \sim 14$, which is most likely contributed by K/M dwarf stars. Applying the selection function **T1**, the map of the number density of mock spectroscopic stars, $n_{\rm sp}(J - K, K, l, b)$, is displayed in the middle panel. The right panel shows the map of $S^{-1}(J - K, K, l, b)$, which is obtained by dividing values in the left by those in the middle panel of Figure 2. It is apparent that $S^{-1}(J - K, K, l, b)$ is roughly separated

**Table 1** The Test Populations Selected from the *Galaxia* Mock Data

Population	Criteria	Number for **T1**	Number for **T2**
S1	all spectral types	2230±29	1943±22
S2	$M_K < -2$	40±5	23±3
S3	[Fe/H]< -0.3	979±23	998±22
S4	[Fe/H]> -0.3	1244±29	946±17

into two platforms; one is at $K < 13$ with a value of about 5 and the other is at $K > 13$ with a value of about 10. This pattern implies that the selection function **T1** does not induce selection effects in $J - K$ at each K bin, although the overall distributions in $J - K$ (as shown in the bottom panel of Figure 1) are slightly different than the complete dataset.

In order to mimic a more realistic situation, i.e., study the stellar density via a selected stellar population, we select four different populations defined in Table 1 for tests. Table 1 also lists the median numbers and scatter of the 50 groups of arbitrarily selected mock spectroscopic samples based on **T1** in the 3rd column. It can be noted that for the sub-population **T1S2**, which is selected with $M_K < -2$, only about 40 stars are involved in the determination of the stellar density profile. This is a good sample to test the performance of the density estimates for a small dataset.

Figure 3 shows the distributions of star counts without a selection function correction for the complete set (black lines) and for one of the 50 groups of the mock spectroscopic samples (red lines). The four panels show the cases for the four different sub-populations, which demonstrate that the selection function **T1** does distort the spatial distributions of the mock spectroscopic stars compared to those of the complete samples.

The panels in Figure 4 show the derived median stellar densities and their 1σ dispersions (blue circles with error bars) over the 50 groups of mock spectroscopic stars at the spatial positions in which the mock spectroscopic stars are located. In contrast, the corresponding stellar densities for the complete dataset, i.e., the "true" profiles, are superposed as black solid lines in the panels. In the bottom of each panel, the relative $\Delta\nu/\nu$ is also displayed. It is seen that the derived stellar density profiles from the mock spectroscopic data are perfectly consistent with the "true" profiles. This confirms that when the selection function is very simple, the stellar density can be reconstructed well. In particular for the case **T1S2**, the accuracy of derived stellar density is still very high even though the sample contains only about 40 stars (see the top-right panel of Fig. 4).
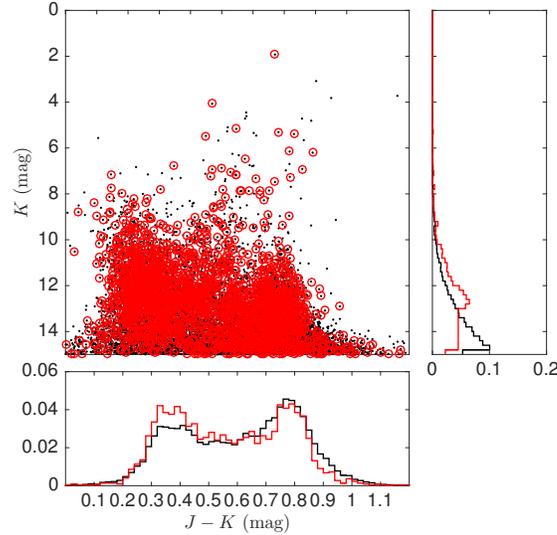
Fig. 1 The black dots show the distribution in $J - K$ vs. K for the complete sample of mock stars generated from *Galaxia*. The red circles indicate mock spectroscopic stars according to the selection function **T1**. The right panel shows normalized distributions of K for the complete and selected mock stars with black and red lines, respectively. The bottom panel indicates the normalized distributions of $J - K$ for the two populations with the same colors as in the right panel.
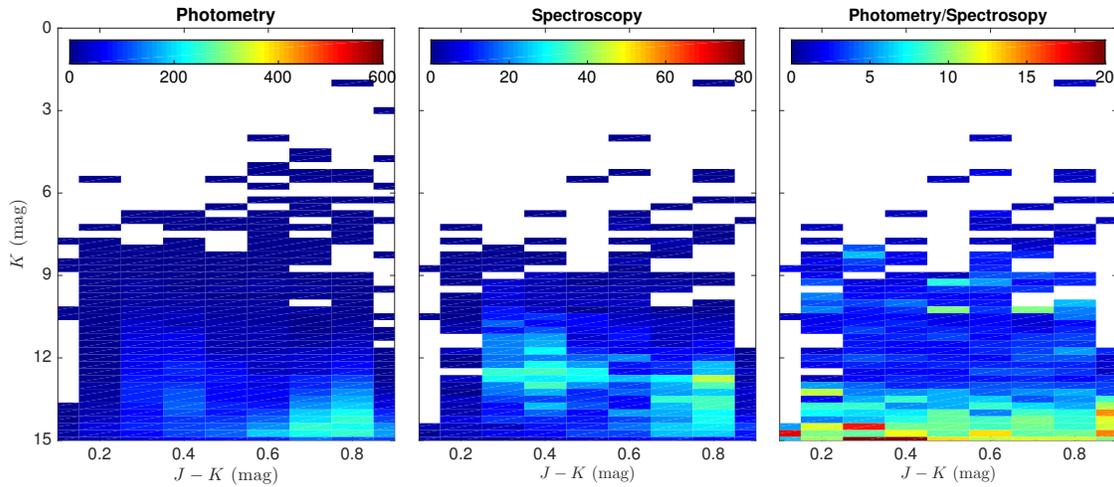

Fig. 2 The left panel shows the number density of stars in a $J - K$ vs. K diagram for the complete set of mock stars. The colors code the number of stars within each small $J - K$ vs. K bin. The middle panel shows a similar map of the number density in the $J - K$ vs. K plane for the mock spectroscopic stars based on selection function **T1**. The right panel displays the map of $S^{-1}(J - K, K)$ for **T1**. The color scales are indicated in colorbars located above the panels.

3.2 Tests with Selection Function T2

We now change to a more complicated selection function. Because the LAMOST survey selects its targets using optical-band photometry, the bright/faint limiting magnitude cuts in an optical band lead to a sloped cut in K magnitude. Thus a rectangle region selected in the optical color-magnitude diagram turns out to be a wedge region in IR bands.

Figure 5 provides the selection function for a sample plate[2] of the LAMOST survey. The middle panel shows that the spectroscopic stars are distributed in a wedge re-

---

[2] A plate in the LAMOST survey refers to an observed pointing along a line-of-sight. LAMOST has a field-of-view of 20 square degrees. Thus a plate covers the same solid angle. Because LAMOST contains 4000 fibers, a plate can simultaneously observe 4000 objects at most. In practice, a typical number of targets in a plate is around 2500, with some fibers dedicated to observing the sky and flux standard stars, plus a few fibers that are broken and a small fraction of fibers that are not used.
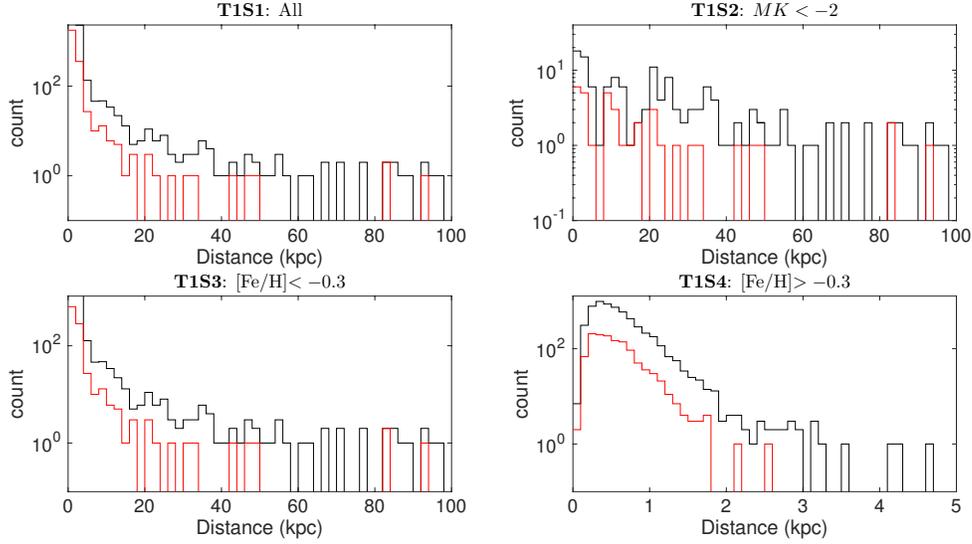
Fig. 3 The four panels display histograms of D for the mock spectroscopic (*red lines*) and complete (*black lines*) datasets. The top-left, top-right, bottom-left and bottom-right panels display histograms for the populations **T1S1**, **T1S2**, **T1S3** and **T1S4**, respectively.
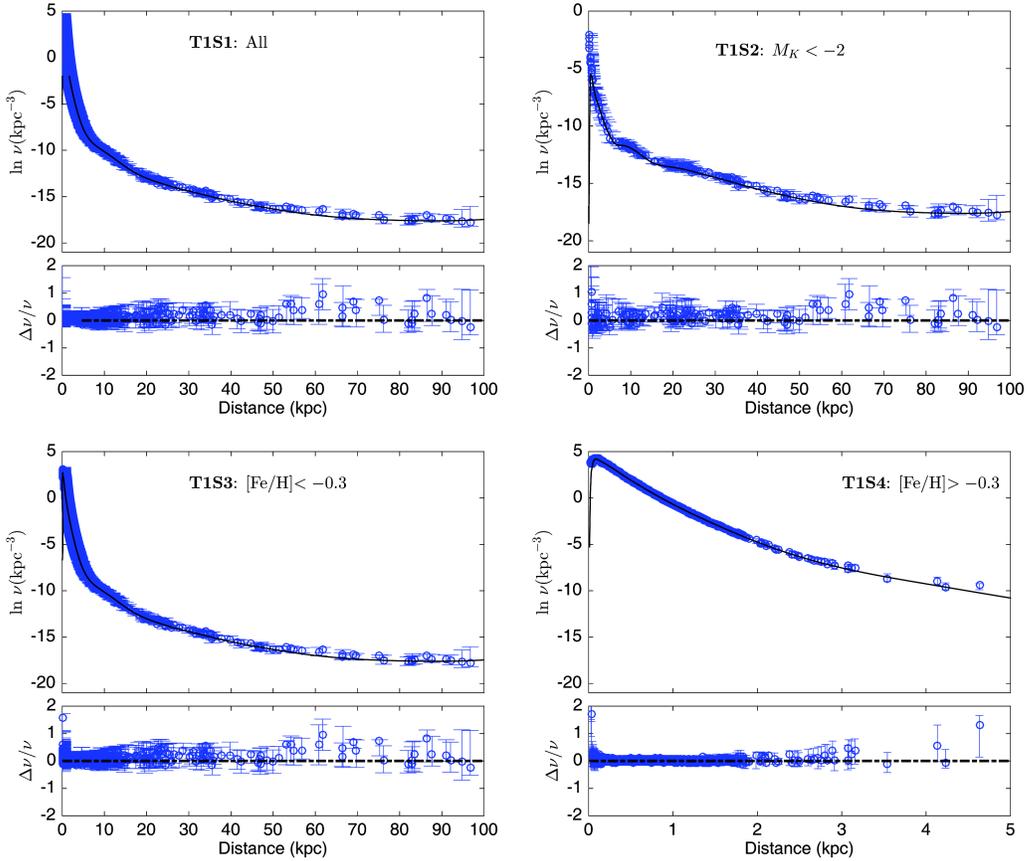
Fig. 4 The four panels plot the derived density profiles versus distances. The top-left, top-right, bottom-left and bottom-right panels display profiles derived from the mock spectroscopic stars (*blue circles* with error bars) and the "true" profile from the complete dataset (*black lines*) for populations **T1S1**, **T1S2**, **T1S3** and **T1S4**, respectively. The relative residual density, $\Delta\nu/\nu$, for each case is displayed in the bottom of the corresponding panel.
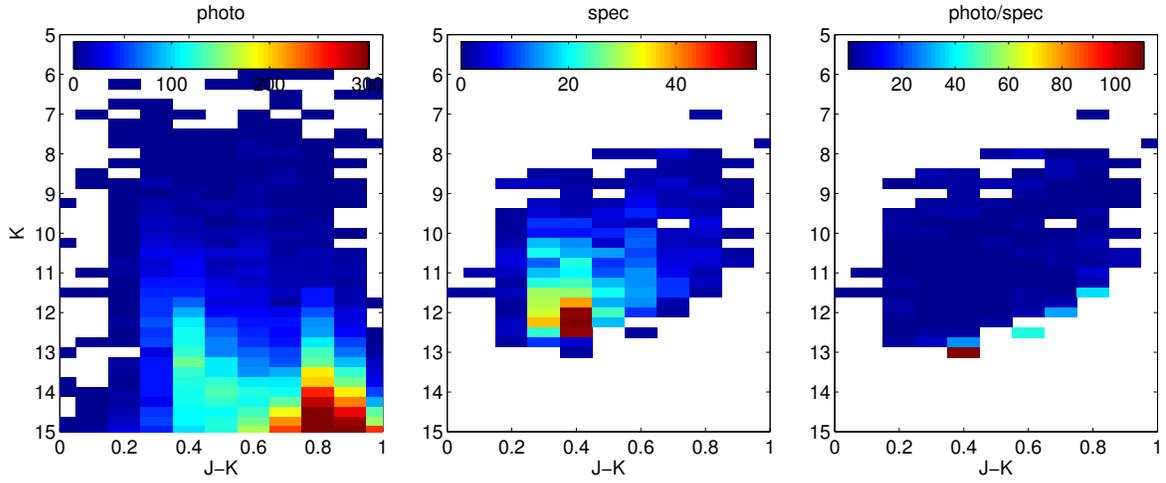
Fig. 5 A sample of the selection function for the LAMOST plate "HD103843N315834V01." The left panel is the $J - K$ vs. K diagram for photometric data in the same line-of-sight. The middle panel displays spectroscopically selected data. The right panel shows S^{-1}.
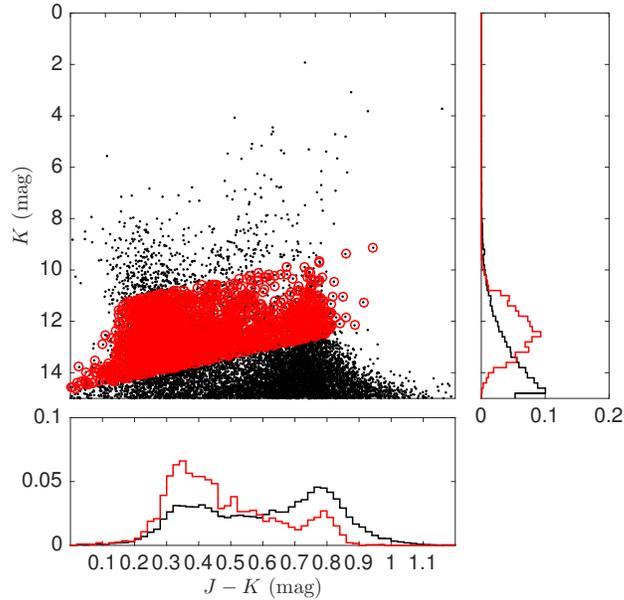
Fig. 6 Similar to Fig. 1, but with the selection function **T2**.

gion in the $J - K$ vs. K diagram. Thus the selection function S^{-1} in the right panel also displays a similar shape.

To mimic such a selection function, we randomly select stars satisfying the following criterion

$$-3(J - K) + 12 < K < -3(J - K) + 15. \quad (18)$$

We denote this selection function as **T2**. Figure 6 shows a randomly drawn sample based on **T2** in the $J - K$ vs. K diagram, in which the selected spectroscopic stars (red circles) are distributed in a wedge area. The median

numbers of stars and their scatters over the 50 groups of arbitrarily drawn mock spectroscopic stars based on **T2** are separately listed in the last column of Table 1 for the four sub-populations.

Figure 7 features the density of stars in the color-magnitude diagram for the same sample group of mock spectroscopic data as in Figure 6. It can be seen that the selected stars exhibit a tilted distribution in the $J - K$ vs. K plane, as does the map of S^{-1} according to the definition of **T2**.
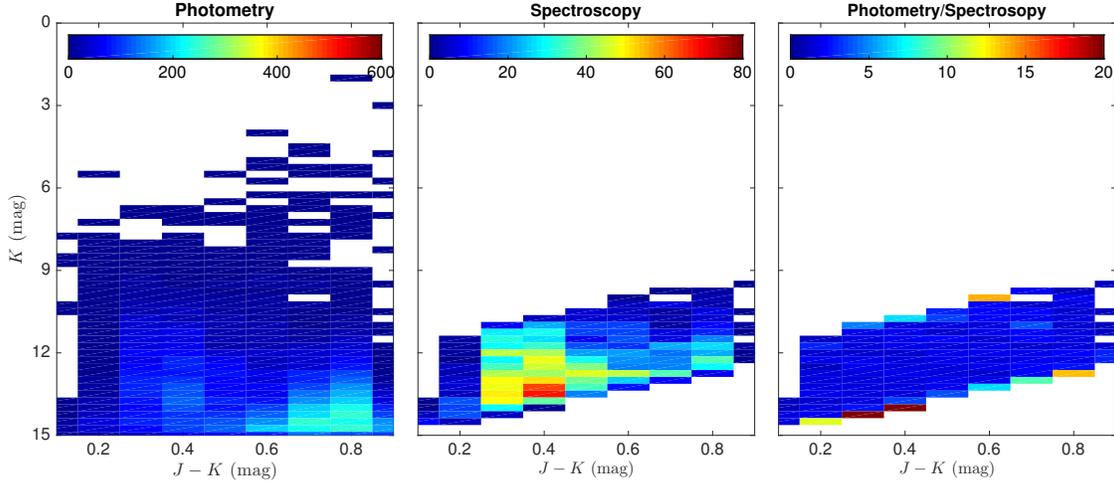
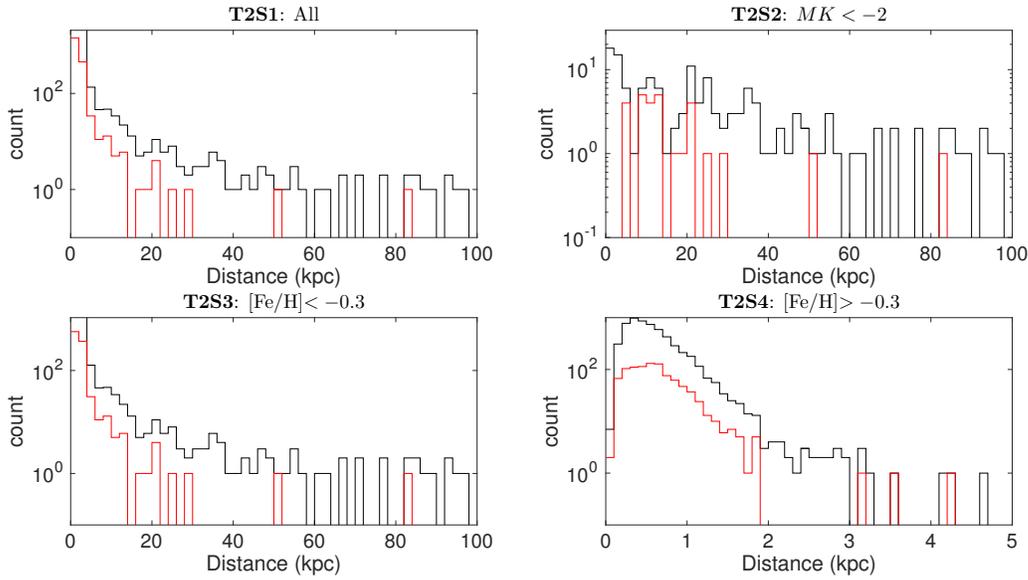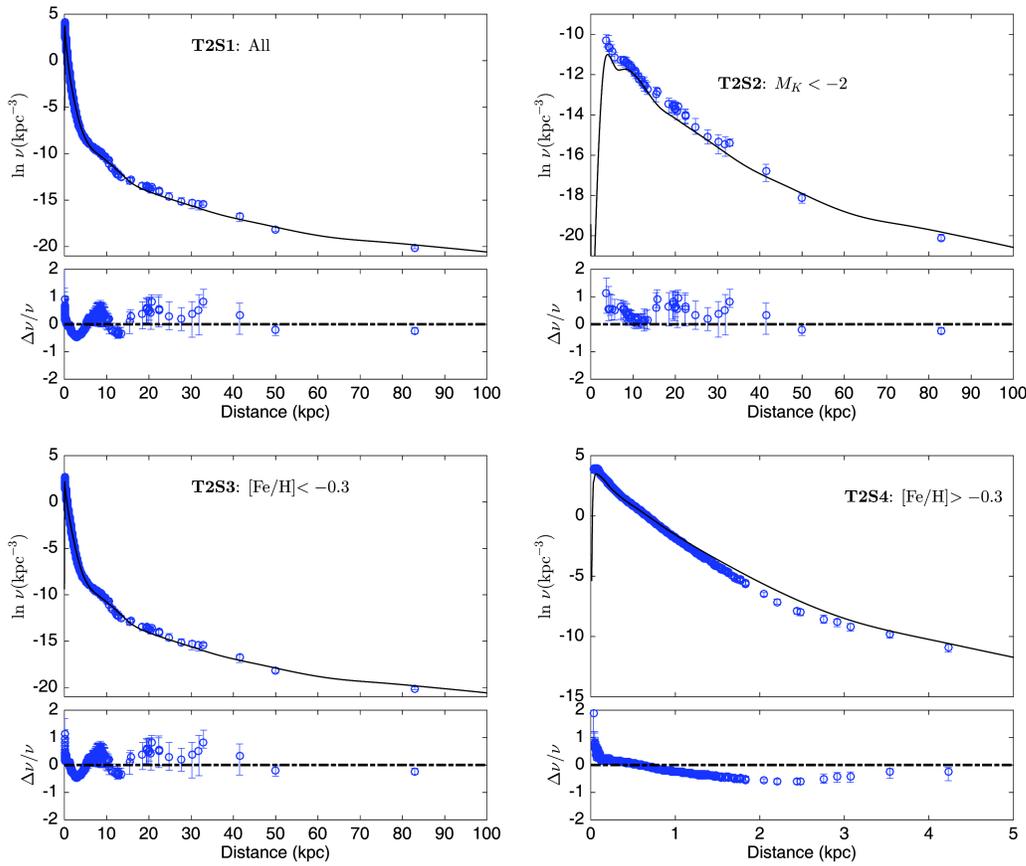**Fig. 7** Similar to Figure 2, but with the selection function **T2**.

Fig. 8 Similar to Figure 3, but with the selection function **T2**.

Figure 8 displays the spatial distributions of star counts without selection function correction for the four sub-populations from one of the 50 groups of mock spectroscopic stars (red lines). In contrast, the black lines indicate the distributions of stars for the corresponding complete datasets. It can be seen that for **T2S1**, **T2S2** and **T2S3**, distances further than 30 kpc are significantly under-sampled.

Figure 9 shows the derived median stellar density profiles and their 1σ dispersions over the 50 randomly drawn groups (blue circles with error bars). In general, the derived profiles are quite consistent with their corresponding "true" values within an uncertainty of $\Delta\nu/\nu <$

1. However, a few systematic biases of about $\Delta\nu/\nu \sim 0.5 - 1$ appear in the derived density profiles: one occurs at the distance of $\sim 20 - 30$ kpc for **T2S1**, **T2S2** and **T2S3**; the other is at $D \sim 2.5$ kpc for **T2S4**.

Considering that properties of Galactic spatial structures, such as the scale height/length of the disks and the power-law index of the stellar halo, are mostly measured in logarithmic density, such systematic biases may not lead to severe effects. For instance, a simple exponential fit within 0.3–2.5 kpc gives a scale height of 219 pc with the true density profile of **T2S4** (black line in the bottom-right panel of Fig. 9), while it decreases to 191 pc obtained from the corresponding derived stellar density

Fig. 9 Similar to Fig. 4, but with the selection function **T2**.

profile (blue circles in the same panel). This means that the derived density based on **T2** may induce a systematically smaller exponential scale height of only $\sim 13\%$.

For a realistic survey such as LAMOST, many plates are overlapped with each other, and the selection function for different plates varies. Then the possible systematic bias of the stellar density measurement in one plate, as shown in Figure 9, may be compensated by other overlapped plates. Consequently, cumulation of the systematic bias from each individual plate may not induce an overall systematic bias to the final density profiles, but may increase random uncertainty in the resulting density profiles.

4 THE STELLAR DENSITY PROFILE OF THE MILKY WAY

4.1 The Selection Function Used By the LAMOST Survey

In general, the targeting strategy of the LAMOST survey is rather simple. Carlin et al. (2012) designed an elegant targeting algorithm, which tries to make the distribution of target stars flat in color-magnitude planes. However,

in practice, this technique has not been fully used for a few reasons. First, the dynamical range of magnitudes observed in a single LAMOST plate is limited to only about 3 magnitudes so that bright spectra will not saturate or cross-talk between fibers. Second, the actual limiting magnitude of $r \sim 18$ mag is brighter than the designed goal by at least 2 mag. Both situations lead to insufficient source stars for targeting. For instance, the bright plates used in the LAMOST survey cover $14 < r < 16.8$ mag, which may only contain less than 200 stars per square degree at high Galactic latitudes. Such a sparse stellar sampling is not sufficient to apply the selection function designed by Carlin et al. (2012) for targeting, because the targeting algorithm needs far more source stars than available fibers to achieve a flat distribution of targets in the color-magnitude plane.

Finally, the LAMOST survey separates the targets into different plates with different ranges of magnitudes in each line-of-sight. The *VB* plates cover $9 < r < 14$, the *B* plates cover $14 < r < 16.8$, the *M* plates cover $r < 17.8$ and the *F* plates cover $r < 18.5$. For the *VB* and *B* plates, no specific selection function is applied, i.e., the stars are randomly selected. For the *M* and *F* plates,
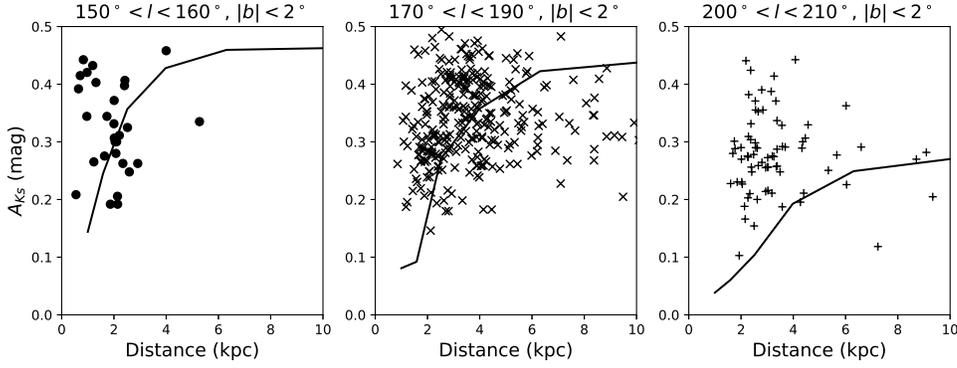
Fig. 10 The circles, crosses and plus symbols indicate the RGB stars with $|b| < 2°$ along the lines-of-sight $150° < l < 160°$, $170° < l < 190°$ and $200° < l < 210°$, respectively. Three extinction-distance relations at $l = 150°$ (line in the left panel), $l = 180°$ (line in the middle panel) and $l = 210°$ (line in the right panel) queried from the 3D extinction map of Green et al. (2014, 2015) are superposed on the observed data.

we only apply the selection function in r magnitude, and leave the selection in color index to be arbitrary.

This simplified selection function is applied to all of the sky area observed by LAMOST except for the Galactic anti-center region, which covers $150° < l < 210°$ and $|b| < 30°$. For the main survey regions, the UCAC4 catalog (Zacharias et al. 2013) with $r < 14$ mag and PanSTARRS-1 catalog (Schlafly et al. 2012) with $r > 14$ are adopted as the source catalogs for targeting. For the Galactic anti-center region, the selection strategy follows Yuan et al. (2015), and uses the 2MASS, UCAC4 and Xuyi survey catalogs (Liu et al. 2014b) as the sources for targeting.

Because the different source catalogs are based on different photometric systems, they cannot be unified unless they are cross-calibrated with each other. To avoid the complicated calibrations between at least four different systems, we finally adopt the 2MASS catalog, which covers most of the LAMOST observed stars, as the photometric dataset. Indeed, for LAMOST K giant stars, about 96% of them are within $K < 14.3$ mag, which is the limiting magnitude of the 2MASS catalog. Subsequently, we use the $J - K$ vs. K map to derive the selection function S for each observed plate in the LAMOST survey. The selection function S^{-1} for all LAMOST DR3 plates and the *python* codes used for measurement of the stellar density profiles can be found at *https://github.com/liuchaonaoc/LAMOST_density*.

4.2 LAMOST RGB Stars

In order to derive the density profiles for the stellar disk and halo of the Galaxy, we select RGB stars. Although RGB stars are not the standard candles that red clump stars are, the accuracy of distance using the former ones

is still around 20%–30% (Liu et al. 2014a; Carlin et al. 2015), which is sufficient for mapping Galactic structure in the outer regions. Moreover, RGB stars are intrinsically brighter and occur in stellar populations with a broader range of metallicity than red clump stars. Therefore, they are suitable tracers for the outer disk and halo, which are dominated by metal-poor stars.

For LAMOST survey data, we adopt stellar parameters, i.e., the effective temperature, surface gravity and metallicity, provided by the LAMOST pipeline (Wu et al. 2011, 2014; Luo et al. 2015). K giant stars, including the RGB and red clump stars, are selected based on criteria proposed by Liu et al. (2014a). Red clump stars are identified by Wan et al. (2015) and Tian et al. (2016). The RGB stars are then selected from K giants by excluding the red clump stars.

The absolute magnitudes at K_s band, M_{Ks}, for the RGB stars are estimated following the method suggested by Carlin et al. (2015). Interstellar extinction for individual RGB stars is estimated using the Rayleigh-Jeans color excess (RJCE) approach introduced by Majewski et al. (2011) and later revised to incorporate WISE bandpasses by Zasowski et al. (2013). Subsequently, the distances to RGB stars are estimated by combining the apparent K_s magnitude, the absolute magnitude M_{Ks} and the RJCE extinction A_{Ks}.

The RJCE extinction is compared with the 3D extinction map provided by Green et al. (2014, 2015), who derived the 3D dust map from PanSTARRS-1 data.

Figure 10 shows the RJCE extinction vs. distance for the RGB stars located at ($150° < l < 160°$, $|b| < 2°$), ($170° < l < 190°$, $|b| < 2°$) and ($200° < l < 210°$, $|b| < 2°$) using circles, crosses and pluses, respectively. The Green et al. mean extinction-distance relations at similar lines-of-sight are superposed in the same figure.

We see that the RJCE extinctions for individual RGB stars are roughly consistent with Green's 3D extinction map within about 0.2 mag, which may lead to only about 10% uncertainty in distance. Therefore, the distances for the RGB stars should not be substantially affected by extinction, even in the Galactic mid-plane.

In the following subsections we demonstrate two samples: (1) RGB stars with all ranges of metallicity as tracers for the Milky Way structures, especially for the Galactic disk, and (2) metal-poor RGB stars as tracers for the stellar halo.

4.3 The Stellar Disk

We select 21 954 RGB stars with $M_{Ks} < -3.5$ mag and distances larger than 0.5 kpc as probes to map the Galaxy. Because these samples are dominated by metal-rich stars, they mostly trace the Galactic disk. The volume completeness for these RGB stars is within 40 kpc. The volume completeness in this paper is defined such that the range of luminosity for the stars does not significantly change within the volume. Plate by plate, we apply the method to derive stellar density and finally obtain density values for the distances at which sample stars are located.

Table 3 lists the first three rows of the samples. The columns in the table are described in Table 2. The full RGB data can be found at *https://github.com/liu chaonaoc/LAMOST_density*.

Note that the samples contain about 20% of stars that have been observed more than twice. Because the duplicated stars were observed in different plates and selection effects for each observation can be considered separately, the duplication would not affect stellar density estimation. In fact, these data can be used to test the internal uncertainty, or the precision, of the distance and stellar density estimations. The left panel of Figure 11 shows the distribution of relative scatter in the distance estimates (solid line) of stars observed multiple times. The scatter is defined by the standard deviation of the distance. The relative scatter is the scatter divided by the mean distance estimate. The distribution of relative scatter in distance is fitted with a Gaussian (dashed line) with the best fit parameter of $\sigma = 0.038$, while the median value of the distribution is 0.028. This means that the internal uncertainty of distance estimates for the RGB stars is around 3%–4%.

The right panel of Figure 11 shows the distribution of relative scatter in stellar density (solid line) for the duplicated stars. The definitions of scatter and relative scatter are similar to those for the distance estimates. The dashed

line stands for the best fit Gaussian with $\sigma = 0.37$. The median value for the distribution is 0.25. This is better than the performance of the mock data test described in Sections 3.1 and 3.2, which may reach $\Delta\nu/\nu > 0.5$ in the worst cases. The difference in the derived stellar densities for multiply-observed stars originates from two possible channels: (1) the different selection corrections from the different plates observing the same star and (2) the different averaged sky areas since the plates observing the same star may not be completely overlapped with each other. The small dispersion of the relative scatter implies that these differences are very small for most of the stars.

Figure 12 demonstrates the averaged stellar density map in the R–Z plane[3]. The bin size is 0.5×0.5 kpc. Because this population is dominated by metal-rich stars, it displays a remarkable tomographic map of the Galactic disk from $R \sim 4$ to ~ 20 kpc. Two prominent features are seen in this figure. First, the Monoceros ring, unveiled by Newberg et al. (2002) and later highlighted in SDSS star counting (Jurić et al. 2008) and metallicity distribution mapping (Ivezić et al. 2008), does not show up in our map of the whole Galactic outer disk with LAMOST RGB stars. Second, the iso-density contours (black lines) show moderate north-south asymmetry in the outer part of the disk starting from $R \sim 12$ kpc. The stellar density above the mid-plane is larger than that below at given radii. This is essentially consistent with Xu et al. (2015). The asymmetry about the mid-plane may also be the result of the warp. If that is the case, the stellar disk may bend up to some extent in the Galactic anti-center region between $150° < l < 210°$, in which the LAMOST disk data are mostly concentrated. Both features are very interesting and will be further discussed in an upcoming paper (Wang et al. in preparation).

In Figure 13, black circles show the stellar surface density derived by integrating the volume density over Z in each R slice. In order to avoid issues due to a lack of data in the southern Galactic hemisphere, we use $|Z|$ instead of Z. That is, the mean stellar density at given $|Z|$ is averaged over the values at both $+Z$ and $-Z$. Only $|Z| < 40$ kpc is included in the integration. The width of each R slice is 1 kpc. For some bins without stellar density estimates, interpolated values are used. The red and blue dashed lines are best-fit exponential disk and power-law halo profiles, respectively, and the green dot-dashed line is the sum of the two profiles. We find that the surface density profile can be fitted well with an exponential disk with scale length of 1.6 ± 0.1 kpc and a power law

---

[3] R and Z are Galactocentric cylindrical coordinates with adopted solar position at $R_0 = 8$ kpc and $Z_0 = 0.027$ kpc (Chen et al. 2001).

Table 2 Explanation of Data Columns in the Catalogs used for Disk and Halo Density Mapping

Column name	Type	Unit	Comments
obsid	integer		The identifier of the LAMOST spectra
RA	double	degree	Right ascension in epoch J2000.0
DEC	double	degree	Declination in epoch J2000.0
$T_{\rm eff}$	float	K	Effective temperature from LAMOST DR3 catalog
$\log g$	float	dex	Surface gravity from LAMOST DR3 catalog
[Fe/H]	float	dex	Metallicity from LAMOST DR3 catalog
$v_{\rm los}$	float	km s^{-1}	Line-of-sight velocity from LAMOST DR3 catalog
M_{Ks}	float	mag	K_s-band absolute magnitude from Carlin et al. (2015)
M_{Ks} lower error	float	mag	-1σ error of M_{Ks}
M_{Ks} upper error	float	mag	$+1\sigma$ error of M_{Ks}
A_{Ks}	float	mag	K_s-band interstellar extinction derived from RJCE (Majewski et al. 2011, Zasowski et al. 2013)
K_s	float	mag	K_s-band apparent magnitude from 2MASS catalog
Distance	float	kpc	Distance estimated from M_{Ks}, A_{Ks} and K_s
Distance lower error	float	kpc	-1σ error of distance
Distance upper error	float	kpc	$+1\sigma$ error of distance
Z	float	kpc	Vertical distance from the star to the Galactic mid-plane
R	float	kpc	Galactocentric radius in cylindrical coordinates
r	float	kpc	Galactocentric radius in spherical coordinates
$\ln \nu$	float	ln kpc^{-3}	Logarithmic stellar density at the position of the star

Table 3 The Disk-Like RGB Stars Selected from the LAMOST DR3 Catalog

obsid	RA	DEC	$T_{\rm eff}$	$\log g$	[Fe/H]	$v_{\rm los}$	M_K	A_K	K	Distance	Z	R	r	$\ln \nu$
	(°)	(°)	(K)	(dex)	(dex)	(km s^{-1})	(mag)	(mag)	(mag)	(kpc)	(kpc)	(kpc)	(kpc)	ln(pc^{-3})
75606117	337.46744	+6.01526	4829	1.11	−1.66	−3	$-3.96^{+0.42}_{-0.44}$	0.148	10.646	$7.79^{+1.75}_{-1.37}$	−5.22	8.26	9.77	−11.40
269507178	3.66002	+55.72871	4159	1.47	−0.47	−93	$-3.82^{+0.24}_{-0.36}$	0.252	9.721	$4.55^{+0.82}_{-0.48}$	−0.51	10.86	10.88	−6.25
51211049	257.70419	+18.38338	4207	2.64	−0.51	−33	$-3.32^{+0.38}_{-0.42}$	0.072	9.687	$3.86^{+0.82}_{-0.62}$	1.97	5.82	6.14	−7.51

Notes: Here we only list the first three rows; the complete table is in the online file (*https://github.com/liuchaonaoc/LAMOST_density*).
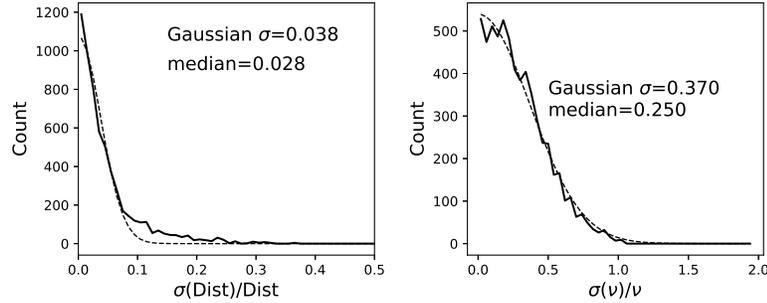

Fig. 11 In the left panel, the solid line shows the distribution of relative scatter in distance for the repeatedly observed RGB stars. The dashed line is the best fit Gaussian for the distribution of the scatter. The σ of the Gaussian is 0.038 and the median of the scatter is 0.028. In the right panel, the solid line displays the relative scatter in ν for the repeatedly observed RGB stars. The dashed line is the best fit Gaussian with $\sigma = 0.37$. The median of the scatter of ν is 0.25.

halo with index of 2.4 ± 0.3. Note that this power-law index is not equivalent to that derived from spherical coordinates, since it is derived from the surface density as a function of cylindrical radius R.

The fraction of stars contributed by the halo, as derived from our models of the surface density, is indicated as the black line which is scaled by the right y-axis in the figure. It shows that, at $R = 14\,{\rm kpc}$, about half of the surface density is contributed by the disk. Therefore, no evidence is found to support that there is a truncation at around $14\,{\rm kpc}$. Moreover, at $R = 19\,{\rm kpc}$, the disk still contributes $\sim 10\%$ to the total stellar surface density, which cannot be negligible because the data points (black circles) are clearly larger than the halo model (blue lines)
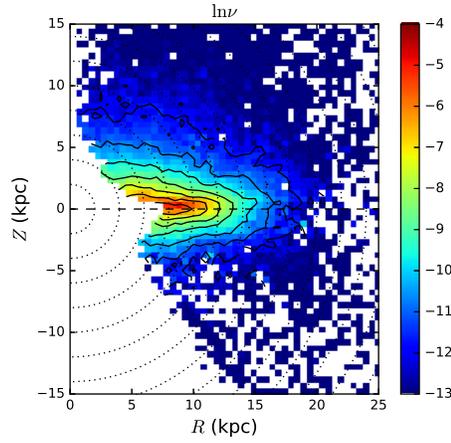
Fig. 12 The figure shows the averaged stellar density map for the RGB stars with all range of metallicity, $M_{Ks} < -3.5$ and $D > 0.5$ kpc. The colors code mean $\ln \nu$ for each R–Z bin. The black contours indicate the iso-surface lines at $\ln \nu = -12, -11,$ $-10, -9, -8$ and -7 from outside to inside, respectively. The dotted circles indicate the Galactocentric radii of 2, 4, ..., 24 kpc from left to right, respectively. The black horizontal dashed line indicates the plane of $Z = 0$.
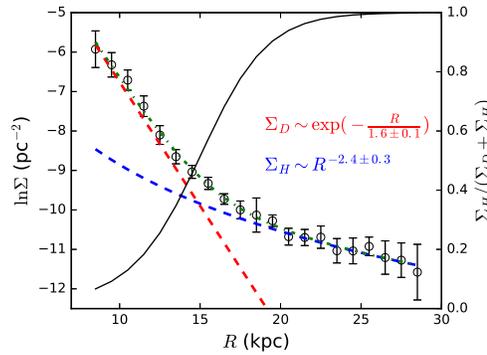
Fig. 13 The black circles display the radial stellar surface density profile integrated over Z for the RGB stars. The bin size of R is 1 kpc. The red and blue dashed lines stand for the best fit exponential disk and power-law halo profiles, respectively. The scale length for the disk is 1.6 ± 0.1 kpc and the power-law index for the halo is 2.4 ± 0.3. The green dot-dashed line is the sum of the two model profiles for the halo and disk. The black line, which is scaled by the right y-axis, shows the fraction of the halo relative to the total surface density as a function of R.
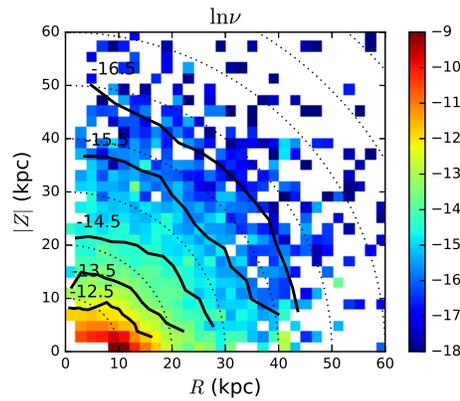
Fig. 14 The figure shows the averaged stellar density map for halo RGB stars in the R–Z plane. The colors codes the mean $\ln \nu$ in each R–Z bin. The black contours indicate the iso-surface lines at $\ln \nu = -16.5, -15.5, -14.5, -13.5$ and -12.5 from outside to inside, respectively. The dotted circles indicate Galactocentric radii of 10, 20, ..., 80 kpc from bottom-left to top-right, respectively.

around the radius. In other words, the disk extends to as far as 19 kpc. Beyond this radius, the observed surface density smoothly transitions to a pure halo profile. The perfect fitting with an exponential disk model implies that the disk displays neither truncation, nor a break, nor an up-bending feature.

The extended larger size of the disk is not likely to be an artificial effect due to incorrect interstellar extinction, according to the discussion in Section 4.2. On the other hand, Carlin et al. (2015) pointed out that the distances may be systematically overestimated for α-enhanced giant stars by about 15%. However, Hayden et al. (2015) show that it is the α-low giant stars that dominate the outer disk. Consequently, the systematic bias in distance due to the α abundance may not significantly affect the distance for the outer disk RGB stars. Therefore, we suggest that the largely extended disk without truncation should be real.

The Galactic disk measured in this work is significantly larger than that in some other works, which claimed that the Milky Way stellar disk truncates at around 14–15 kpc (e.g. Robin et al. 1992; Minniti et al. 2011). Carraro (2015) pointed out that the lines-of-sight investigated by these works are very close to $b \sim 0$ and may not be aligned with the disk plane in the outskirts due to the existence of the warp. In this case, at some radius the in-plane lines-of-sight no longer probe the disk, leading to a false truncation at that radius. It is worth noting that Carraro et al. (2010) and Feast et al. (2014) found some young stars at around 20 kpc from the Galactic center, which is in agreement with our suggestion that the disk extends to such radii. We also point out that Xu et al. (2015) suggested the disk may extend all the way to large radii. More quantitative analysis will appear in an upcoming work by Wang et al. (in preparation).

4.4 The Stellar Halo

In this section we select 5171 halo-like RGB stars with [Fe/H]< -1 and $M_{Ks} < -4$ from the LAMOST DR3 catalog to map the stellar density profile for the stellar halo. The cut in K_s-band absolute magnitude ensures that the data are roughly complete within 50 kpc from the Galactic center. The first three rows of the sample are listed in Table 4. The columns in the table are explained in Table 2. The full catalog can be found at *https://github.com/liuchaonaoc/LAMOST_density*. We then derive stellar densities at the spatial positions at which the halo-like RGB stars are located.

Figure 14 shows the averaged stellar density map in the *R–Z* plane for the halo population. The metal-poor

RGB stars probe the shape of the halo well within a Galactocentric radius of 50 kpc. The contours of $\ln \nu = -12.5$ and -13.5 in the figure show clear oblate shapes within 20 kpc. The iso-density contour at -14.5 displays a slightly larger axis ratio, although it is still oblate. Beyond 30 kpc, the contours of -15.5 and -16.5 are roughly spherical. Although Xue et al. (2015) claimed that a single power-law halo with variable axis ratio can fit their data better than the broken power-law, this work is the first time, to the best of our knowledge, that the variable axis ratio from inner to outer halo is directly illustrated in the tomographic map.

We further demonstrate how the broken power-law appears with the assumption of a constant axis ratio. Figure 15 shows $\ln \nu$ as a function of r for the points shown in Figure 14 with different fixed values of the axis ratio. The top panel presumes the axis ratio $q = 1.0$. Thus r is the same as the Galactocentric radius in spherical coordinates. The $\ln \nu$ profile is fitted with a broken power-law. Within $\sim 40 \pm 2.8$ kpc, the best fit power-law index is -2.9 ± 0.1, but beyond this it is down to -5.0 ± 1.2. In the bottom panel, q is set to be 0.75. In this case, the broken power-law profile shown in the top panel almost disappears. Instead, it shows that the two power-law indices in the inner and outer halos are similar, i.e., a single power-law may be better to fit the data. This means that the presumption of the axis ratio can significantly change the radial profile of $\ln \nu$. More quantitative study about the structure of the stellar halo with LAMOST RGB samples can be found in Xu et al. (2017).

5 DISCUSSIONS

5.1 The Plate-Wide Smoothing Density Profile

In this work, we derive the density profile along a line-of-sight for each LAMOST plate. This implies that the stellar density estimate is smoothed over the solid angle of 20 square degrees, since the field-of-view of the LAMOST telescope is 5 degrees in diameter. Note that the stellar density profile may dramatically change from one end to another in such a large sky area, especially when the distance is large. As a consequence, the plate-wide smoothing density profile may smear out or blur the structure of the Milky Way to some extent. In this section we investigate the influence of plate-wide smoothing on the density profile with toy models.

We adopt a toy model of the vertical stellar density profile with two exponential components: the thin disk component has a scale height of 0.3 kpc and the thick disk has a scale height of 0.9 kpc. At $Z = 0$, the thick disk component is 15% of the thin one. The normalizer,

Table 4 The Halo-Like RGB Stars Selected from the LAMOST DR3 Catalog

obsid	RA	Dec	$T_{\rm eff}$	$\log g$	[Fe/H]	$v_{\rm los}$	M_K	A_K	K	Distance	Z	R	r	$\ln \nu$
	(deg)	(deg)	(K)	(dex)	(dex)	(km s^{-1})	(mag)	(mag)	(mag)	(kpc)	(kpc)	(kpc)	(kpc)	ln(pc^{-3})
75606117	337.46744	+6.01526	4829	1.11	−1.66	−3	$-3.96^{+0.42}_{-0.44}$	0.148	10.646	$7.79^{+1.75}_{-1.37}$	−5.22	8.26	9.77	−11.81
342809095	263.45129	+18.10553	3870	−0.24	−2.17	−22	$-6.12^{+0.02}_{-0.02}$	0.053	11.280	$29.47^{+0.27}_{-0.27}$	12.49	21.36	24.74	−13.47
318815093	233.31128	+12.40625	4386	0.53	−2.32	172	$-5.32^{+0.28}_{-0.16}$	0.287	10.854	$15.04^{+1.15}_{-1.82}$	11.45	3.53	11.98	−12.43

Notes: Here we only list the first three rows; the complete table is in the online file (*https://github.com/liuchaonaoc/LAMOST_density*).
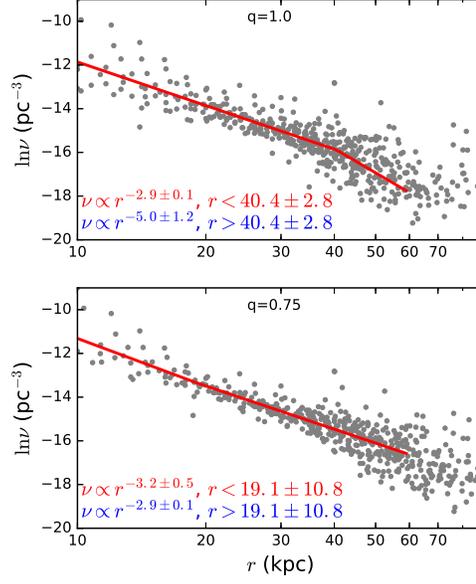
Fig. 15 In the top panel, with the fixed axis ratio of $q = 1.0$, the black dots represent the mean logarithmic stellar density values extracted from Fig. 14 by converting (R, Z) into radii r. The red line is the best fit broken power-law profile with a break radius of 40.4 ± 2.8 kpc. Within the break radius, the power-law index is -2.9 ± 0.1, while beyond this radius the index becomes -5.0 ± 1.2. The bottom panel illustrates the data with $q = 0.75$. In this case the broken power-law does not work quite well in the sense that the break radius has a large uncertainty of 10.0 kpc and the inner and outer power-law indices, which are -3.2 ± 0.5 and -2.9 ± 0.1 respectively, are quite similar. Therefore, it seems that a single power-law may be more suitable for this case.
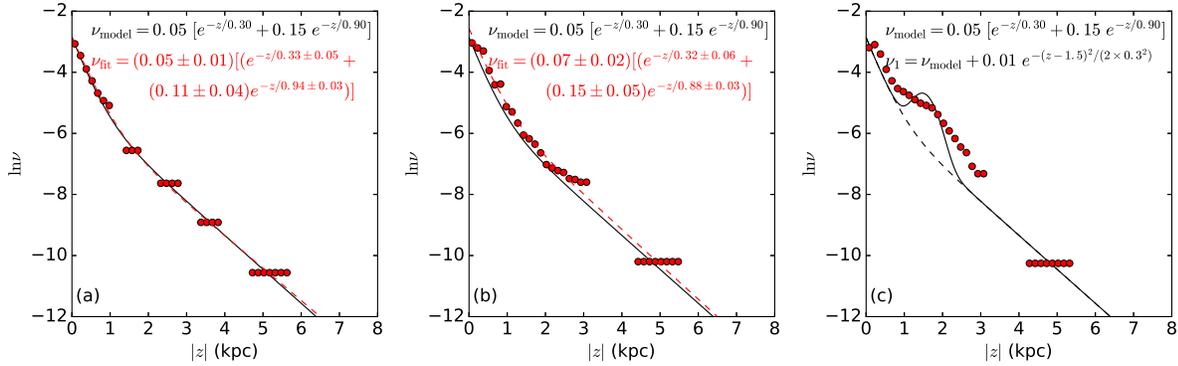
Fig. 16 Panel (a) shows the result of the vertical density profile estimates at $R = 12$ kpc for the toy model. The black solid line traces the profile from the toy model (see details in the text), while the red filled circles display the averaged stellar densities over the mock stars at various Z bins. The red dashed line represents the best fit result for the red filled circles. The "true" parameters of the toy model are shown in the top line of text in the panel and the best fit parameters are shown in the second line of text. Panel (b) displays a similar result of the vertical density profile at $R = 20$ kpc. The symbols in this panel are the same as in Panel (a). Panel (c) provides a similar result with an additional Gaussian-like substructure, which is centered at $z = 1.5$ kpc, superposed on the double-exponential toy model at $R = 20$ kpc. The first line of text indicates the toy model of the double-exponential profile, which corresponds to the black dashed line, while the second line of text expresses the analytical form of the Gaussian-like substructure.

i.e. the stellar density at $Z = 0$, is set to $0.05\,\mathrm{pc}^{-3}$. Such a configuration is similar to the Milky Way disk in the solar neighborhood.

Considering that LAMOST plates usually overlap each other in most of the surveyed sky areas, we draw mock stars from the toy model in the partly superposed LAMOST-like plates, i.e. circles with 5-degree-field-of-view, centered at $b = 0°, 1°, 2°, \ldots, 12°, 22°, \ldots, 82°$ with $l = 180°$. Obviously, these plates correspond to different vertical heights above the Galactic mid-plane; the larger the distance, the larger the vertical height.

For each plate, at a given Galactocentric radius, we randomly draw 50 mock stars evenly distributed within the circular area of the plate. Note that, even though the mock stars have the same radius, their Z values are different because they are distributed at various Galactic latitudes within the plate. According to Section 2, the stellar densities at the positions of the mock stars are derived by averaging over the model density covering the area of the plate at the given radius. Thus, although the Z values are different for the mock stars drawn from the same plate at a given radius, the stellar densities assigned to them are the same. This is a similar but simplified way to determine the stellar density as compared to the real observed data applied in Section 4. Although uncertainties in the distance estimates and the selection function corrections are ignored, the simplified method for stellar density determination is sufficient for investigating the effect induced by plate-wide smoothing.

5.1.1 The smoothing effect in the vertical disk profile

We first investigate the effect in the vertical density profile. Figure 16(a) shows the result of the vertical density profile test at $R = 12\,\mathrm{kpc}$. The black solid line indicates the "true" profile from the toy model. The red filled circles represent the averaged density values over the mock stars at different Z bins. The red dashed line displays the best-fit double-exponential profile with red symbols. It is seen that the derived stellar densities perfectly follow the "true" stellar density profile and the best-fit parameters are quite consistent with the "true" values with uncertainties less than 20%.

We then move to a larger distance at $R = 20\,\mathrm{kpc}$, where the effect of plate-wide smoothing should be more substantial. The result in Figure 16(b) shows that the shape of the best-fit density profile (red dashed line) derived from the "observed" data (red filled circles) is in agreement with the "true" profile (black solid line). Similar to the test at $R = 12\,\mathrm{kpc}$, the uncertainty of the derived model parameters is less than 20% without sub-

stantial systematic bias. In both cases at $R = 12$ and $20\,\mathrm{kpc}$, plate-wide smoothing due to the large field-of-view does not significantly change the result.

5.1.2 The smoothing effect in the Monoceros ring-like substructure

In Section 4.3, the well known Monoceros ring is not very apparent in the outer Galactic disk, according to stellar density profile estimates. Could it be possible that smoothing in the large field-of-view of a LAMOST plate smears out this kind of substructure? We superpose an additional Gaussian-like substructure with the peak located at $1.5\,\mathrm{kpc}$ and σ of $0.3\,\mathrm{kpc}$ in the toy model with a double-exponential vertical profile. The peak density of the Gaussian-like substructure is 20% of the disk density at $Z = 0$. The analytic form of the Gaussian-like substructure is indicated in the second line of text in Figure 16(c). The "true" vertical density profile with the Gaussian-like substructure located at $R = 20\,\mathrm{kpc}$ is shown as the black solid line in Figure 16(c). The red filled circles in the figure are the approximated density values obtained from the mock data. Compared with the derived best-fit stellar density in panel (b) (red dashed line), it is seen that the approximated stellar density values do show a broader bump between $Z = 1$ and $2\,\mathrm{kpc}$, which corresponds to the Gaussian-like substructure. This means that even though the substructure has a relatively smaller typical scale of $0.3\,\mathrm{kpc}$, it can still be identified after the spatial smoothing induced by the large field-of-view, although the resulting substructure may be broadened to some extent. If the substructure is very weak, however, its signal may also be weakened by the broadening effect.

5.1.3 Summary

To summarize, the stellar density profile derived from spatial averaging over the 5-degree-field-of-view may not significantly distort the derived Galactic structure. Moreover, although smoothing with relatively large sky area may broaden existing substructures, small-scale substructure with a typical scale of $0.3\,\mathrm{kpc}$ can still be identified from the resulting stellar density profile. Therefore, the disappearance of the Monoceros ring in our result may not be due to spatial smoothing induced by the applied technique.

5.2 The *post-observation* Selection Function

The selection function used in this work is determined by comparison between the color-magnitude diagram from

the photometric data and that from the final spectroscopic catalog. The selection function in the latter one is not only composed of targeting selection, but also considers selection effects induced during observations and data reductions. The original selection function is always altered during observations and data processing. For instance, one line-of-sight is designed to be observed multiple times with different ranges of magnitudes so that a larger range of magnitude can be covered. But due to weather, some designed plates may never be observed. Then the *post-observation* selection function for the line-of-sight would never be identical to the designed one. The site conditions, which allows more observation time in winter but less in summer, can also induce selection function in the sky area (Yao et al. 2012). Another instance is that the data reduction may lose spectra with very low signal-to-noise ratio, which may more frequently occur for fainter sources. Therefore, only using the originally designed targeting selection for the correction cannot take into account effects from observations and data reductions.

6 CONCLUSIONS

In this paper, we introduce a statistical method to measure the stellar density profile along a given line-of-sight using spectroscopic survey data. This technique can flexibly deal with stellar density for different stellar populations.

Our validation tests based on *Galaxia* mock data demonstrate that we can reconstruct the stellar density profile well for a sub-population. Moreover, even if the sub-population contains quite a few stars, we can still obtain reasonable density values at the spatial positions at which stars are located. Tests show that in the worst cases, the derived stellar density may have a systematic bias of about $\Delta\nu/\nu \sim 1$. However, large surveys like LAMOST can observe many plates covering a wide area of the sky. Thus, statistically, the final averaged stellar density in the small discrepant regions can reduce or cancel different systematic biases occurring in different lines-of-sight, and finally achieve relatively high precision when averaged over large areas.

Finally, we apply the method to LAMOST DR3 RGB stars. For the Galactic outer disk, we find that (1) the disk component still contributes to the total stellar surface density by about 10% at $R = 19$ kpc, implying that our Galaxy has quite a large stellar disk; (2) we do not observe any significant truncation of the disk, but only see a smooth transition from the disk to the halo at around 20 kpc; (3) the Monoceros ring is not seen in

our density map of the outer disk; and (4) the disk seems vertically asymmetric beyond $R \sim 12$ kpc.

We confirm that the stellar halo is oblate in the inner halo (within $r \sim 30$ kpc) and becomes spherical in the outer part. If we presume a constant axis ratio of 1.0, the stellar density profile shows a clear broken power-law similar to some previous works (e.g., Deason et al. 2011). However, if we change the constant axis ratio to 0.75, the broken power-law disappears. This means that whether the halo profile is broken or not depends on the axis ratio. The tomographic map of the stellar halo displayed in this work shows evidence that the halo has a radially varying axis ratio. Therefore, either using a single or broken double power-law with constant axis ratio does not properly reflect the real structure of the stellar halo.

Appendix A: TEST OF THE METHOD WITH SIMULATION DATA AT $L = 180°$ AND $B = 0°$

In this section we further provide an additional test for the method using the *Galaxia* simulation data located at $l = 180°$ and $b = 0°$, at which the extinction is significant, with a field of view of 20 square degrees.

Figure A.1 shows the test result with the selection function **T**1, while Figure A.2 shows the test result with the selection function **T**2. Under both selection functions, the performance of the stellar density estimates for stars with higher extinction do not exhibit significant differences with the previous tests, as displayed in Figures 4 and 9.
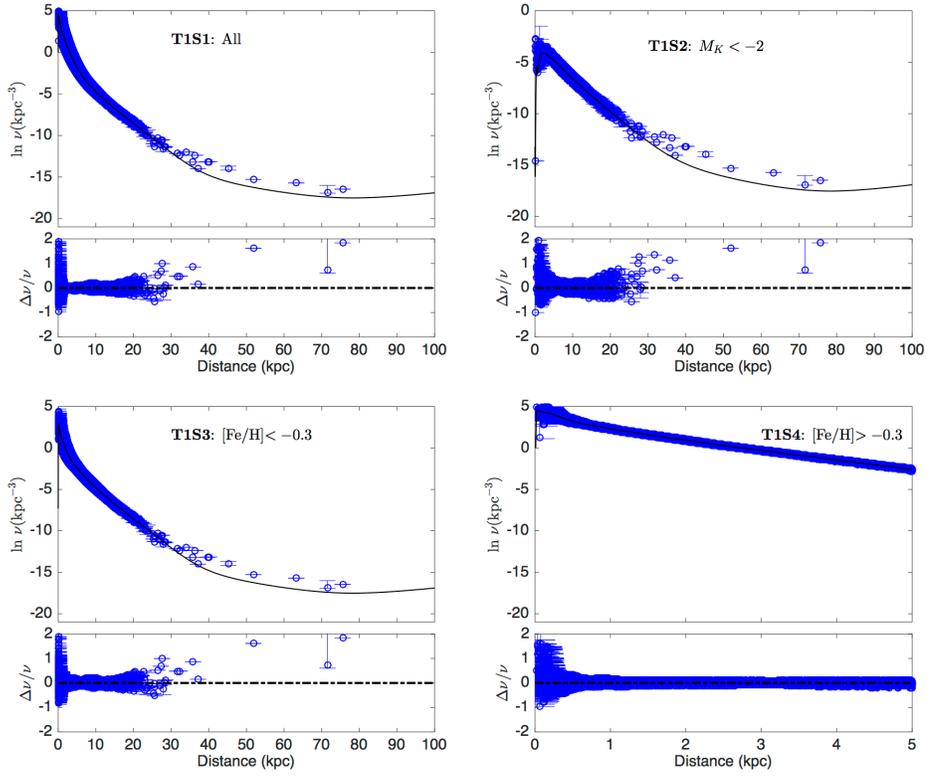
Fig. A.1 Similar to Fig. 4, but with the selection function **T1** for the *Galaxia* simulation data located at $l = 180°$ and $b = 0°$.
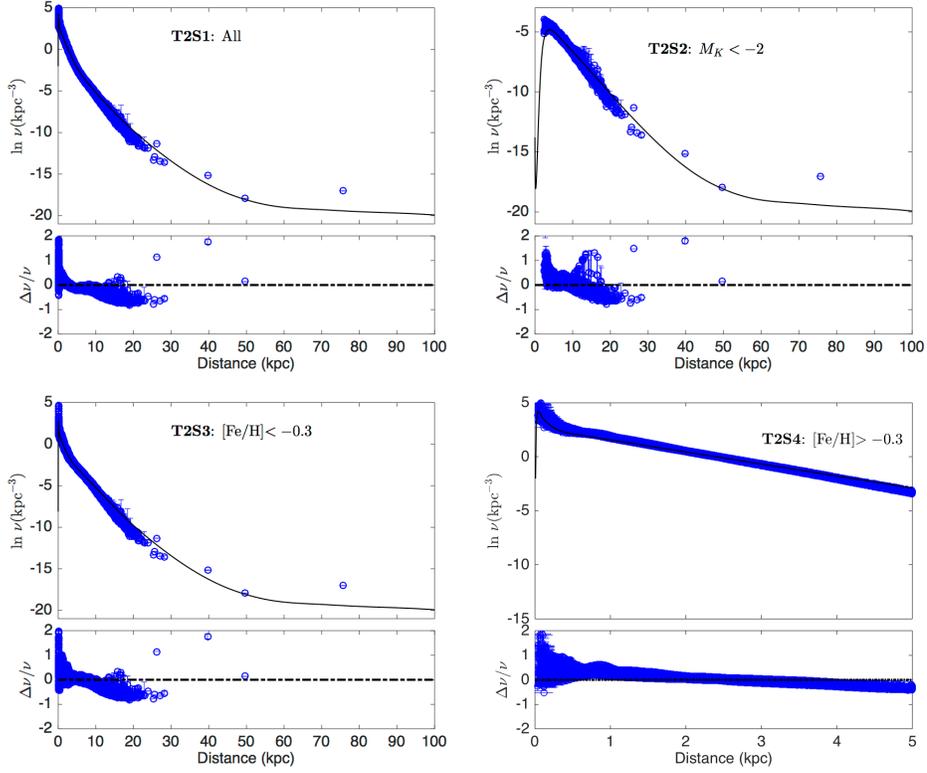
Fig. A.2 Similar to Fig. 4, but with the selection function **T2** for the *Galaxia* simulation data located at $l = 180°$ and $b = 0°$.

References

Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529

Bovy, J., Rix, H.-W., Liu, C., et al. 2012, ApJ, 753, 148

Bovy, J., Rix, H.-W., Schlafly, E. F., et al. 2016, ApJ, 823, 30

Carlin, J. L., Lépine, S., Newberg, H. J., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 755

Carlin, J. L., Liu, C., Newberg, H. J., et al. 2015, AJ, 150, 4

Carraro, G., Vázquez, R. A., Costa, E., Perren, G., & Moitinho, A. 2010, ApJ, 718, 683

Carraro, G. 2015, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 57, 138

Chen, B., Stoughton, C., Smith, J. A., et al. 2001, ApJ, 553, 184

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 1197

Deason, A. J., Belokurov, V., & Evans, N. W. 2011, MNRAS, 416, 2903

Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 735

Feast, M. W., Menzies, J. W., Matsunaga, N., & Whitelock, P. A. 2014, Nature, 509, 342

Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2014, ApJ, 783, 114

Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2015, ApJ, 810, 25

Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, ApJ, 808, 132

Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864

Liu, C., & van de Ven, G. 2012, MNRAS, 425, 2144

Liu, C., Deng, L.-C., Carlin, J. L., et al. 2014a, ApJ, 790, 110

Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al. 2014b, in IAU Symposium, 298, Setting the Scene for Gaia and LAMOST, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 310

López-Corredoira, M., Cabrera-Lavers, A., Garzón, F., & Hammersley, P. L. 2002, A&A, 394, 883

Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA (Research in Astronomy and Astrophysics), 15, 1095

Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, ApJ, 739, 25

Minniti, D., Saito, R. K., Alonso-García, J., Lucas, P. W., & Hempel, M. 2011, ApJ, 733, L43

Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, ApJ, 569, 245

Robin, A. C., Creze, M., & Mohan, V. 1992, ApJ, 400, L25

Schlafly, E. F., Finkbeiner, D. P., Jurić, M., et al. 2012, ApJ, 756, 158

Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, ApJ, 730, 3

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Tian, H.-J., Liu, C., Wan, J.-C., et al. 2016, arXiv:1603.06262

Wan, J.-C., Liu, C., Deng, L.-C., et al. 2015, RAA (Research in Astronomy and Astrophysics), 15, 1166

Watkins, L. L., Evans, N. W., Belokurov, V., et al. 2009, MNRAS, 398, 1757

Widrow, L. M., Gardner, S., Yanny, B., Dodelson, S., & Chen, H.-Y. 2012, ApJ, 750, L41

Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA (Research in Astronomy and Astrophysics), 11, 924

Wu, Y., Du, B., Luo, A., Zhao, Y., & Yuan, H. 2014, in IAU Symposium, 306, Statistical Challenges in 21st Century Cosmology, eds. A. Heavens, J.-L. Starck, & A. Krone-Martins, 340

Xia, Q., Liu, C., Mao, S., et al. 2016, MNRAS, 458, 3839

Xu, Y., Deng, L. C., & Hu, J. Y. 2006, MNRAS, 368, 1811

Xu, Y., Deng, L. C., & Hu, J. Y. 2007, MNRAS, 379, 1373

Xu, Y., Liu, C., Xue X-X. et al. 2017, submitted to MNRAS, arXiv:1706.08650

Xu, Y., Newberg, H. J., Carlin, J. L., et al. 2015, ApJ, 801, 105

Xue, X.-X., Rix, H.-W., Ma, Z., et al. 2015, ApJ, 809, 144

Yao, S., Liu, C., Zhang, H.-T., et al. 2012, Research in Astronomy and Astrophysics, 12, 772

Yuan, H.-B., Liu, X.-W., Huo, Z.-Y., et al. 2015, MNRAS, 448, 855

Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, AJ, 145, 44

Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, AJ, 146, 81

Zhang, L., Rix, H.-W., van de Ven, G., et al. 2013, ApJ, 772, 108

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA (Research in Astronomy and Astrophysics), 12, 723