CrossMark

Determination of Distance, Extinction, Mass, and Age for Stars in LAMOST DR7

Jianling Wang[1], Zihuang Cao[1], Yang Huang[1], and Haibo Yuan[2]

[1] CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Beijing 100101, China; wjianl@bao.ac.cn
[2] Department of Astronomy, Beijing Normal University, Beijing 100875, China

Abstract

Large scale spectroscopic surveys such as that using Large-sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) have collected spectra of millions stars in the Milky Way. Utilizing this huge sample of stars to study the assembling history and structure of our Galaxy requires accurate estimates of distance, extinction, age, and mass for individual stars. Combining the parallax constraint from Gaia EDR3 with Bayesian inference, we have estimated the distance and extinction for stars observed in LAMOST DR7, as well as the stellar mass and age for evolved stars in this data release. We validated the accuracies of the stellar parameters by comparing our results against various measurements, including the star-pair technique, asteroseismology, globular clusters, and isochrone fits to main sequence stars and subgiants. This is a valuable catalog of stellar parameters under a Bayesian framework estimated using the data from Gaia EDR3 and LAMOST spectroscopic data. With this data set we explored the stellar population of the Galactic massive substructure Gaia-Sausage-Enceladus (GSE). The kinematically selected members of GSE have a median metallicity of $[Fe/H] = -1.29$ and a median age of 11.6 Gyr.

Key words: stars: distances – stars: fundamental parameters – stars: kinematics and dynamics – Galaxy: formation – (Galaxy:) globular clusters: general

1. Introduction

We have entered a golden era of Galactic archaeology with many large-scale sky surveys and huge data sets available in both photometry and spectroscopy, e.g., the Two Micron All-Sky Survey (2MASS, Skrutskie et al. 2006), the Sloan Digital Sky Survey (SDSS, York et al. 2000), the Sloan Extension for Galactic Understanding and Exploration (SEGUE, Yanny et al. 2009), the Large-sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST, Deng et al. 2012; Liu et al. 2014), the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), and the Gaia mission (Gaia Collaboration et al. 2016, 2018, 2021). These ambitious surveys have provided valuable information for a large number of stars, and are essential for understanding the structure, formation, and evolution of our Galaxy. It is expected that in the next decade the ongoing (e.g., LAMOST and Gaia) and upcoming (e.g., 4MOST, WEAVE; Dalton et al. 2014; de Jong et al. 2019) surveys will continually revolutionize our view on the Galactic structure and its assembling history.

To make full use of these surveys to study the assembling history of our Milky Way, it is crucial to derive and use the multi-dimensional information on individual stars. These astrophysical properties include the six-dimensional phase space coordinates (i.e., three-dimensional (3D) position and 3D velocity), metallicity and chemical abundances, photometries,

age, and extinction. The combination of these astrophysical parameters can be used to reconstruct the assembly history of the Galaxy. Chemo-dynamical studies have led to many new findings, such as stellar streams, dwarf satellites, and clumps. One of the major contributions of Gaia to this field is the discovery of the so-called Gaia-Sausage-Enceladus (GSE; Belokurov et al. 2018; Helmi et al. 2018; Xiang & Rix 2022), a major merger event of the Milky Way 8–11 Gyr in age. Now our inner halo is almost dominated by the relics of this massive merger event. This finding is achieved by the analysis of kinematics, chemistry, age, and spatial distribution of stars, which demonstrate the importance of the combination of multi-dimensional information on individual stars.

The LAMOST Galactic survey began its regular survey in the fall of 2012, which was the first spectroscopic survey to obtain spectra of over 10 million stars, and regularly released basic stellar parameters measured with the official LAMOST stellar parameter pipeline (LASP; Wu et al. 2011; Luo et al. 2015). These stellar parameters include effective temperature T_{eff}, metallicity $[Fe/H]$, surface gravity $\log g$, and the line-of-sight velocity V_r. To fully make use of this huge data set for Galactic archaeology, it is of vital importance to provide an estimate of distance, extinction, age, and stellar mass. In addition to those basic stellar parameters, in the Value-Added

Catalog, more physical parameters (like the distance, extinction, mass, and age) from the LAMOST surveys are provided to the community (e.g., Wang et al. 2016b; Xiang et al. 2017a). For example, using the isochrone fitting method, Xiang et al. (2017b) derived stellar ages for around 1 million main sequence turnoff stars and subgiants in the LAMOST Spectroscopic Survey of the Galactic Anti-center (LSS-GAC; Liu et al. 2014). Xiang & Rix (2022) derived ages for subgiants in LAMOST Data Release 7 (DR7) with the Bayesian isochrone fitting method. Huang et al. (2020) derived distances, masses, and ages for ∼140,000 primary red clump stars with a kernel principal component analysis method. The measurements of the previous studies were usually based on different methods, different models, and different prior functions, which usually caused systematic errors (Pinsonneault et al. 2018). It would be useful to check those results from different approaches.

It has been demonstrated that Bayesian inference is a powerful and important method in astrophysics, especially in deriving stellar parameters (Burnett & Binney 2010; Burnett et al. 2011; Binney et al. 2014; McMillan et al. 2018; Queiroz et al. 2018; Aguirre Børsen-Koch et al. 2022; Anders et al. 2022; Wang et al. 2022b). With this approach, we have derived distance and extinction for the LAMOST spectroscopy survey (Wang et al. 2016b) and SDSS/APOGEE survey (Wang et al. 2016a). In the current work, we improved our former calculation by imposing constraints with parallaxes from Gaia Early Data Release 3 (EDR3). With extensive comparisons with independent measurements, we show that accurate parallaxes are important for obtaining accurate stellar parameters, especially for significantly improving the accuracies of stellar ages and masses.

The paper is structured as follows. In Section 2, we briefly describe the methodology and data employed in the current analysis. In Section 3, we compare our estimates of distance, extinction, mass, and age with independent measurements in the literature. In Section 4, we use these data to study the age and metallicity of GSE substructures. Finally, a summary is given in Section 5.

2. Methodology and Data

In Wang et al. (2016a, 2016b), we applied the Bayesian method to estimate the distance and extinction to stars observed by both the LAMOST and APOGEE surveys. The observables used in this method include stellar parameters $T_{\rm eff}$, $\log g$, and [Fe/H], from LAMOST/APOGEE spectra, and infrared photometries J, H, Ks from the 2MASS survey. In this work, we include the constraint of Gaia parallax (ϖ). These observables form an observed vector

$$\boldsymbol{O} = (T_{\rm eff}, [{\rm M/H}], \log g, J, H, Ks, \varpi). \quad (1)$$

Each star can be characterized by a set of "intrinsic" parameters: metallicity [M/H], age τ, initial mass $\mathcal{M}$, position

on the sky l, b, and distance from the Sun d. These quantities also form another vector

$$\boldsymbol{X} = ([{\rm M/H}], \log \tau, \mathcal{M}, l, b, d). \quad (2)$$

With trivial Bayesian theory, we can determine the posterior probability of $P(\boldsymbol{X}|\boldsymbol{O})$, which is the conditional probability of the parameter set $\boldsymbol{X}$ given $\boldsymbol{O}$

$$P(\boldsymbol{X}|\boldsymbol{O}) = \frac{P(\boldsymbol{X})}{P(\boldsymbol{O})} P(\boldsymbol{O}|\boldsymbol{X}), \quad (3)$$

where $P(\boldsymbol{O}|\boldsymbol{X})$ is the likelihood of $\boldsymbol{O}$ given the parameter set $\boldsymbol{X}$. $\boldsymbol{O}$ and $\boldsymbol{X}$ can be connected by theoretical isochrones. $P(\boldsymbol{X})$ is the prior probability ascribed to the set of parameters. A Gaussian function is used to associate the measured observables for each component of $\boldsymbol{O}$ with a mean $\tilde{\boldsymbol{O}}$ and standard deviation $\sigma_{\tilde{O}}$.

$$P(\tilde{\boldsymbol{O}}|\boldsymbol{X}) = G(\tilde{\boldsymbol{O}}|\boldsymbol{O}(\boldsymbol{X}), \sigma_{\boldsymbol{o}}) = \prod_{(i=1)}^{n} G(\tilde{O}_i|\boldsymbol{O}(\boldsymbol{X}), \sigma_{oi}). \quad (4)$$

An extinction model and a three-component model of the Galaxy for the distribution functions of metallicity, density, and age are used for the prior probability, $P(\boldsymbol{X})$

$$P(\boldsymbol{X}) = p(\mathcal{M}) \sum_{i=1}^{3} p_i([{\rm M/H}]) p_i(\tau) p_i(\boldsymbol{r}) A_{V\,{\rm prior}}(\ell, b, d). \quad (5)$$

Here $i = 1$, 2, 3 correspond to thin- and thick-disk, and stellar halo, respectively.

The extinction prior model ($A_{V{\rm prior}}(\ell, b, d)$) is taken from Sharma et al. (2011) with normalization to infinity adopted from Schlegel et al. (1998). The Padova stellar isochrones[3] are used. With the posterior distribution function $P(\boldsymbol{X}|\boldsymbol{O})$ in hand, it is trivial to derive the moments for each component x_i of $\boldsymbol{X}$. The first and second moments for each component x_i of $\boldsymbol{X}$ are the expectation and uncertainties for each stellar parameter respectively. In the current work, we derived the expectation values and their associated uncertainties for four stellar parameters: stellar distance, extinction, stellar mass, and age. For details on the method, please refer to Wang et al. (2016a, 2016b), Burnett & Binney (2010), Burnett et al. (2011), Binney et al. (2014).

In the current work, we follow the same method as used in Wang et al. (2016a, 2016b), but with several improvements. First, the metallicity resolution of isochrones is increased to a step of 0.05 dex ranging from −2.2 to 0.5. These improved resolutions in metallicity result in 55 metallicity isochrones, which are a factor of 2.5 larger than those used in Wang et al. (2016a, 2016b). Even though the uncertainty of observed stellar metallicity is larger than this value, higher resolution of metallicity in isochrones helps break the degeneracy between metallicity and age for the internal precision, especially for the stellar age. We note that this metallicity bin size is within the ranges of the literature using a Bayesian framework for stellar
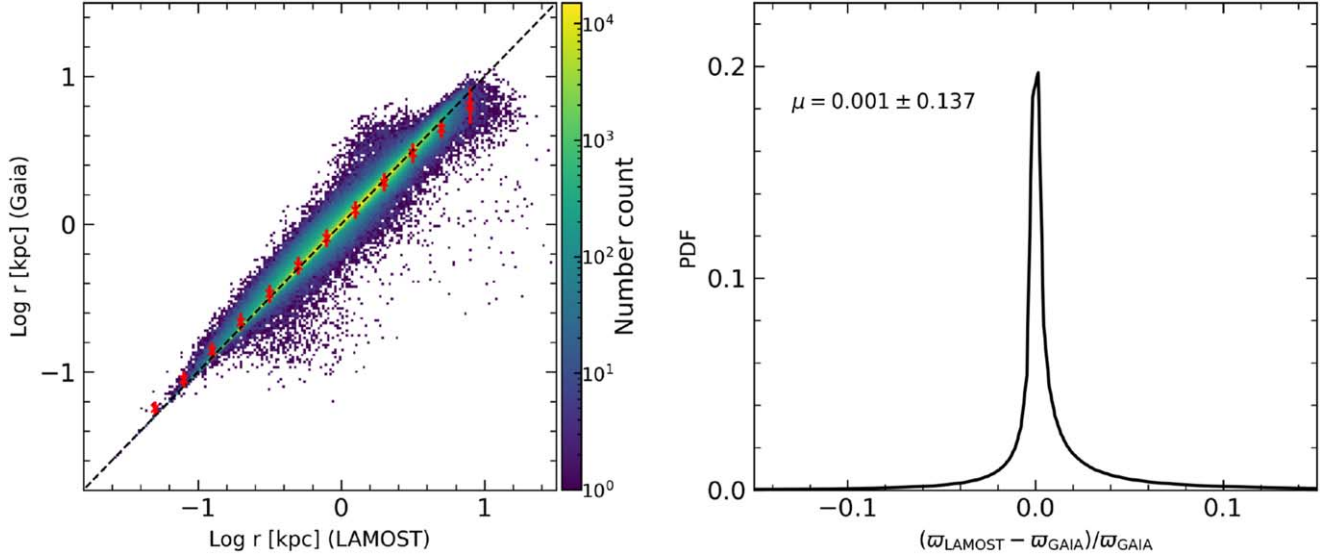
---

3 http://stev.oapd.inaf.it/cgi-bin/cmd

Figure 1. (Left) Comparing distance derived by our Bayesian method with that of Gaia EDR3 parallax. The black-dashed line shows the one-to-one line. The red crosses indicate the median values, and the error bars show the 1σ dispersion. The color bar on the right indicates the number density. (Right) The fractional difference distribution of parallax. The values of median and dispersion are labeled on the top left. In the comparison with Gaia EDR3 data, the criterion of $\sigma_\varpi/\varpi < 0.1$ has been used to select parallax with high accuracy.

parameter estimation. Sanders & Das (2018) relied on a bin size of 0.01 dex in metallicity following similar methodology as us for stellar age estimation, but considering Gaia Data Release 2 (DR2) parallax. Anders et al. (2022) used 0.1 dex in metallicity bin size in their "StarHorse" code. Lin et al. (2022) adopted 0.04 and 0.03 dex bin sizes for metallicity in the range [−3, −1] and [−1, 0.5], respectively. Second, we imposed parallax constraints from Gaia EDR3 (Gaia Collaboration et al. 2021) in the observable vector $\boldsymbol{O}$, the errors of which are considered in Equation (4). We note that the systematic errors in the Gaia EDR3 parallaxes were corrected following the method suggested by Lindegren et al. (2021), which are also validated by independent checks (Huang et al. 2021; Zinn 2021; Wang et al. 2022a). These improvements lead to accurate estimates of stellar parameters as demonstrated in the following sections.

In this work, we utilize LAMOST DR7 with atmospheric parameter measurements for around 5 million stars (Wu et al. 2011; Luo et al. 2015). We cross-matched this catalog with Gaia EDR3 to obtain parallax. There are 204,878 stars without parallaxes available in Gaia EDR3, for which we set their parallax errors to infinite. This corresponds to a flat prior in the parallax constraint. In Appendix A, we discuss the uncertainties of distance, extinction, stellar mass, and age yielded by our method.

3. Comparison and Validation with External Measurements

In this section we compare the distance, extinction, mass, and age for stars measured with our Bayesian inference to those

in the literature. In this way, we have an overview on the accuracy and any systematic uncertainties of our measurements. In the following comparisons, we will select LAMOST data with signal-to-noise ratio (S/N) greater than 20 unless otherwise stated.

3.1. The Distance

The left panel of Figure 1 compares the distance derived by our Bayesian method with LAMOST data with that derived with Gaia EDR3 parallaxes. In this comparison, 10% uncertainties of Gaia EDR3 parallax data are selected for cross-matching with LAMOST data. The red crosses indicate the median values, and the error bars show the 1σ dispersion in each bin. Thanks to the constraints from the Gaia EDR3 parallax used in our Bayesian method, the distance derived is well correlated with the distance of Gaia EDR3 with tiny uncertainties. There are many stars with large deviation from the one-to-one straight line. There could be several reasons for these stars, e.g., large uncertainties in stellar atmospheric parameters from spectroscopic observation, and the mismatch of stars between LAMOST data and Gaia EDR3 due to large proper motion (PM).

The right panel of Figure 1 displays the distribution of fractional difference in the parallaxes between our measurement and Gaia EDR3, where the values of mean and dispersion are labeled on the top left. There are large extended wings in the distribution of fractional difference of parallaxes, which are linked to the stars with large deviation from the one-to-one
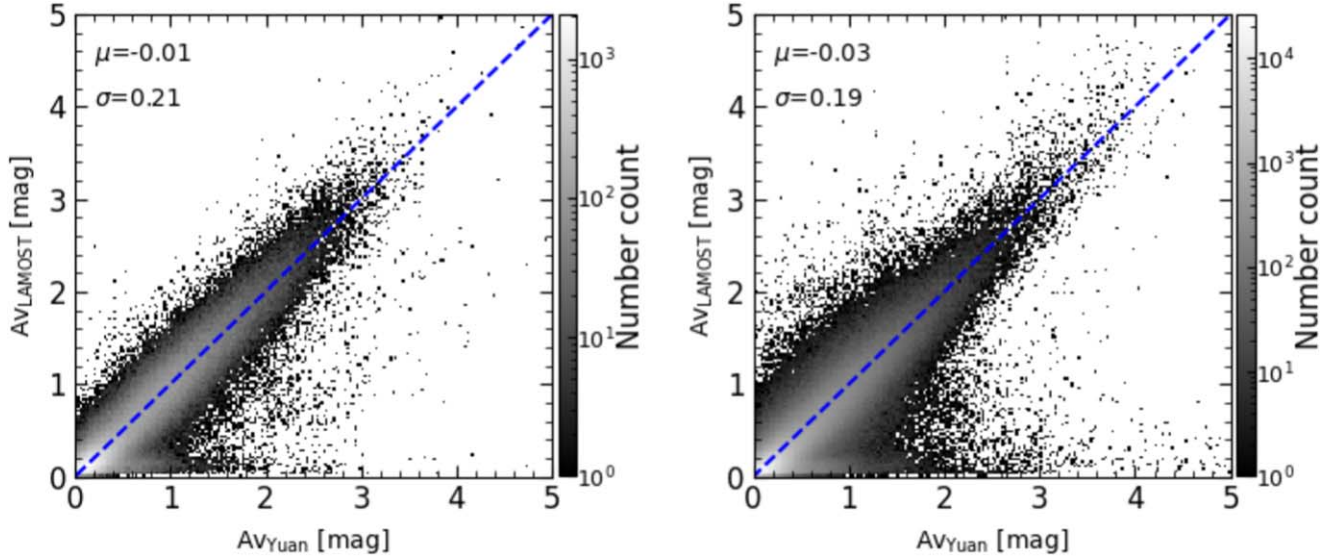
Figure 2. Comparing the extinction A_v derived by our Bayesian method with that from the star-pair method (Yuan et al. 2013). The left panel shows results of giants with $\log g < 3.5$, and the right panel exhibits dwarf stars with $\log g > 3.5$. The median difference and 1σ dispersion are labeled on the top left of each panel. The blue-dashed line in each panel indicates the one-to-one correspondence. The background gray-scale image in each panel signifies the number density of stars.

correspondences in the left panel of Figure 1. Several reasons may cause this.

3.2. The Extinction

In this section we compare the extinction measured by our Bayesian method with measurement with the star-pair technique (Yuan et al. 2013). This technique has been applied to LAMOST data to estimate the extinction of individual stars, which is provided in the Value-Added Catalog of the LAMOST survey (e.g., Xiang et al. 2017b). In our method, we used the Cardelli et al. (1989) extinction curve with $R_V = 3.1$ to calculate the extinction of the Ks band (A_{K_s}), which provides a relation $A_{K_s} = 0.117A_v$ (for details please refer to Wang et al. 2016b). With the star-pair technique, Yuan et al. (2013) suggest the extinction coefficient of $R(Ks) = 0.306$. To make a fair comparison, we used the same extinction curve (Cardelli et al. 1989) to convert their A_{K_s} to Av.

Figure 2 features the comparison of Av between our extinction and that derived by the star-pair method in the LAMOST Value-Added Catalog DR2 (Xiang et al. 2017b), separated into giant ($\log g < 3.5$, left panel) and dwarf ($\log g > 3.5$, right panel) stars. The median difference of Av derived from the two independent methods is very small, i.e., 0.01 and 0.03 mag for giants and dwarfs, respectively, with a dispersion of ∼0.2 mag.

3.3. The Stellar Mass

In this section, we compare the stellar mass to other measurements. We note that it is only possible for mass and age

estimation for evolved stars, since the identification of mass and age using theoretical isochrones is very difficult for main sequence stars. We will hereafter consider evolved stars with $\log g < 3.9$ for their mass and age validation in this section and the following section.

Pinsonneault et al. (2018) derived masses and ages for 6676 evolved stars ($1 < \log g < 3.5$) with Kepler asteroseismic data and APOGEE spectroscopic parameters. We cross-matched this catalog with LAMOST DR7, and found that most of these stars were observed by LAMOST.

Figure 3 compares our Bayesian stellar mass to that derived by Pinsonneault et al. (2018). There is good agreement between the stellar mass estimated by these two independent measurements, while there is a general trend that at high mass the M_{Kepler} tends to be larger. The mean difference for the whole sample is 0.06 $M_\odot$, with a dispersion of 0.23 $M_\odot$. The median value and 68 percentile interval of fractional difference are $0.06^{+0.22}_{-0.12}$.

Xiang et al. (2017a) estimated stellar age and mass for ∼1 million main sequence turnoff and subgiant stars from the LAMOST spectroscopic survey. They also adopted a Bayesian algorithm with observation data derived from LAMOST spectra, including T_{eff}, absolute magnitude M_v, [Fe/H], and [α/Fe]. Even though they adopted the Bayesian algorithm as we did, there are great differences, e.g., the prior function, isochrone models, and, more importantly, constraints given by Gaia parallax. Therefore, it is interesting to compare the results from these two methods. Figure 4 compares the estimated stellar masses, which show good correlation. The red stars indicate the median value at each bin, which shows that the
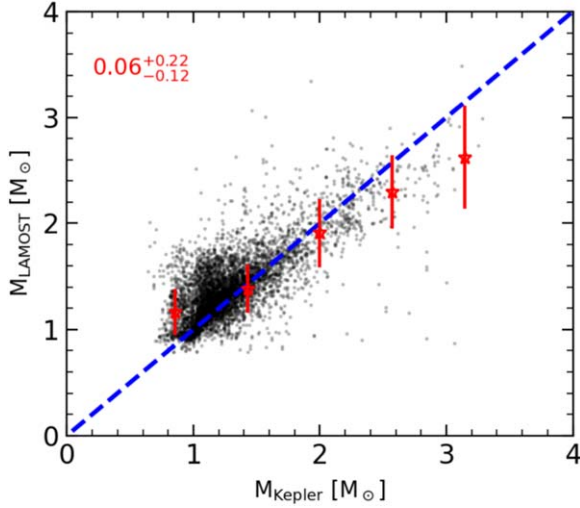
Figure 3. Comparing stellar mass estimated by our Bayesian method to that derived with asteroseismic data for giant stars (Pinsonneault et al. 2018). The red stars indicate the median value for each bin, and error bars signify the dispersion in each bin. The blue-dashed line indicates one-to-one correspondence. The median value and 68 percentile interval of fractional difference distribution are labeled on the top left with red color.
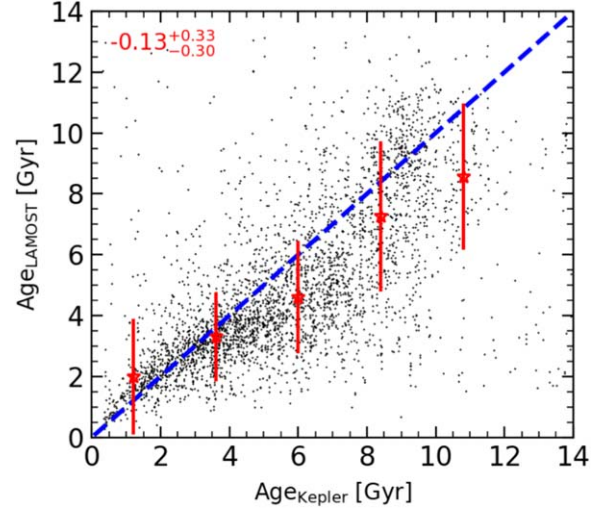

Figure 5. Comparing stellar age measured in current work for giant stars to that from Pinsonneault et al. (2018) with asteroseismic data. The blue-dashed line indicates the one-to-one correspondence, and the red stars and error bars show the median value and 1σ dispersion in each bin respectively. The red numbers on the top left express the median value and 68 percentile interval of fractional difference distribution.
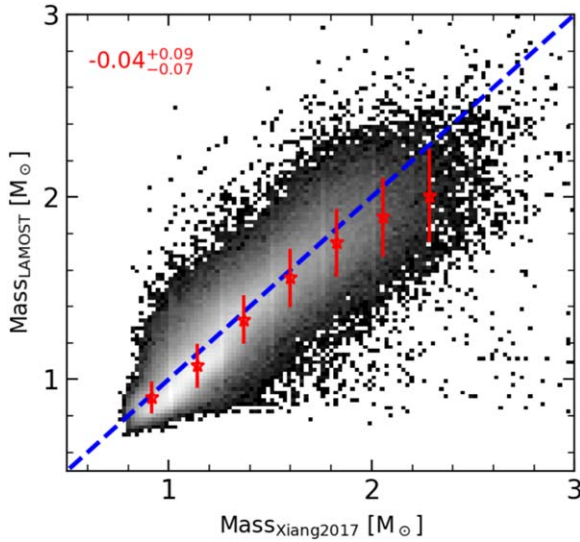
Figure 4. Comparing stellar mass estimated with our Bayesian method to that from Xiang et al. (2017a) for main sequence turnoff stars and subgiants. The spectral data with $S/N > 20$ were used in this comparison. The blue dashed line indicates the one-to-one correspondence. The median value and 68 percentile interval of fractional difference distribution are labeled on the top left with red color.
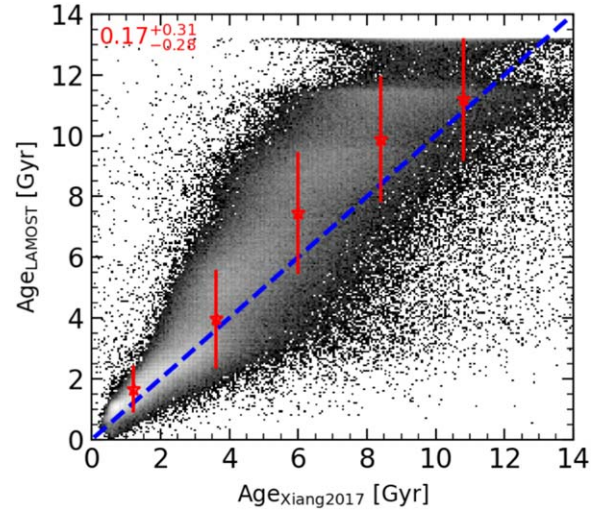

Figure 6. Comparing the stellar age for main sequence turnoff stars and subgiants in the current work to that from Xiang et al. (2017a). The blue-dashed line shows the one-to-one correspondence, and red stars and error bars indicate the median and dispersion in each bin respectively. The median value and 68 percentile interval of fractional difference distribution are labeled on the top left with red color.

mass from Xiang et al. (2017a) is slightly larger than ours at large mass range. The median difference is only 0.05 $M_\odot$, and dispersion is 0.12 $M_\odot$. The median value of fractional difference is 0.04, and the 68 percentile interval of the fractional difference is $[-0.11, 0.05]$.

3.4. The Age

In this section, we compare our stellar age to those estimated from other independent methods. As noted in the above section, only evolved stars with $\log g < 3.9$ are considered in this section.
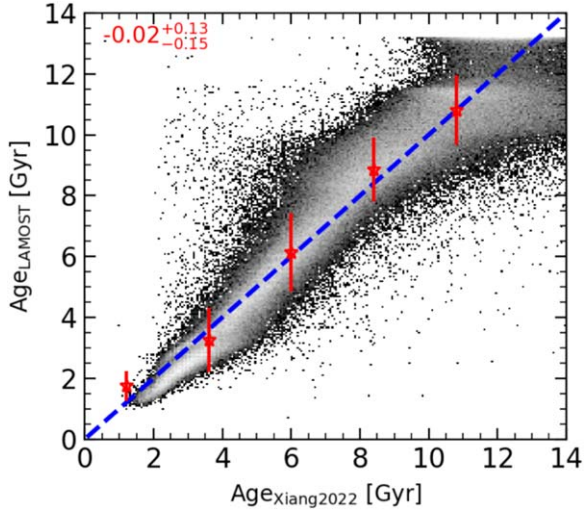
Figure 7. Comparing the stellar age for subgiant samples in the current work to that from Xiang & Rix (2022). The blue-dashed line indicates the one-to-one correspondence. The red stars and error bars signify the median and dispersion in each bin. The median value and 68 percentile interval of the fractional difference distribution are labeled on the top left.
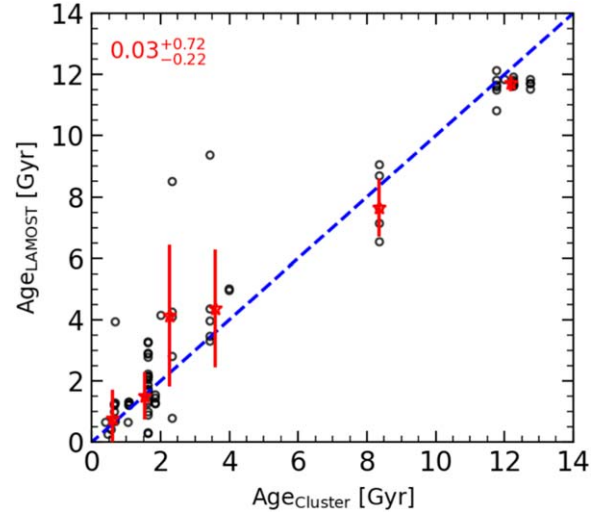

Figure 8. Comparing individual star ages of cluster members measured in the current work to the age of clusters in the literatures. The blue-dashed line indicates the one-to-one correspondence. The red stars and error bars signify the median value and dispersion in each bin respectively. The median value and 68 percentile interval of the fractional difference distribution are labeled on the top left.

Table 1
Age and Reference for the Star Clusters Used in the Current Work

Cluster Name	Age (Gyr)	Reference
Berkeley 17	3.98	Donor et al. (2020)
Berkeley 71	1.05	Donor et al. (2020)
Berkeley 9	2.00	Donor et al. (2020)
NGC 1245	1.06	Donor et al. (2020)
NGC 1664	0.56	Donor et al. (2020)
NGC 1857	0.47	Donor et al. (2020)
NGC 1907	0.40	Donor et al. (2020)
NGC 2304	0.68	Donor et al. (2020)
NGC 2420	2.32	Donor et al. (2020)
NGC 2682	3.43	Donor et al. (2020)
NGC 4147	12.25	VandenBerg et al. (2013)
NGC 5024	12.25	VandenBerg et al. (2013)
NGC 5272	11.75	VandenBerg et al. (2013)
NGC 6205	12.00	VandenBerg et al. (2013)
NGC 6341	12.75	VandenBerg et al. (2013)
NGC 6791	8.3	Brogaard et al. (2021)
NGC 6811	0.64	Donor et al. (2020)
NGC 6819	1.62	Donor et al. (2020)
NGC 7078	12.75	VandenBerg et al. (2013)
NGC 7789	1.84	Donor et al. (2020)

The asteroseismic data are powerful for constraining stellar mass, and then stellar age. Figure 5 compares stellar age estimated with asteroseismic data for giant stars from Pinsonneault et al. (2018). The stellar ages estimated with these two different methods are consistent with each other, while the ages from Pinsonneault et al. (2018) are larger than

ours when ages are greater than 5 Gyr. The median value of fractional difference is 13%.

The ages of main sequence turnoff stars and subgiants can be relatively accurately estimated, since in these evolution stages, the atmospheric parameters of these stars vary significantly with age. However, for the main sequence stars it is difficult to estimate their ages as the isochrones of different ages are tightly crowded together (Xiang et al. 2017a). Xiang et al. (2017a) measured the ages for ~1 million stars' main sequence turnoff and subgiant stars selected from the Value-Added Catalog of the LAMOST Galactic survey (Xiang et al. 2017b). Figure 6 compares our stellar ages to those in Xiang et al. (2017a). The stellar ages are correlated well, while there is systematic offset for age larger than 4 Gyr. We discuss this large systematic offset in Appendix B. We note there is a cutoff of age at ~13.2 Gyr in our age estimation, which results from an imposed limitation by the Padova isochrones.

Recently, Xiang & Rix (2022) applied a Bayesian method on LAMOST DR7 to derive stellar age for subgiant stars. There are several differences between theirs and ours. First, the input parameters are different. They used M_K, $T_{\rm eff}$, [Fe/H], and [α/Fe], Gaia parallax, and Gaia and 2MASS photometries. We note that stellar labels used in their method are derived from the data-driven Payne (DD-Payne) approach, while what we used are from the LAMOST pipeline LASP (Wu et al. 2011; Luo et al. 2015). Second, they adopted *YY* stellar isochrones, while the Padova isochrones are utilized in our method. Third, the extinction estimation is different. They estimated the extinction with a methodology of the intrinsic colors empirically inferred
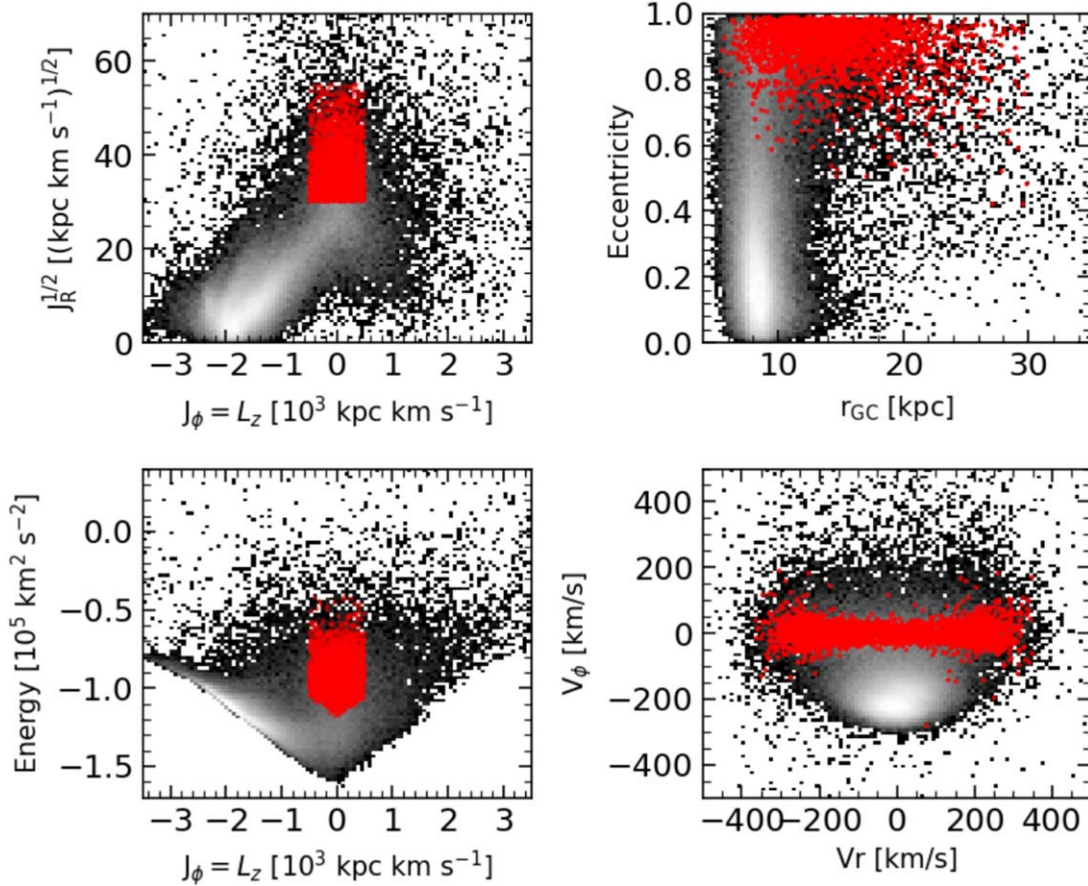
Figure 9. The kinematic distribution of stars. GSE stars are selected following Feuillet et al. (2021) labeled by red points. The black density shows our main samples. For details on the sample selection please refer to the text.

from their stellar parameters, while our extinction is calculated self-consistently in the Bayesian inference. Figure 7 compares the stellar ages estimated in the current work to those from Xiang & Rix (2022). Even though there are many differences in deriving the stellar age, both methods produce a well correlated stellar age. This may reflect the facts that the age of subgiant stars is well determined with high precision, and that the constraints from Gaia parallax are important for retrieving accurate age. We note there is only a 0.11 Gyr systematic offset in this comparison, with fractional difference being $-0.02^{+0.13}_{-0.15}$. In Figure C1 we also check the relation between stellar age and metallicity for comparison with Xiang & Rix (2022).

Star members for a star cluster are believed to form from a single molecular cloud at the same time. Therefore, the star members share a single age which is the cluster age. A star cluster provides unique opportunities to test the age accuracies of stellar age estimates. In this section, we compare individual stellar ages to cluster ages from the literature.

In order to identify the member stars for a given cluster, we use two well studied star cluster catalogs in which star members for each cluster have been carefully identified. One is

from Mészáros et al. (2013), which provide 20 star clusters including both open and globular clusters. These clusters served as calibration to the pipeline of APOGEE in Sloan Digital Sky Survey-III (SDSS-III). The other is the Open Cluster Chemical Abundances and Mapping survey from Donor et al. (2020), which provides large uniform, infrared-based spectroscopic data for 128 open clusters to constrain the key Galactic dynamical and chemical evolution parameters. The cluster member stars have been selected based on stellar radial velocities, PMs, spatial location, and derived metallicity as membership discriminators. A visual inspection by several authors for each cluster is performed to guarantee the quality of member selection.

We cross-matched these two star cluster catalogs with LAMOST DR7 entries, and selected cluster member stars with $\log g < 3.9$. Finally, we obtain 79 members belonging to 20 clusters. In Table 1 we list the clusters used in the current work and the age adopted for each cluster and their references. For the open clusters, the ages are adopted from Donor et al. (2020), while for the globular clusters their ages are taken from VandenBerg et al. (2013). One special case is the open cluster
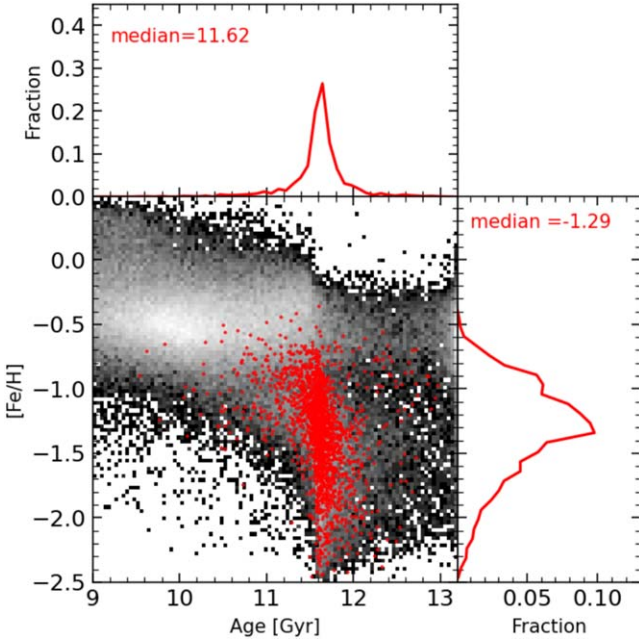
Figure 10. (Bottom left) Age–metallicity relation for giant stars with log $g < 3.9$. The GSE stars selected following Feuillet et al. (2021) are displayed with red points. (Top) Age distribution for GSE stars. (Right) Metallicity distribution of GSE stars.

NGC 6791. The age of this old super-metal-rich ([Fe/H] = 0.4) open cluster has been studied by Grundahl et al. (2008) from comparisons with theoretical isochrones in the mass–radius diagram. They found the cluster is old with age from 7.7 to 9.0 Gyr, depending on the adopted models. This age is consistent with the estimate from the eclipsing binary method from Brogaard et al. (2021), 8.3 ± 0.3 Gyr. Therefore, 8.3 Gyr is adopted in our analysis.

Figure 8 compares individual star ages measured by our Bayesian method to the cluster age. Both ages agree well generally, even though there are some stars with large deviations from the one-to-one correspondence line. There are two clusters with large age deviations from our estimate, which are NGC 2420 and NGC 2682. We note the age of open cluster NGC 2682 is 3.43 Gyr estimated by Kharchenko et al. (2013) with stellar data from PPMXL and 2MASS. Based on Gaia DR2 its age is estimated to be 3.64 Gyr by Bossini et al. (2019) and 4.3 Gyr by Cantat-Gaudin et al. (2020). The new results based on Gaia data lead this cluster age to be closer to our estimation, ~4.0 Gyr.

4. The GSE Substructure in LAMOST Survey

In the above section we have extensively compared the distance, extinction, mass, and age, measured with our Bayesian method to those derived with completely independent methods. These comparisons validate our results. These data

will be valuable for studies on Galactic structure, formation, and evolution. Below we use these data to explore the properties of the Galactic massive substructure GSE.

In the following we will use these data to illustrate properties of the Galactic massive substructure, GSE, which is believed to be the relic of an ancient major merger of the Milky Way (i.e., Belokurov et al. 2018). In order to select samples with better kinematic and age measurements, we use strict selection criteria as the following: to have better kinematic data from Gaia EDR3, (1) renormalized unit weight error < 1.2, (2) ipd_gof_harmonic_ amplitude $\leqslant 0.1$, and (3) ipd_frac_multi_ peak < 2. To avoid disk star contamination, we require $|b| > 30°$. To have better age measurement from LAMOST data, we select giant stars with $\log g < 3.9$ and $S/N > 30$. Finally, there are 161,523 stars selected, which are our main samples for comparison with GSE members.

To select a clean sample of GSE substructure, we followed the method of Feuillet et al. (2021) based on the radial action and angular momentum, $\sqrt{J_R}$ versus L_z. They found that the simple selection criteria $30 \leqslant \sqrt{J_R} \leqslant 55$ (kpc km s^{-1})$^{1/2}$ and $-500 \leqslant L_z \leqslant 500$ kpc km s^{-1} provide a clean and least-contaminated sample (Feuillet et al. 2020, 2021). We also impose $|z| > 5$ kpc to avoid disk star contamination. It is well known that GSE dominates the inner halo region, so we further constrain the sample stars with $r < 30$ kpc (Naidu et al. 2020). Finally, there are 2371 stars selected as members of GSE.

In order to calculate the physical quantities, we have used a Milky Way potential of Eilers et al. (2019), which is similar to Wang et al. (2022b) and Jiao et al. (2021). The values of J_R, energy, and orbital information are calculated with AGAMA (Vasiliev 2019). Figure 9 shows the kinematic distribution of main samples (black points) and GSE stars (red points). The top-left panel of Figure 9 depicts the selection of GSE members. The top-right panel affirms that most GSE member stars have eccentricity greater than 0.8, which is consistent with former studies (e.g., Naidu et al. 2020).

Figure 10 features the metallicity–age relation. The GSE stars have median [Fe/H] = −1.29, which is consistent with recent results (Helmi et al. 2018; Mackereth et al. 2019; Matsuno et al. 2019; Sahlholdt et al. 2019). In the literature, ~0.1–0.2 dex higher metallicity is also observed, e.g., Naidu et al. (2020) found the GSE has a higher metallicity [Fe/H] = $-1.15^{+0.24}_{-0.33}$, and Feuillet et al. (2020) found [Fe/H] = -1.17 ± 0.34.

Most GSE stars have ages in the range of 11.2–12.2 Gyr, and these old ages may indicate that the merger was completed around 11 Gyr ago, which is consistent with the results by Xiang & Rix (2022), and 1 Gyr older than Helmi et al. (2018). The median age of GSE stars is 11.6 Gyr. This median age is younger than the estimation from Gallart et al. (2019), who find a median age of 12.37 Gyr. We note that there is a ridge at age ~11.6 Gyr, which could reflect the prior function adopted in our Bayesian

method. This indicates that further improvement is required to adopt a more sophisticated prior function. On the other hand, we note that the comparison with Xiang & Rix (2022) in Figure 7 has shown that there is a clear correlation in the ages, while there is no prior imposed in the age estimation of Xiang & Rix (2022) (which actually means a flat prior function).

5. Summary

We have measured distance and extinction for around 5 million stars observed in LAMOST DR7 with Bayesian inference, as well as stellar mass and age for giant stars. Compared to former work (Wang et al. 2016a, 2016b), we have imposed the parallaxes from Gaia EDR3 to constrain these parameters, which result in accurate results achieved, in particular for stellar age and mass. Comprehensive comparisons with measurements from independent methods are performed to validate these results. We have kinematically selected GSE member stars in LAMOST data, and studied their metallicity and age distribution, with which we demonstrated that this data set is valuable for studying Galactic archaeology. We found that GSE stars have median value of metallicity [Fe/H] = −1.29, which is consistent with literature. This corresponds to star formation in GSE being dominant 11.6 Gyr ago.

This huge data set is vital for constraining Galactic formation and evolution, and it will be released to the community as a Value-Added Catalog on the LAMOST website.

Acknowledgments

Appendix A
The Uncertainties of Stellar Parameters Output by our Method

In Figure A1 we show the uncertainties of stellar parameters output by our Bayesian method as a function of S/N. There are clear trends that the uncertainties decrease with increasing S/N. At S/N ∼ 20, the median values of uncertainties of distance, extinction are 4% and 0.1 mag. At this S/N the uncertainties of stellar mass and age for the stars with $\log g < 3.9$ are 10% and 25% respectively. Higher accuracy of stellar age can be achieved for subgiants as shown with blue squares in the bottom right panel of Figure A1. At S/N ∼ 40, the age uncertainty of subgiants is around 10%, and it decreases with increasing S/N.
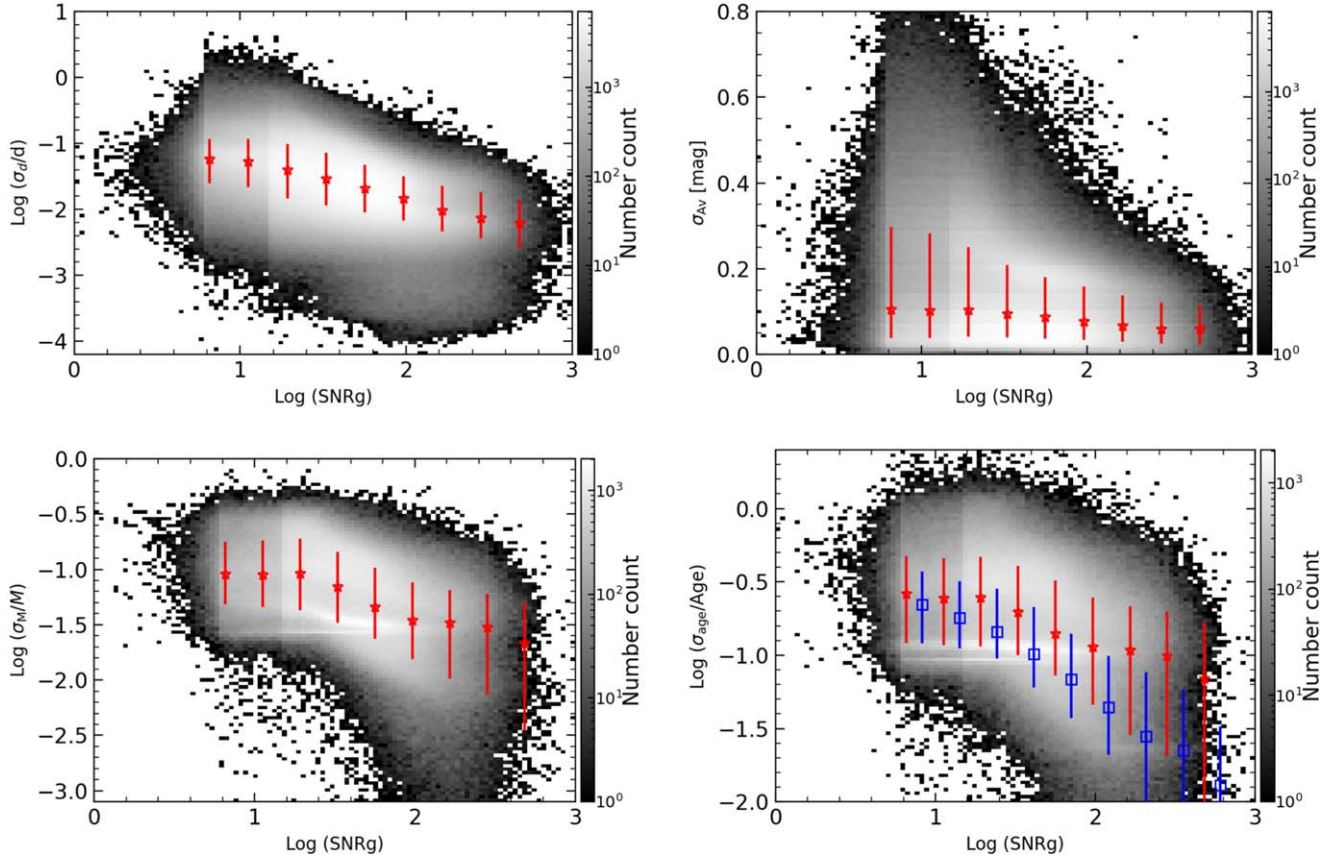
Figure A1. Uncertainties of distance, extinction, stellar mass, and age output by our method as a function of S/N. The red stars in each panel indicate the median value for each bin, and the error bars signify the 68 percentile distribution. All of the LAMOST DR7 samples are used in the distance and extinction panels, while in the stellar mass and age panels only stars with $\log g < 3.9$ are used. The blue squares and error bars in the stellar age panel show the median value and the 68 percentile distribution for subgiants with $3.4 < \log g < 3.9$ respectively.

Appendix B
Comparing the Stellar Ages Estimated by Xiang et al. (2017a) to those by Xiang et al. (2022) and this Work

Figure B1 compares the stellar ages estimated by Xiang et al. (2017b) to those estimated by Xiang & Rix (2022) and this work. There are systematic offsets in star age estimations between Xiang et al. (2017b) and Xiang & Rix (2022) and between Xiang et al. (2017b) and this work. The offset in age estimation between Xiang et al. (2017b) and this work is slightly larger than that between Xiang et al. (2017b) and Xiang & Rix (2022) in the age range from 6 to 11 Gyr.

Even though the work of Xiang et al. (2017b), Xiang & Rix (2022), and this paper used the LAMOST spectra, the stellar atmospheric parameters utilized in these three works are different. In our work the metallicity ([Fe/H]) is produced with LASP (Wu et al. 2011; Luo et al. 2015). Xiang & Rix (2022) follow a DD-Payne approach to derive the metallicity. Xiang et al. (2017b) derived the stellar metallicity with the

pipeline of LSP3, which was developed at Peking University. These different methods may lead to different scales in metallicity. Figure B2 compares the metallicities derived with these three approaches.

The [Fe/H] considered in Xiang & Rix (2022) is systematically lower than that of Xiang et al. (2017b). This may partly explain that even though Xiang & Rix (2022) and Xiang et al. (2017b) applied similar methods to derive stellar age, the age of Xiang & Rix (2022) is higher than that in Xiang et al. (2017b).

The [Fe/H] used in this work is well correlated with that of Xiang et al. (2017b). We note that we relied on Padova isochrones to derive stellar parameters, in which the α abundance cannot be accounted for. In the work of Xiang et al. (2017b) and Xiang & Rix (2022) the isochrones of *YY* are used, in which the α abundance has been taken into account. The ignorance of α could lead to a systematic overestimation of age (Xiang & Rix 2022), but as
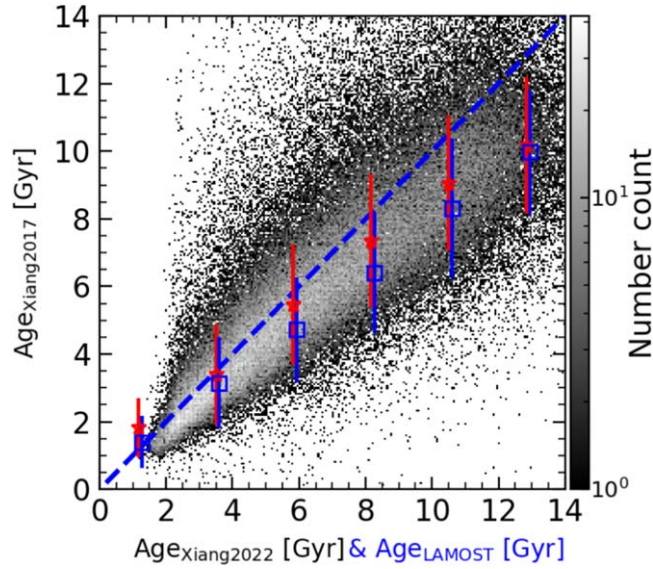
Figure B1. Comparing star ages estimated by Xiang et al. (2017b) to Xiang & Rix (2022) and the current work. The gray color map indicates comparison between Xiang et al. (2017b) and Xiang & Rix (2022), and the red stars and error bars signify the median value and dispersion in each bin for this comparison respectively. Also, blue squares and blue error bars correspond to the median value and dispersion in each bin for comparison respectively between Xiang et al. (2017b) and this work. The blue-dashed line indicates the one-to-one correspondence.
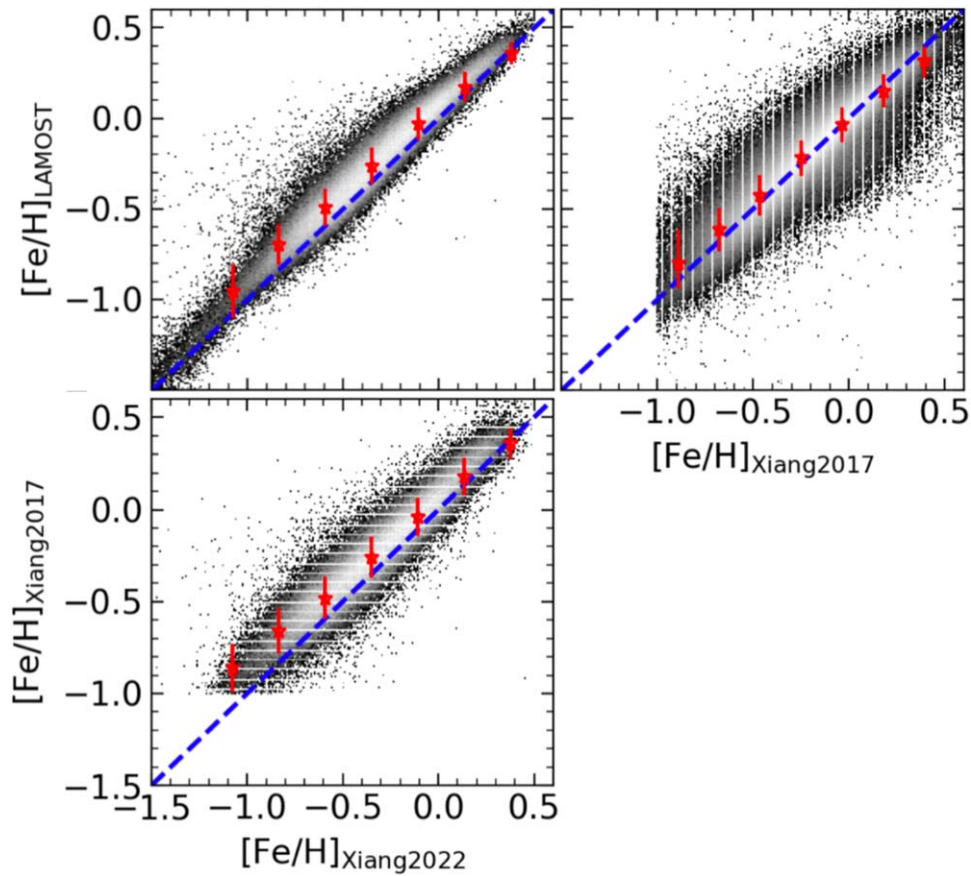

Figure B2. Comparing metallicities used in Xiang et al. (2017b), Xiang & Rix (2022), and this work.

demonstrated by Xiang & Rix (2022) this ignorance can only lead to a ~15% overestimation under the assumption that all stars have alpha abundance 0.2 dex. There is around a 30% difference in ages between this work and those in Xiang et al. (2017b). We note that there is also around a 20%–30% difference in ages for old stars between Xiang & Rix (2022) and Xiang et al. (2017b), both of which accounted for α abundance.

Even though the [Fe/H] used in Xiang & Rix (2022) is lower than that used in this work, the age estimated in both works is in good agreement. This may be due to the fact that our ignorance of α abundance is well compensated for by the higher metallicity.

Appendix C
The Stellar Age and Metallicity Relation

Figure C1 compares the relations of stellar age and metallicity by using our estimated age and that from Xiang & Rix (2022). There is general consistency between the two color maps, e.g., from 2 to 8 Gyr there are two branches in the metallicity relation, one has flat [Fe/H] ~ 0 dex, and the other has a [Fe/H] decrease with increasing age from 2 to 8 Gyr. At age greater than 8 Gyr, the metallicity decreases with age increasing in both maps. The similarity between the two maps reflects the fact that the ages estimated in this work and those in Xiang & Rix (2022) are very consistent as affirmed in Figure 7.
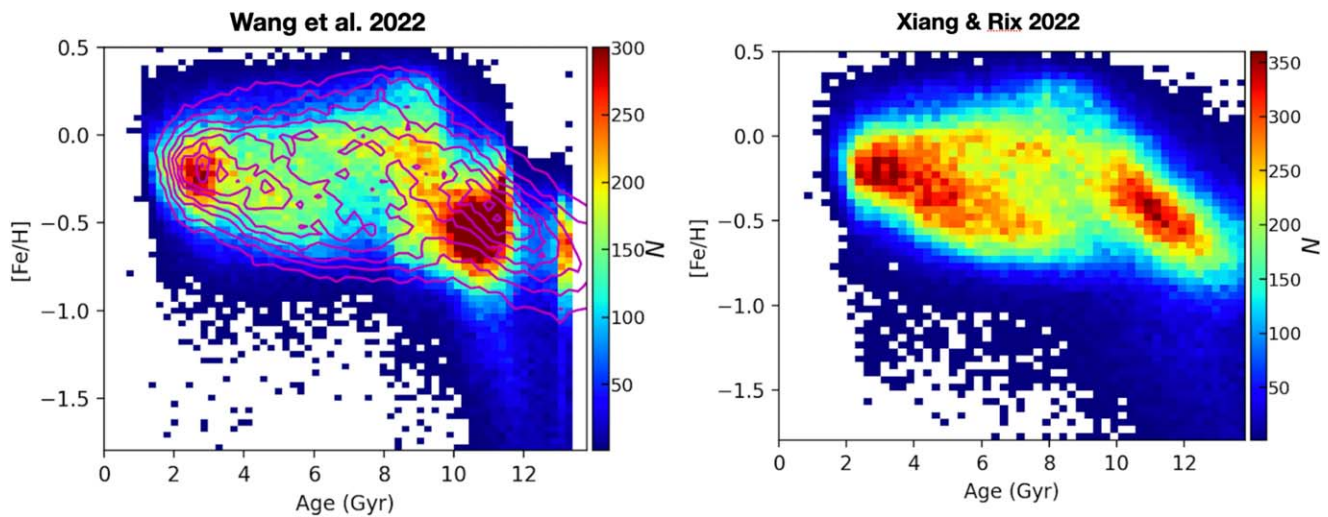

Figure C1. The stellar age and metallicity relation. The left color map shows the results from this work, while the right panel features the results from Xiang & Rix (2022). The magenta color contours overplotted in the left panel indicate the results from the right panel for comparison.

References

Aguirre Børsen-Koch, V., Rørsted, J. L., Justesen, A. B., et al. 2022, MNRAS, 509, 4344

Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91

Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611

Binney, J., Burnett, B., Kordopatis, G., et al. 2014, MNRAS, 437, 351

Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, A&A, 623, A108

Brogaard, K., Grundahl, F., Sandquist, E. L., et al. 2021, A&A, 649, A178

Burnett, B., & Binney, J. 2010, MNRAS, 407, 339

Burnett, B., Binney, J., Sharma, S., et al. 2011, A&A, 532, A113

Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Dalton, G., Trager, S., Abrams, D. C., et al. 2014, Proc. SPIE, 9147, 91470L

de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, Msngr, 175, 3

Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 735

Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2020, AJ, 159, 199

Eilers, A.-C., Hogg, D. W., Rix, H.-W., & Ness, M. K. 2019, ApJ, 871, 120

Feuillet, D. K., Feltzing, S., Sahlholdt, C. L., & Casagrande, L. 2020, MNRAS, 497, 109

Feuillet, D. K., Sahlholdt, C. L., Feltzing, S., & Casagrande, L. 2021, MNRAS, 508, 1489

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, 595, A2

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1

Gallart, C., Bernard, E. J., Brook, C. B., et al. 2019, NatAs, 3, 932

Grundahl, F., Clausen, J. V., Hardis, S., & Frandsen, S. 2008, A&A, 492, 171

Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Natur, 563, 85

Huang, Y., Schönrich, R., Zhang, H., et al. 2020, ApJS, 249, 29

Huang, Y., Yuan, H., Beers, T. C., & Zhang, H. 2021, ApJL, 910, L5

Jiao, Y., Hammer, F., Wang, J. L., & Yang, Y. B. 2021, A&A, 654, A25

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R. D. 2013, A&A, 558, A53

Lin, J., Casagrande, L., & Asplund, M. 2022, MNRAS, 510, 433

Lindegren, L., Bastian, U., Biermann, M., et al. 2021, A&A, 649, A4

Liu, X. W., Yuan, H. B., Huo, Z. Y., et al. 2014, in Setting the Scene for Gaia and LAMOST, Proc. IAU Symp., Vol. 298, ed. S. Feltzing et al., 310

Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095

Mackereth, J. T., Schiavon, R. P., Pfeffer, J., et al. 2019, MNRAS, 482, 3426

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94

Matsuno, T., Aoki, W., & Suda, T. 2019, ApJL, 874, L35

McMillan, P. J., Kordopatis, G., Kunder, A., et al. 2018, MNRAS, 477, 5279

Mészáros, S., Holtzman, J., García Pérez, A. E., et al. 2013, AJ, 146, 133

Naidu, R. P., Conroy, C., Bonaca, A., et al. 2020, ApJ, 901, 48

Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, ApJS, 239, 32

Queiroz, A. B. A., Anders, F., Santiago, B. X., et al. 2018, MNRAS, 476, 2556

Sahlholdt, C. L., Casagrande, L., & Feltzing, S. 2019, ApJL, 881, L10

Sanders, J. L., & Das, P. 2018, MNRAS, 481, 4093

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

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

VandenBerg, D. A., Brogaard, K., Leaman, R., & Casagrande, L. 2013, ApJ, 775, 134

Vasiliev, E. 2019, MNRAS, 482, 1525

Wang, C., Yuan, H., & Huang, Y. 2022a, AJ, 163, 149

Wang, J., Hammer, F., & Yang, Y. 2022b, MNRAS, 510, 2242

Wang, J., Shi, J., Pan, K., et al. 2016a, MNRAS, 460, 3179

Wang, J., Shi, J., Zhao, Y., et al. 2016b, MNRAS, 456, 672

Wu, Y., Luo, A. L., Li, H.-N., et al. 2011, RAA, 11, 924

Xiang, M., Liu, X., Shi, J., et al. 2017a, ApJS, 232, 2

Xiang, M., & Rix, H.-W. 2022, Natur, 603, 599

Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2017b, MNRAS, 467, 1890

Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

Yuan, H. B., Liu, X. W., & Xiang, M. S. 2013, MNRAS, 430, 2188

Zinn, J. C. 2021, AJ, 161, 214