Research in Astronomy and Astrophysics

Red clump stars from the LAMOST data I: identification and distance

Jun-Chen Wan¹, Chao Liu¹, Li-Cai Deng¹, Wen-Yuan Cui², Yong Zhang³, Yong-Hui Hou³, Ming Yang¹ and Yue Wu¹

Received 2015 April 1; accepted 2015 April 13

Abstract We present a sample of about 120 000 red clump candidates selected from the LAMOST DR2 catalog based on the empirical distribution model in the effective temperature vs. surface gravity plane. Although, in general, red clump stars are considered as standard candles, they do not exactly stay in a narrow range of absolute magnitude, but may have a range of more than one magnitude depending on their initial mass. Consequently, conventional oversimplified distance estimations with the assumption of a fixed luminosity may lead to systematic bias related to the initial mass or age, which can potentially affect the study of the evolution of the Galaxy with red clump stars. We therefore employ an isochrone-based method to estimate the absolute magnitude of red clump stars from their observed surface gravities, effective temperatures and metallicities. We verify that the estimation removes the systematics well and provides initial mass/age estimates that are independent of distance with accuracy better than 10%.

Key words: stars: general — stars: horizontal-branch — stars: statistics — stars: distances — Galaxy: stellar content

1 INTRODUCTION

Red clump (RC) stars are metal-rich stars in the evolution phase of helium-core burning (Cassisi & Salaris 1997). They play important roles in the study of the Galactic disk because they are widespread in the thin disk, are usually considered to be luminous standard candles (Paczyński & Stanek 1998; Girardi et al. 1998; Alves 2000; Groenewegen 2008) and form a prominent population in the color magnitude diagram, which makes them easily identifiable from multi-band photometry (López-Corredoira et al. 2002).

Identification of individual field RC stars, however, is not trivial because their properties are very similar to red giant branch (RGB) stars. Paczyński & Stanek (1998) used a Gaussian to model the distribution of magnitudes for RC stars and a quadratic polynomial for RGB stars. Then the stars located in the Gaussian dominated region are very likely to be RC stars. This method was applied by

¹ Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China; *junchenwan@bao.ac.cn*

² Physics Department, Hebei Normal University, Shijiazhuang 050024, China

³ Nanjing Institute of Astronomical Optics & Technology, National Astronomical Observatories, Chinese Academy of Sciences, Nanjing 210042, China

Nataf et al. (2013) to select RC stars in the Galactic bulge. Recently, Bovy et al. (2014) employed a new method to identify RC stars based on their distribution with a model that incorporated color, effective temperature, surface gravity, metallicity and stellar evolution. However, these authors only identify the primary RC stars and remove the secondary population to simplify the distance estimation.

Compared to identification, distance estimation of RC stars is relatively simple with the assumption that the absolute magnitude of RC stars is around a fixed value with a small dispersion (Paczyński & Stanek 1998; Girardi et al. 1998; Alves 2000; Groenewegen 2008; Li & Cao 2012).

However, the stellar evolution model demonstrates that RC stars do not always stay at the same luminosity. They are separated into two subclasses: primary RC stars, which have electron-degenerate cores, and secondary RC stars, which contain non-degenerate He cores (Girardi 1999). In general, the primary RC stars have low mass and are hence older, while the secondary RC stars are massive and therefore younger than 1 Gyr. Most of the primary RC stars are generally brighter than the secondary RC stars, thus the former have smaller log g than the later (see Stello et al. 2013). In the context of evolution of the Galactic disk, we intend to obtain a sample of RC stars with a wide range of ages so that we can trace the evolution of the Galaxy from present day back to ~ 10 Gyr. For instance, Salaris & Girardi (2002) fitted the distribution of RC stars observed by *Hipparcos* (Perryman et al. 1997), including both the primary and secondary populations, in a color-magnitude diagram with a stellar evolution model and derived the distribution of age in the solar neighborhood. Although keeping both the primary and secondary RC stars in the sample is important for the study of disk evolution, the distance estimation may turn out to be more complicated since the RC stars would not act as standard candles in this case.

As the first paper in a series of works on studying the evolution of the Galactic disk with RC stars from the LAMOST catalog, we initially develop a new method of identification for both the primary and secondary RC stars and use the normal approach to distance estimation for both populations. This paper is organized as follows. In Section 2, we briefly introduce data from the LAMOST survey and describe the empirical method of identification that is applied to RC stars. In Section 3, we develop a new method of distance estimation. The performance of the estimation method is then assessed in the same section. Finally, we further discuss the accuracy of our distance estimation in Section 4 and draw a brief conclusion in Section 5.

2 IDENTIFICATION OF RC STARS

In this section we identify RC stars from LAMOST data using an empirical method.

2.1 LAMOST Data

LAMOST, also called the Guo Shou Jing Telescope, is a 4-meter reflecting Schmidt telescope with 4000 fibers on a 20-square degree focal plane (Cui et al. 2012; Zhao et al. 2012). The LAMOST survey will observe more than 5 million low resolution stellar spectra during its 5-year survey (Deng et al. 2012; Liu et al. 2014b; Liu et al. 2015). According to Yao et al. (2012), winter is the best season to acquire data due to site conditions where LAMOST is located, which is best suited for the Galactic Anticenter region. Therefore, there will be lots of disk populations located in the region that will be observed by LAMOST. In this work, we adopt the derived $T_{\rm eff}$ directly from the LAMOST pipeline (Wu et al. 2011a,b, 2014; Luo et al. in prep) and the estimated log g from Liu et al. (2014a), who estimate log g using support vector regression with the training dataset from asteroseismic studies by the *Kepler* mission. The uncertainty in the log g estimates is only about 0.1 dex, which is a factor of two better than that from the LAMOST pipeline for metal-rich ([Fe/H]> -0.6) giant stars, including RC stars.

We select stars with $\log g$ between 0.9 and 3.5, [Fe/H] between -0.6 and 0.4, and $T_{\rm eff}$ between 3600 and 6000 K from the LAMOST Data Release 2 (DR2) catalog, which provides good coverage

of the RC region. We also remove all spectra with signal-to-noise ratio lower than 10 and finally obtain 279 423 stars.

2.2 Identification of RC Stars

It is expected that many RC stars in the disk population will be sampled by the LAMOST survey. Indeed, Figure 1 shows that RC stars from the LAMOST DR2 catalog are prominent in the $T_{\rm eff}$ vs. $\log g$ plane. In this section, we establish an approach for identification of RC stars from LAMOST data.

Paczyński & Stanek (1998) used a Gaussian and quadratic polynomial to model the distribution of magnitudes for RC and RGB stars, respectively. We expand this method into two dimensions in the T_{eff} vs. log g plane.

First, we empirically build a 2-D distribution model for the RGB stars in the T_{eff} vs. $\log g$ plane with various metallicity bins ([Fe/H] = (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4)). We mask out the region between $\log g = 2.0$ and 3.0 to avoid RC stars and fit the distribution of remaining RGB stars with the following empirical model,

$$N_{\rm RGB}(T_{\rm eff}, \log g) = N(\log g) \times \exp\left[-\frac{(T_{\rm eff} - T(\log g))^2}{\sigma^2(\log g)}\right],\tag{1}$$

where $N(\log g)$ and $\sigma^2(\log g)$ are three smoothing cubic spline functions with different metallicity bins ([Fe/H] = (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4)), respectively; and

$$T(\log g) = \begin{cases} 150.2\log^3 g - 946.3\log^2 g + 2417\log g + 2188 \ [Fe/H] \in (-0.6, -0.3) \\ 49.8\log^3 g - 374.1\log^2 g + 1435\log g + 2632 \ [Fe/H] \in (-0.3, 0.0) \\ 81.65\log^3 g - 590.7\log^2 g + 1921\log g + 2107 \ [Fe/H] \in (0.0, 0.4). \end{cases}$$
(2)

Figure 2 shows the best fit curves for the terms in Equations (1)–(2). From left to right, the range of metallicity is (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4), respectively. $N(\log g)$ and $\sigma^2(\log g)$



Fig. 1 The observed distribution of metal-rich ([Fe/H] > -0.6) RGB stars from LAMOST DR2 in the log g vs. $T_{\rm eff}$ plane. The bin size is $\Delta \log g = 0.02$ and $\Delta T_{\rm eff} = 20$ K.



Fig.2 Fitting the coefficients in Eq. (1). The solid lines are the best fit curves, while the cross symbols are measured from RGB stars with $\log g < 2$ and > 3. From left to right, the metallicity bins are (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4), respectively. The top panels show the best fit smoothing splines for N and the bottom panels show the best fit splines for σ^2 . The other panels show the best fit polynomials, the coefficients of which appear in Eq. (2).

are the best fit smoothing spline functions shown in the top panels and bottom panels of Figure 2, respectively. The other panels of Figure 2 show the best fit polynomials for the terms in Equation (2).

Assuming that RGB stars are smoothly distributed along $\log g$, we can *interpolate* the distribution of RGB stars between $\log g = 2.0$ and 3.0 with the best fit model shown in Equations (1)–(2). The middle panels of Figure 3 show the distribution of RGB stars according to our model in the T_{eff} vs. $\log g$ plane for various [Fe/H] bins in different rows.

Second, we subtract the smoothed distribution model of RGB stars for $\log g = 2$ and 3 and the residuals are mostly contributed by the RC stars, as shown in the right panels of Figure 3, in which the contours represent 68% (red) and 95% (yellow) completeness in the residual distribution. A compromise has to be made between the completeness for both primary and secondary RC stars and the fraction of contamination from the RGB stars. We find that the 68% completeness contour



Fig. 3 From top to bottom, the metallicity bins are (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4), respectively. The left panels display distributions of the full sample of giant stars with different metallicities. The middle panels show the best fit distribution models of RGB stars. The right panels present the distributions of residuals from the full sample in the left panels after subtracting the RGB distributions in the middle panels. They are mostly contributed by RC stars. The red and yellow contours show the 68% and 95% completeness of the RC candidates, respectively. The top horizontal, bottom horizontal and vertical white lines give cuts for stars with $\log g > 2.9$ dex, $\log g < 2.1$ dex and $T_{\rm eff} > 5200$ K, respectively. The slanted white line located at $\log g = 0.0016 T_{\rm eff} - 4.7170$ gives another cut for removing RGB bump stars located in the bottom-right corner of the 95% contour in the bottom-right panel. The color represents the stellar count in the bins.

cannot cover the region with the most secondary RC stars, therefore, we select 95% completeness as the recommended selection criterion for RC stars. With this empirical distribution of RC stars, other users can freely adjust the selection criterion to identify a different sample of RC stars to meet their specific requirements. It can be noted that there are a few fractional areas far from the empirical RC star region that are also within the 95% completeness level. They may be artificial features because the data are quite sparse in these regions. We then manually exclude three artificial areas with $T_{\rm eff} > 5200$ K, $\log g < 2.1$ dex and $\log g > 2.9$ dex. Moreover, the bottom-right corner of the



Fig. 4 The colors show the fractions of true RC stars with different metallicity bins, (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4), from the left to right panels, respectively, in the log g vs. T_{eff} plane.

[Fe/H]	Completeness level	Contour level	$N_{\rm RC}$	Total ratio of RC stars
(-0.6,-0.3)	68%	51.39	36099	89.68%
	95%	15.46	54890	80.47%
(-0.3,0.0)	68%	42.26	31553	83.32%
	95%	10.46	48064	76.29%
(0.0,0.4)	68%	12.12	11291	84.87%
	95%	3.01	15757	81.13%

Table 1 The Location of the RC Candidates in the Right Panel of Fig. 3

Notes: Contour level: the contour levels in Fig. 3 correspond to 68% or 95% completeness. $N_{\rm RC}$: number of RC candidates enclosed in the contour (with the additional cut around the edges, see white lines in the right panels of Fig. 3). Ratio of RC stars: the total fraction of RC stars to the all metal-rich giant stars within the contour level.

95% completeness level in the bottom-right panel is apparently contributed by the RGB bump stars rather than the RC stars. Therefore, a fourth cut at $\log g < 0.0016 T_{\rm eff} - 4.7170$ is added to remove contamination from the RGB bump stars. These additional data cuts are shown as white lines in the right panels of Figure 3. Finally, we find 118711 RC candidates with a refined 95% completeness level.

Although most stars located in the refined 95%-level region are RC stars, some contaminations might still be included. Assuming that the residual distribution in T_{eff} vs. $\log g$ planes of Figure 3 are the distributions of the true RC stars, we can give the percentage of the true RC stars by dividing the residual distributions by the full distributions, which contain both RC and RGB stars, although this method cannot identify individual RC stars.

Figure 4 shows the fractions of RC stars in the $T_{\rm eff}$ vs. log g planes for metallicity bins (-0.6, -0.3), (-0.3, 0.0) and (0.0, 0.4) from left to right, respectively, demonstrating that the fractions of true RC stars are mostly larger than 75%, even for some regions with secondary RC stars in the candidate catalog. The contour levels for 68% and 95% completeness, the numbers of RC candidates under such completeness levels, and the total fraction of the true RC stars are listed in Table 1.

Compared with the method provided by Bovy et al. (2014), our method does not depend on the stellar model but rather only on the specific observations. Moreover, although the number of secondary RC stars is less than that of primary RC stars, we can still discriminate them from the background RGB stars with a fraction of 75% \sim 85%, as shown in Figure 4. Therefore, the identification of RC stars in this work is suitable for both primary and secondary RC stars, ensuring that the study of the evolution of the Milky Way can be extended from <1 Gyr to around 10 Gyr based on this sample.

3 DISTANCE ESTIMATIONS

Most of the RC stars are located in the Galactic disk and their apparent magnitudes and color indices are significantly affected by interstellar extinction. Therefore, the distance and reddening have to be determined simultaneously. In the next subsection, we introduce a likelihood method to determine these quantities, and then we apply this method with fixed and varying absolute magnitude, respectively, in the following two subsections.

3.1 A Likelihood Method to Estimate Distance and Reddening

The first step to estimate the distance of RC stars is to estimate the reddening from the observed color index. Currently, LAMOST spectra do not have good optical multi-band photometry. The input catalog used is a combination of UCAC4 (Zacharias et al. 2013), PanSTARRS1 (Tonry et al. 2012), SDSS (Ahn et al. 2014), and the Xuyi Schmidt Telescope Photometric Survey of the Galactic Anti-center (XSTPS-GAC) (Yuan et al. 2015). Although all of these source catalogs contain g, r and i bands, they are still not well calibrated with each other. Therefore, at this stage, we use the 2MASS catalog (Skrutskie et al. 2006) as the input catalog to derive the reddening and distance for RC stars.

The likelihood of E(J - K) for an RC star given the observed J - K and the intrinsic color index of RC stars can be written as

$$Pr(E(J-K)|J-K,\sigma_{J-K},(J-K)_{RC},\sigma_{RC,J-K}) \sim \exp\left[-\frac{(E(J-K)-((J-K)-(J-K)_{RC}))^2}{2(\sigma_{J-K}^2+\sigma_{RC,J-K}^2)}\right],$$
(3)

where $(J - K)_{RC}$ is the intrinsic color index of RC stars, $\sigma_{RC,J-K}$ the dispersion of intrinsic color index for RC stars, and σ_{J-K} the measurement error of the observed J - K. To convert the reddening in J - K to extinction in K band, we adopt an expression from Indebetouw et al. (2005)

$$A_{\rm K} = 0.67E(J - K). \tag{4}$$

Then, the likelihood of the distance modulus, DM, for an RC star given the observed K magnitude and the fixed absolute magnitude M_K can be written as

$$Pr(DM|K, \sigma_K, A_K, M_K, \sigma_{M_K}) \sim \exp\left[-\frac{(DM - (K - A_K - M_K + 5))^2}{2(\sigma_K^2 + \sigma_{M_K}^2)}\right],$$
 (5)

where M_K is the fixed absolute magnitude in K band for RC stars, σ_{M_K} the intrinsic dispersion of absolute magnitude in K band for RC stars, and σ_K is the measurement error of the apparent K magnitude.

3.2 The Fixed Absolute Magnitude and the Intrinsic Color Index of RC Stars

Combined with Equations (3)–(5), the likelihood of both E(J - K) and DM for RC stars can be derived. The last ingredients that need to be put in are the absolute magnitude, intrinsic color index and their intrinsic dispersions for RC stars. Although some literatures have provided the absolute

 Table 2
 The Derived Absolute Magnitude and Intrinsic Color Index of RC Stars

	M _K (mag)	$\sigma_{\mathrm{M}_{K}}$ (mag)	J - K (mag)	σ_{J-K} (mag)
Reddened Dereddened	-1.529 ± 0.003 -1.549 ± 0.003	$\begin{array}{c} 0.075 \pm 0.003 \\ 0.076 \pm 0.003 \end{array}$	$\begin{array}{c} 0.681 \pm 0.003 \\ 0.658 \pm 0.001 \end{array}$	$\begin{array}{c} 0.104 \pm 0.005 \\ 0.072 \pm 0.001 \end{array}$

magnitude in the K band and the intrinsic absolute magnitude in J - K for RC stars (Alves 2000; Groenewegen 2008; Zasowski et al. 2013, etc.), the intrinsic dispersions for both quantities, which are necessary in our likelihood method, are not self-consistently provided. Therefore, we estimate the absolute magnitude in K band, the intrinsic color index in J - K, and their intrinsic dispersions with the *Hipparcos* data.

The *Hipparcos* catalog provides parallaxes for more than 100000 bright stars, thousands of which are located in the region of RC stars in the HR diagram. In order to estimate the absolute magnitude of RC stars, we need to correct the reddening as the first step. Bailer-Jones (2011) estimated the extinction parameters for about 47 000 *Hipparcos* stars which are sparsely distributed in the sky. We extend the extinction to all *Hipparcos* stars using spatial interpolation. For a star of interest, we select all stars with reddening parameters derived by Bailer-Jones (2011) within a 10-degree-radius circle and 20 pc in distance around it. Then we assign the median reddening value for all selected stars as the reddening value for the star of interest. In order to ensure the accuracy of the absolute magnitude, we select stars with errors in parallax smaller than 20% and errors in 2MASS photometry smaller than 0.5 mag. Figure 5 shows the J - K vs. M_K diagrams without (top-left panel) and with (middle-left panel) dereddening for about 900 giant stars with $M_K < 0$ mag.

We adopt the empirical model of the distribution in M_K from Paczyński & Stanek (1998), which has the following form

$$F = a + b(M_K - c)^2 + d \exp\left[-\frac{(M_K - e)^2}{2f^2}\right],$$
(6)

where a, b, c, d, e and f are the free parameters. The quadratic polynomial in Equation (6) models the stellar distribution of the RGB stars and the Gaussian term models the RC stars.

Similarly, we use this model for the marginalized distribution of color index

$$F' = a' + b'(J - K - c')^2 + d' \exp\left[-\frac{(J - K - e')^2}{2f'^2}\right].$$
(7)

We fit the models for the absolute magnitude and color index both without and with dereddening. The top-right panel of Figure 5 shows the best fit for the marginalized reddened absolute magnitude with Equation (6). The middle-right panel shows the best fit model for the marginalized dereddened absolute magnitude. In addition, the bottom panel shows the best fit for the reddened (dashed line) and dereddened (solid line) color index J - K. Table 2 lists the best fit absolute magnitude and intrinsic color index and their dispersions. It shows that the dereddened absolute magnitude and intrinsic color index are brighter and bluer than the reddened values by about 2%, respectively. In this work we adopt the dereddened M_K and J - K as the standard value in the estimation of the distances for RC stars.

3.3 The M_K Based on Synthetic Isochrones

The assumption that the RC stars have a fixed magnitude can only work for primary RC stars, which are mostly composed of relatively older RC stars compared with the secondary ones. When we want to trace the evolution of the Galactic disk with RC stars, we cannot only use the primary RC stars



Fig. 5 The top-left panel shows the reddened J - K vs. $M_K + A_K$ diagram for about 900 giant stars selected from the *Hipparcos* catalog. The middle-left panel shows a similar plot but with a dereddened color index and absolute magnitude. The bottom panel shows the marginalized distribution of the reddened J - K for the stars (corresponding to the top-left panel) with cross symbols and that of the dereddened color index $(J - K)_0$ (corresponding to the middle-left panel) with square symbols. The dashed and solid lines are the best fit model of Eq. (7) for the J - K and $(J - K)_0$, respectively. The top-right panel shows the marginalized distribution of $M_K + A_K$ with cross symbols. The dashed line stands for the best fit model according to Eq. (6). The middle-right panel shows the marginalized distribution of the dereddened M_K with square symbols, and the best fit model with the solid line.

and ignore the secondary RC stars. Therefore, we need to improve the approach used to estimate the distance to RC stars so that the secondary RC stars are also taken into account.

We turn to use a special isochrone fitting process to estimate the absolute magnitude for all kinds of RC stars. After quickly reviewing the isochrones, we realize that the absolute magnitude of RC stars is a function of log g, [Fe/H] and $T_{\rm eff}$. Panel (a) of Figure 6 shows the synthetic RC stars from the PARSEC library (Bressan et al. 2012) in the log g vs. M_K plane. It shows that M_K is strongly dependent on log g. Further separations of the data into different initial stellar masses are shown in the right panels. We find that the RC stars with stellar masses of $0.8 \sim 1.1 M_{\odot}$ (panel (b) in Fig. 6) are mostly concentrated within $M_K = -1.4 \sim -1.6$ mag, while the M_K for RC stars with initial mass of $1.1 \sim 1.4 M_{\odot}$ increases to $-1.6 \sim -1.8$ mag (panel (c)). Then the M_K for RC stars with initial mass at $1.4 \sim 1.7 M_{\odot}$ moves back to the range $-1.5 \sim -1.7$ mag (panel (d)). For RC stars with $1.7 \sim 2.0 M_{\odot}$, M_K dramatically extends from ~ -1 to more than -2 mag (panel



Fig. 6 Panel (a): The synthetic $\log g$ vs. M_K diagram for RC stars from the PARSEC stellar evolution track (Bressan et al. 2012). The range of age is from 4 Myr to 13 Gyr in steps of $\Delta(\log t) = 0.05$. The ranges of metallicity [Fe/H] and $M_{\rm ini}$ are $-0.6 \sim 0.3$ ($Z_{\odot}=0.0152$) and $0.8 \sim 2M_{\odot}$ respectively. Panels (b)-(e): The log g-M_K relation for RC stars with $M_{\rm ini}$ between $0.8 \sim 1.1M_{\odot}$, $1.1 \sim 1.4M_{\odot}$, $1.4 \sim 1.7M_{\odot}$, and $1.7 \sim 2.0M_{\odot}$, respectively.

(e)). To further investigate how the M_K varies, we separate the synthetic stars into two groups at [Fe/H] = -0.3. For the RC stars with [Fe/H] > -0.3 dex, the relation between M_K and $\log g$ can be empirically modeled with a quadratic polynomial (see the related panels in Fig. 7)

$$M_K = P_1 \log^2 g + P_2 \log g + P_3.$$
(8)

The best fit coefficients P_i (i = 1, 2, 3) are listed in Table 3. For stars with [Fe/H] < -0.3 dex, when $\log g < 2.45$, the M_K is no longer a function of $\log g$ (see the related panels in Fig. 7). Figure 8 shows that for these stars, the absolute magnitude is roughly a linear function of effective temperature. Then we have the following more complicated relation

$$M_K = \begin{cases} P_1 T_{\text{eff}} + P_2 & \log g < 2.45\\ P_1 \log^2 g + P_2 \log g + P_3 & \log g > 2.45 \end{cases}.$$
(9)

Table 4 shows the best fit coefficients of P_i (i = 1, 2, 3). It can be noted that for both groups of metallicity, the synthetic data (dots) shown in Figures 7 and 8 are not exactly located on a narrow line, but are spread out with varying dispersions. We then measure the dispersions of the residuals of M_K for the synthetic RC stars with respect to the best fit models at different bins of metallicity and show them in the column for σ_{M_K} in Tables 3 and 4.

Then, for each observed RC star, we firstly derive the M_K and σ_{M_K} from its T_{eff} , $\log g$ and [Fe/H], and reinsert them into Equation (5) to derive the likelihood of the distance modulus. Because the intrinsic color index of the primary and secondary RC stars is quite similar, we adopt this value derived from Section 3.2.

3.4 Performance Assessment

Before applying the improved distance estimation results to the LAMOST data, we assess the performance of the improved M_K model demonstrated in Section 3.3.



Fig.7 For RC stars with $\log g > 2.45$, M_K is modeled as a quadratic polynomial of $\log g$ for each [Fe/H] bin. The dots are the synthetic data and the lines are the best fit quadratic polynomials.

We arbitrarily select 1000 points from the synthetic dataset and add random Gaussian errors to the true values of $T_{\rm eff}$, $\log g$ and [Fe/H]. For each synthetic RC star, we create 20 mock stars with different random errors. In total, we create 20 000 mock stars with errors. Then, we derive the absolute magnitude for the mock data based on the method described in Section 3.3. We create a total of nine mock datasets with various measurement errors of $\log g$, $T_{\rm eff}$ and [Fe/H]. In the first three mock datasets, we simulate errors with $\sigma_{\log g} = 0.1$, 0.2 and 0.3 dex, respectively, with $\sigma_{T_{\rm eff}} = 120$ K and $\sigma_{\rm [Fe/H]} = 0.1$ dex. The second three mock datasets are simulated with random errors of $\sigma_{T_{\rm eff}} = 120$, 150 and 200 K, respectively, with $\sigma_{\log g} = 0.1$ dex and $\sigma_{\rm [Fe/H]} = 0.1$ dex. The second three mock datasets are simulated with random errors of $\sigma_{T_{\rm eff}} = 120$ K. We compare the derived absolute magnitudes with the true values in the nine mock samples. The residuals of the derived absolute magnitudes as functions of the errors in the stellar parameters are shown in Figure 9.

The left panels of Figure 9 show the residuals of M_K , denoted as δM_K , as a function of the errors in $\log g$ at $\sigma_{\log g} = 0.1$, 0.2 and 0.3 dex from top to bottom, respectively. It demonstrates that the larger the uncertainty in $\log g$, the larger the errors in absolute magnitude. When the random errors in $\log g$ are larger, the derived M_K seems more overestimated. However, slightly increasing



Fig.8 For RC stars with log g < 2.45 and [Fe/H] < -0.3, the M_K is modeled linearly as a function of $T_{\rm eff}$ in each [Fe/H] bin. The dots are the synthetic data and the lines are the best fit lines.

Table 3 The Coefficients from Different [Fe/H] Bins forthe M_K Model of Metal-rich RC Stars

	$M_K = P_1 \log^2 g + P_2 \log g + P_3$						
[Fe/H]	P_1	P_2	P_3	σ_{M_K}			
$\begin{array}{c} (-0.3,-0.2) \\ (-0.2,-0.1) \\ (-0.1,0.0) \\ (0.0,0.1) \\ (0.1,0.2) \end{array}$	6.72 6.67 6.21 6.17 6.19	-33.94 -33.55 -31.25 -31.00 -31.05	41.15 40.49 37.64 37.26 37.26	0.12 0.10 0.11 0.10 0.09			

Table 4 The Coefficients from Different [Fe/H] Bins for the M_K Model of Metal-poor RC Stars

	$\label{eq:main_state} \begin{split} \log g < 2.45 \\ \mathrm{M}_K = P_1 T_\mathrm{eff} + P_2 \end{split}$			$\label{eq:MK} \begin{split} \log g &> 2.45\\ \mathbf{M}_K &= P_1 \log^2 g + P_2 \log g + P_3 \end{split}$				
[Fe/H]	P_1	P_2	σ_{M_K}	l	P_1	P_2	P_3	σ_{M_K}
(-0.6, -0.5) (-0.5, -0.4) (-0.4, -0.3)	0.0024 0.0026 0.0027	-13.65 -14.50 -14.73	0.15 0.14 0.14		7.87 6.28 6.54	-39.98 -31.74 -33.09	49.06 38.41 40.16	0.09 0.08 0.08



Fig. 9 Scatter plots showing the residuals of derived M_K , denoted by δM_K , for 20 000 simulated data with various uncertainties in log g (left panels), [Fe/H] (middle panels) and T_{eff} (right panels).



Fig. 10 The relationship between the standard deviation of the residuals in M_K , denoted by σ_{M_K} , and the uncertainty in log g (σ_{logg}), [Fe/H] ($\sigma_{[Fe/H]}$) and T_{eff} ($\sigma_{T_{eff}}$).

the uncertainties in [Fe/H] (the middle panels) and $T_{\rm eff}$ (the right panels) would not significantly increase uncertainties in the M_K estimates.

Figure 10 presents the standard deviation of δM_K in terms of the σ of the best fit Gaussian from the histogram of δM_K , as functions of the uncertainties of $\log g$ (the left panel), [Fe/H] (the middle panel) and T_{eff} (the right panel). Again, this figure shows that the accuracy in the M_K estimates mostly relies on the accuracy of $\log g$, rather than that of [Fe/H] and T_{eff} . Therefore, an accurate $\log g$ calibrated with asteroseismic $\log g$ from Liu et al. (2014a) is very important in a highly accurate distance estimation.

Figure 10 also shows that with the typical uncertainty of 0.1 dex in $\log g$ the uncertainty of the derived absolute magnitude is better than 0.2 mag, corresponding to ~10% in distance.

4 DISCUSSION

4.1 Comparison between Two Absolute Magnitude Models

In Section 3, we discuss two approaches to estimating the absolute magnitude of RC stars. It is worth directly comparing the distance estimates based on the two different methods. Figure 11 shows the difference in the distance estimates for the LAMOST RC stars between the fixed absolute magnitude-based and the isochrone-based method as a function of log g. It can be seen that the fixed magnitude-based method tends to underestimate the values for RC stars with smaller log g and also tends to significantly overestimate those with larger log g. The overestimation in data with large log g is because the fixed absolute magnitude is dominated by the primary RC stars and hence may not be suitable for secondary RC stars. The underestimation when log g < 2.7 dex is likely because of the slight inconsistency between the M_K used in the synthetic isochrones and the one derived in Section 3.2. In Figure 11, we find that the systematic bias can shift by more than 20% in large log g given a fixed absolute magnitude. When log g < 2.3 dex, the isochrone-based method may not give reliable M_K estimates for RC stars since this is very close to the boundary of the isochrone data (see Fig. 8). Therefore the errors increase in this region, as shown in Figure 11.

4.2 External Uncertainty in the Distance Estimation

In Section 3.4, we use the mock data created from the synthetic data to test the performance of distance estimation based on isochrones. This, however, can only give the internal error but not the external one. We then cross-identify the RC stars from the LAMOST data with the *Hipparcos* catalog. Unfortunately, we only obtain less than 10 common RC stars with parallax error less than



Fig. 11 The comparison between the the isochrone-based (D_{iso}) and fixed absolute magnitude-based (D_{fixM_F}) distance estimates with various log *q*.

50%. Because most of these RC stars suffer larger uncertainty in parallax, they cannot be treated as standard stars to investigate the external error in the isochrone-based distance estimates. To resolve this issue, we need to wait for the coming *Gaia* data (Bailer-Jones 2009), which will release its first catalog in 2016.

5 CONCLUSIONS

In this work, we set up an empirical model in the $T_{\rm eff}$ vs. log g plane to identify the RC stars in data from the LAMOST DR2. The employed approach identifies not only the primary, but also the secondary RC stars. This will be very helpful for the study of the evolution of the Milky Way, because the range of RC stars can be extended from <1 Gyr (contributed by the secondary RC stars) to 10 Gyr (contributed by the primary RC stars). Finally, we identify 118711 RC stars from LAMOST DR2 with a 95% completeness level. The sample of selected RC stars may be contaminated by RGB stars with a fraction of about 20%.

After identifying the RC stars, we develop two different approaches to estimate their distances as well as the interstellar extinction. The first one is based on a fixed absolute magnitude and intrinsic color index. The accuracy of the fixed absolute magnitude-based approach relies on the accuracy of determining the absolute magnitude and the dispersion of the absolute magnitude. Consequently, we revisit the absolute magnitude and its dispersion for RC stars in *K* band. We adopt an empirical model that is similar to the one introduced by Paczyński & Stanek (1998), but take into account interstellar extinction for *Hipparcos* RC stars. Although the extinction is small, it leads to an underestimation of the absolute magnitude by about 2%.

Although this method is sufficiently accurate for most of the primary RC stars, it is not suitable for secondary RC stars, which significantly vary in $\log g$ and hence also in absolute magnitude. We then develop a second approach considering both types of RC stars based on the isochrones. With the empirical model, we associate the absolute magnitude of an RC star with its [Fe/H], $\log g$ and $T_{\rm eff}$. The more refined model can reduce the uncertainty in the distance estimates to 10% for almost all types of RC stars given the error in $\log g$ of around 0.1 dex.

This sample of RC stars gives a good representation of the Galactic disk, particularly the outer disk, allowing us to map the structure, kinematics and evolution of the Galactic disk in future works.

Acknowledgements This work is supported by the Strategic Priority Research Program "The Emergence of Cosmological Structures" of the Chinese Academy of Sciences (Grant No. XDB09000000) and the National Key Basic Research Program of China (2014CB845700). CL acknowledges the National Natural Science Foundation of China (NSFC, Grant Nos. 11373032, 11333003 and U1231119). The Guo Shou Jing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by National Astronomical Observatories, Chinese Academy of Sciences.

References

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17

Alves, D. R. 2000, ApJ, 539, 732

Bailer-Jones, C. A. L. 2009, in IAU Symposium, 254, eds. J. Andersen, J. Bland-Hawthorn, & B. Nordström, 475

Bailer-Jones, C. A. L. 2011, MNRAS, 411, 435

Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, ApJ, 790, 127

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

- Cassisi, S., & Salaris, M. 1997, MNRAS, 285, 593
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 1197
- Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA (Research in Astronomy and Astrophysics), 12, 735
- Girardi, L., Groenewegen, M. A. T., Weiss, A., & Salaris, M. 1998, MNRAS, 301, 149
- Girardi, L. 1999, MNRAS, 308, 818
- Groenewegen, M. A. T. 2008, A&A, 488, 935
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, ApJ, 619, 931
- Li, S.-L., & Cao, X. 2012, ApJ, 753, 24
- Liu, C., Fang, M., Wu, Y., et al. 2014a, arXiv:1411.0235
- Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al. 2014b, in IAU Symposium, 298, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 310
- Liu, X. W., Zhao, G., & Hou, J. L. 2015, RAA (Research in Astronomy and Astrophysics), 15, 1089
- López-Corredoira, M., Cabrera-Lavers, A., Garzón, F., & Hammersley, P. L. 2002, A&A, 394, 883
- Nataf, D. M., Gould, A., Fouqué, P., et al. 2013, ApJ, 769, 88
- Paczyński, B., & Stanek, K. Z. 1998, ApJ, 494, L219
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49
- Salaris, M., & Girardi, L. 2002, MNRAS, 337, 332
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
- Stello, D., Huber, D., Bedding, T. R., et al. 2013, ApJ, 765, L41
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, ApJ, 750, 99
- Wu, Y., Singh, H. P., Prugniel, P., Gupta, R., & Koleva, M. 2011a, A&A, 525, A71
- Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011b, RAA (Research in Astronomy and Astrophysics), 11, 924
- Wu, Y., Luo, A., Du, B., Zhao, Y., & Yuan, H. 2014, arXiv:1407.1980
- Yao, S., Liu, C., Zhang, H.-T., et al. 2012, RAA (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
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA (Research in Astronomy and Astrophysics), 12, 723