



Atmospheric Parameters and Kinematic Information for the M Giant Stars from LAMOST DR9

Dan Qiu^{1,2}, Hao Tian^{3,1} , Jing Li⁴, Chao Liu^{1,2}, Lin Long⁴, Jian-Rong Shi^{2,5}, Ming Yang¹, and Bo Zhang¹ 
¹ Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China
tianhao@nao.cas.cn

² University of Chinese Academy of Sciences, 100049, China

³ Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, China

⁴ School of Physics and Astronomy, China West Normal University, Nanchong 637002, China

⁵ Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Received 2022 November 17; revised 2023 January 30; accepted 2023 February 7; published 2023 April 19

Abstract

A catalog of more than 43,000 M giant stars has been selected by Li et al. from the ninth data release of LAMOST. Using the data-driven method SLAM, we obtain the stellar parameters (T_{eff} , $\log g$, $[M/H]$, $[\alpha/M]$) for all the M giant stars with uncertainties of 57 K, 0.25 dex, 0.16 dex and 0.06 dex at $\text{SNR} > 100$, respectively. With those stellar parameters, we constrain the absolute magnitude in the K – band, which brings distance with relative uncertainties around 25% statistically. Radial velocities are also calculated by applying cross correlation on the spectra between 8000 and 8950 Å with synthetic spectra from ATLAS9, which covers the Ca II triplet. Comparison between our radial velocities and those from APOGEE DR17 and Gaia DR3 shows that our radial velocities have a system offset and dispersion around 1 and 4.6 km s^{-1} , respectively. With the distances and radial velocities combining with the astrometric data from Gaia DR3, we calculate the full 6D position and velocity information, which are able to be used for further chemo-dynamic studies on the disk and substructures in the halo, especially the Sagittarius Stream.

Key words: methods: data analysis – stars: atmospheres – Galaxy: disk – stars: abundances

1. Introduction

M giant stars are the kind of stars with high luminosity and low temperature, such as the tip of the red giant branch stars (rRGB stars), C-, O- and extreme asymptotic giant branch stars (AGB stars) and red supergiant stars. On one hand, the brightness means that they are able to be used to trace the distant volumes, which makes them a good tracer to reveal the accretion and merger events in the Milky Way by discovering and identifying the remnants of the relatively metal-rich stellar streams in the halo, especially for the Sagittarius system (Ibata et al. 1994), which is still suffering tidal disruption (Newberg et al. 2002; Belokurov et al. 2014; Koposov et al. 2015; Li et al. 2016, 2016b). Majewski et al. (2003) selected the M giant stars from 2MASS. Those samples clearly represented the Galactic disk and satellite galaxies, such as the Magellanic Clouds and the Sagittarius dwarf spheroidal galaxy. Li et al. (2016) also used the M giant stars to map the Sagittarius Stream and revealed more distant structure. On the other hand, the low temperatures indicate that most of the flux are distributed at the long wavelength bands, such as the K -band in the 2MASS system. Therefore, a further advantage is that the M giant stars suffer less extinction. This provides the opportunity to study the outer volumes of the disk with low latitude.

Though the M giant stars have significant advantages, they are not widely used for the halo and disk studies, especially comparing with the K giant stars (Liu et al. 2017; Xu et al. 2018; Tian et al. 2019, 2020; Xu et al. 2020). The first reason is that there are much fewer M giant stars than the K giant stars. The other one is that, because of the low temperature, there are molecular absorption bands in their spectra, which bring difficulties to constrain the stellar physical parameters, such as the abundances and the radial velocities which are quite important for further studies on the structures in the Milky Way. In recent years, with the development of observation equipment, many large surveys like Large sky Area Multi-Object fiber Spectroscopic Telescope (LAMOST; Wang et al. 1996; Su & Cui 2004; Cui et al. 2012; Deng et al. 2012; Luo et al. 2012; Zhao et al. 2012; Yan et al. 2022), Sloan Digital Sky Survey (SDSS; Ahumada et al. 2020), have collected a large amount of photometric and low resolution spectral data of M-type stars. Meanwhile kinds of stellar parameter pipelines have been developed. However, the Stellar Parameter Pipelines, like the LAMOST Stellar Parameter Pipelines (LASP; Wu et al. 2014; Luo et al. 2015), which designed based on the University of Lyon Spectroscopic Analysis Software (ULYSS; Koleva et al. 2009), is unable to derive accurate stellar parameters for these M giant with low resolution spectra

because of those molecular bands. Special attention should be paid on those low temperature stars, such as the pipeline for the Apache Point Observatory Galactic Evolution Experiment (APOGEE DR17; Abdurro'uf et al. 2022).

The most common and efficient method to determine the stellar parameters is to fit the observed spectra with the synthetic spectra, which has been successfully applied on the RGB stars for decades. Bizyaev et al. (2006, 2010) determined the stellar parameters of hundreds of red giants with high resolution spectra by comparing the observed spectra to a synthetic stellar spectra library ATLAS9 (Kurucz 1993), including few low temperature stars. With a similar method, Carlin et al. (2018) determined T_{eff} , [Fe/H] and $\log g$ of 42 K/M giants with high resolution ($R \sim 67,500$) spectra from Gemini Remote Access to CFHT ESPaDOnS Spectrograph (GRACES; Tollestrup et al. 2012; Chene et al. 2014), using a synthetic spectrum atmospheric model generated with Castelli & Kurucz (2003) and Dartmouth isochrones (Dotter et al. 2008). The typical uncertainties on T_{eff} , [Fe/H] and $\log g$ are 115 K, 0.09 dex and 0.18 dex, respectively. More recently, Ding et al. (2022) derived stellar atmospheric parameters of LAMOST M-type stars from MILES library interpolator by applying the χ^2 minimization performed by the ULYSS package. For M giants, the uncertainties of T_{eff} , $\log g$ and [Fe/H] are 58 K, 0.19 dex and 0.26 dex, respectively.

With rapid development, the machine learning has been applied on deriving the stellar parameters frequently in recent years (Howard 2017; Antoniadis-Karnavas et al. 2020; Galgano et al. 2020; Zhang et al. 2020). More recently, Wang et al. (2020) have designed a neural network model, named SPCANet, to determine T_{eff} , $\log g$ and 13 chemical abundances for medium resolution spectroscopy from LAMOST Medium Resolution Survey (MRS) data sets ($R \sim 7500$) (Liu et al. 2020), including many M giant spectra. The precision of T_{eff} , $\log g$ and [Fe/H] are 119 K, 0.17 dex and 0.06 dex, respectively.

In this work, we use a data-driven method Stellar LAbel Machine (SLAM), which is developed by Zhang et al. (2020) to derive the stellar parameters of M giants from low resolution ($R \sim 1800$) spectra of LAMOST, including T_{eff} , [M/H] and $\log g$. SLAM has shown good performance in deriving stellar parameters. e.g., Zhang et al. (2020) used SLAM to determine T_{eff} , $\log g$ and [Fe/H] from low-resolution spectra for ~ 1 million LAMOST DR5 K giants with random uncertainties are 50 K, 0.09 dex and 0.07 dex, respectively. Li et al. (2021) measured T_{eff} and [M/H] of M dwarfs by training SLAM with LAMOST low-resolution spectra and APOGEE stellar labels, T_{eff} and [M/H] are in agreement to within 50 K and 0.12 dex compare to the APOGEE observation. Guo et al. (2021) adopted SLAM to predict T_{eff} , $\log g$, [M/H] and projected rotational velocity ($v \sin i$) for 3931 early-type stars from LAMOST low-resolution survey. The uncertainties of T_{eff} , $\log g$ and $v \sin i$ are 1642 K, 0.25 dex and 42 km s^{-1} ,

respectively. They also determined the above four parameters by using SLAM for 578 early-type stars from LAMOST medium-resolution survey (MRS). The uncertainties are 2185 K, 0.29 dex and 11 km s^{-1} for T_{eff} , $\log g$ and $v \sin i$, respectively.

This paper is organized as follows: a brief description about the sample will be presented in Section 2. In Section 3 we will show the results of the radial velocities and the stellar parameters. The validation of the parameters will also be discussed in Section 3. Then the application of this value added catalog will be shown in Section 4. Finally, the summary will be given in Section 5.

2. Data

2.1. M Giants

Millions of spectra have been obtained by LAMOST in the last 10 years (Cui et al. 2012; Deng et al. 2012; Zhao et al. 2012), including thousands of M giant stars. Though there are many molecular absorption bands in the spectra, Zhong et al. (2015) and Li et al. (2019) have successfully selected the M giant stars using a similar method on the spectra from LAMOST. Recently, using the similar method, Li et al. (2023) have selected more than 43,000 reliable M giant stars from LAMOST DR9. Those M giant stars are separated from the M dwarf stars using the spectra index of TiO5 versus that of CaH2+CaH3, which has been proved to be a quite efficient way (Zhong et al. 2015). The M type giant and dwarf stars are behaved at two different clumps in the color-color diagram, $(W1 - W2)_0$ versus $(J - K)_0$. Finally the contamination of few white dwarf stars and those dwarf stars located in the overlapped region with the M giant stars in the spectra index diagram and the color-color diagram can be further reduced by applying the distance provided by Gaia DR3 (Zhong et al. 2019). After all those selections, more than 43,000 M giant stars are left, which will be used in this paper.

There are 28,610 M giant stars both in Li et al. (2023) and LAMOST officially released M giant stars. We cross-match these 28,610 M giants with APOGEE DR17, and obtain 2123 common stars with both signal-to-noise ratios (SNRs) from LAMOST and APOGEE larger than 50. The SNR from LAMOST is defined as $\text{SNR} = (\sum_{i=0}^N \text{flux}_i * \sqrt{\text{invvar}_i})/N$, where flux_i and invvar_i are the flux and inverse variance of the i th pixel of a spectrum, N is the number of pixels of the corresponding spectrum. In Figure 1, we represent the comparison of metallicity, surface gravity and temperature between LAMOST and APOGEE of those 2123 stars, respectively. It indicates that there are great differences in these three stellar parameters between LAMOST and APOGEE, especially for [M/H] and $\log g$. 1840 of those common stars are not provided available parameters, but the constants -1 or 0 for the metallicity. That means the LAMOST pipeline cannot give reliable metallicity and surface gravity for those low temperature stars. Due to the low temperature, the parameters from APOGEE

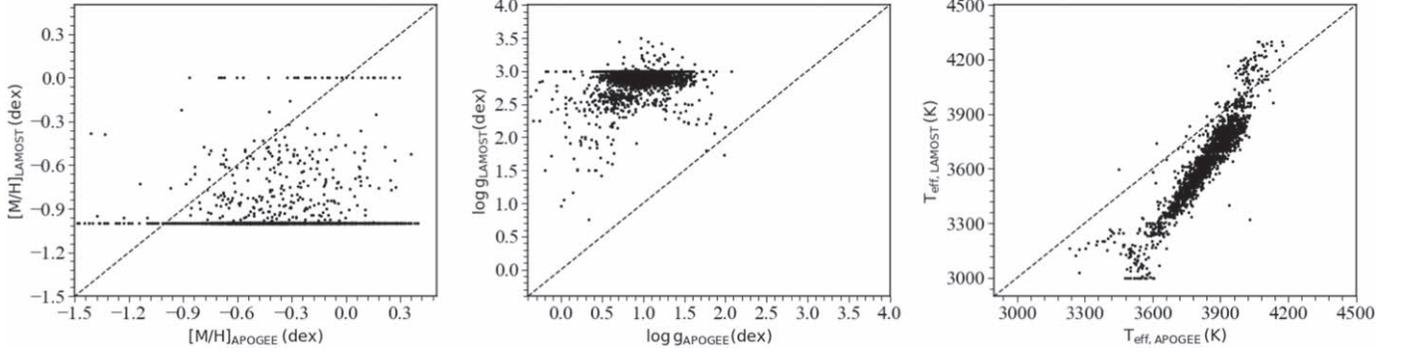


Figure 1. The comparison of metallicity, surface gravity and effective temperature between LAMOST and APOGEE of 2123 M giant stars, respectively.

with near-infrared spectra are more reliable. Therefore, the results from APOGEE are adopted during our constraint on the stellar parameters.

2.2. Training Set

In this work, we train the SLAM model to predict the stellar parameters of M giant stars. The training set is used to train the model of the stellar labels versus the spectra. This requires that the spectra of the training set should have high SNRs and the accurate labels to each spectrum, e.g. $[M/H]$, $[\alpha/M]$, T_{eff} and $\log g$ in this work. To this way, we cross-match the whole sample of M giant stars with APOGEE DR17 (Abdurro’uf et al. 2022) and obtain a common catalog of 4473 M giant stars. With following criteria, we further constrain the accuracy of the stellar parameters T_{eff} , $\log g$, $[M/H]$ and $[\alpha/M]$, finally 3670 M giant stars are left for the training set, including the accurate stellar parameters from APOGEE DR17 and high SNR spectra from LAMOST DR9.

1. $\text{SNR} > 50$
2. $\sigma_{[M/H]} < 0.1$
3. $\sigma_{T_{\text{eff}}} < 20$
4. $\sigma_{\log g} < 0.1$
5. $\sigma_{[\alpha/M]} < 0.05$

where SNR is the mean signal-to-noise ratio of LAMOST spectra as described in Section 2.1, $\sigma_{[M/H]}$, $\sigma_{T_{\text{eff}}}$, $\sigma_{\log g}$ and $\sigma_{[\alpha/M]}$ are the uncertainties of the metallicity $[M/H]$, the effective temperature T_{eff} , the surface gravity $\log g$ and the alpha abundance $[\alpha/M]$ of the M giant stars, which are provided by APOGEE DR17.

The excluded 803 stars will be used in Section 3.2.4 to verify the self consistency of the trained model in deriving the stellar parameters. Figure 2 shows the Hertzsprung Russell diagram (HRD) of the 3670 common M giants stars, which are color-coded by the metallicity $[M/H]$. We find that all the training samples are located in the ranges of $-1.5 < [M/H] < 0.5$ dex, $3200 < T_{\text{eff}} < 4300$ k, $-0.4 < \log g < 2.5$ dex. For comparison,

we also represent the isochrones from the PAdova and TRIeste Stellar Evolution Code (PARSEC; Bressan et al. 2012) with the dashed lines of the same age of 3 Gyr and different metallicities of 0.3, 0, -0.3 and -0.6 dex. As showed in the later results, this is reasonable for our sample, majority of which are the thin disk members. Statistically speaking, the distribution of the training stellar labels are consistent with the stellar evolution model and the M giant stars are mainly the metal-rich stars.

3. Method and Results

3.1. Radial Velocity

The radial velocity is derived by using cross correlation based method *laspec* (Zhang et al. 2021). It is applied on the spectra of M giant stars with wavelength between 8000 and 8950 Å from LAMOST DR9, where Ca II triplet is included in this band. The results can be evaluated by the following equation,

$$\text{CCF}(v|F, G) = \frac{\text{Cov}(F, G(v))}{\sqrt{\text{Var}(F)\text{Var}(G(v))}} \quad (1)$$

where F is the normalized observed spectrum of the given M giant star, while G is the synthetic spectra from ATLAS9 (Allende Prieto et al. 2018), also with wavelength between 8000 to 8950 Å with a shift caused by the radial velocity. It is noteworthy that the spectra used in Section 3.2 have been shifted back to the rest frame by the radial velocity.

To evaluate the results, we cross match our sample with Gaia DR3 (Gaia DR3; Gaia Collaboration 2022; Katz et al. 2022) and APOGEE DR17 (Abdurro’uf et al. 2022) to obtain two data sets of common stars. There are 32,787 and 4888 common stars with Gaia DR3 and APOGEE DR17, respectively. Then the comparison of our results with Gaia DR3 and APOGEE DR17 are represented by the blue and red symbols in Figure 3, respectively. It shows a good agreement between our results with that from Gaia DR3 and APOGEE DR17. The histogram of the radial velocity difference $\Delta rv = rv - rv_{\text{Ref}}$ are shown in

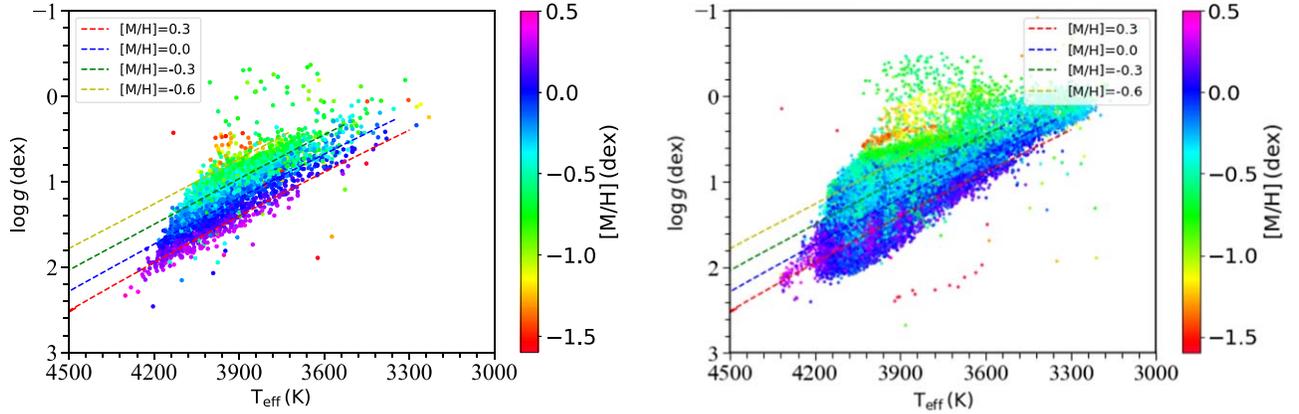


Figure 2. The distributions in the HR diagram of the training set and the prediction M giants with $\text{SNR} > 50$ are shown in the left and right panels, respectively. The color is coded by the metallicity $[M/H]$. In the left panel, the stellar parameters of 3670 M giant stars are from APOGEE. In the right panel, the stellar parameters of the M giant sample with $\text{SNR} > 50$ are predicted by SLAM. The dashed lines in both panels represent the isochrones from PARSEC model with the same age 3 Gyr but different metallicities, i.e., -0.6 , -0.3 , 0 and 0.3 dex for the yellow, green, blue and red lines, respectively.

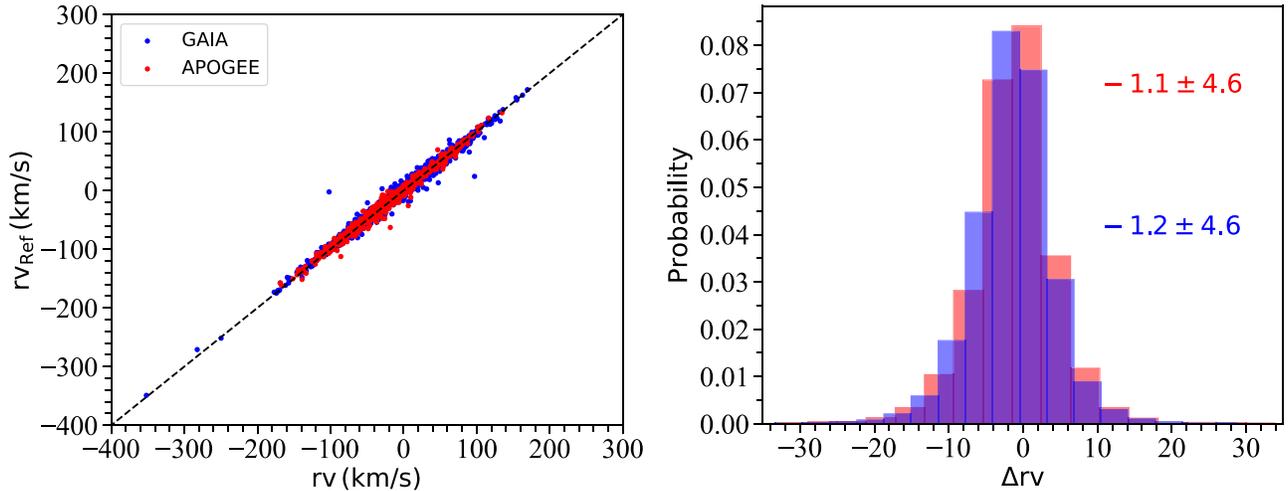


Figure 3. Compared the radial velocity from this work with that in Gaia DR3 (32,787 stars) and APOGEE DR17 (4888 stars), as displayed by the red and blue colors in the left panel, respectively. The right panel displays the corresponding histograms of radial velocity difference Δrv , where rv is the radial velocity derived in this work, and rv_{Ref} is the radial velocity from Gaia DR3 (blue) and APOGEE DR17 (red).

the right panel. The system offset and scatter values of Δrv are around $\sim 1 \text{ km s}^{-1}$ and 4.6 km s^{-1} , respectively, for both comparison results. It indicates that the radial velocity of M giants derived from Ca II Triplet lines is consistent with that of Gaia DR3 and APOGEE DR17.

3.2. Stellar Parameter

The SLAM is a data-driven method based on support vector regression (SVR) (Zhang et al. 2020). SVR is a robust nonlinear regression which has been applied in many fields of astronomy (Liu et al. 2012, 2015). SVR has been particularly widely used in spectral data analysis (Li et al. 2014; Liu et al.

2014; Bu & Pan 2015). This method has been proved to have good performance in determining stellar atmospheric parameters from spectra (Zhang et al. 2020; Guo et al. 2021; Li et al. 2021). In this work, we also adopt SLAM to derive the atmospheric parameters of the M giants.

3.2.1. Stellar Label Model Training

In SLAM, the radial basis function (RBF) is adopted as the kernel of SVR. The hyperparameters C , ε and γ of SLAM represent the penalty level, tube radius and the width of the RBF kernel, respectively. These three hyperparameters can be

automatically determined for each pixel through the training set.

$\vec{\theta}_i$ is denoted as the stellar label vector of the i th star in the training set. $f_j(\vec{\theta}_i)$ and $f_{i,j}$ are defined as the j th pixel of the training spectrum and model output spectrum corresponding to the stars with stellar label vector $\vec{\theta}_i$. The mean squared error (MSE) and median deviation (MD) of the j th pixel can be evaluated with a specific set of hyperparameters, described by Equations (2) and (3)

$$\text{MSE}_j = \frac{1}{m} \sum_{i=1}^m [f_j(\vec{\theta}_i) - f_{i,j}]^2 \quad (2)$$

$$\text{MD}_j = \frac{1}{m} \sum_{i=1}^m [f_j(\vec{\theta}_i) - f_{i,j}] \quad (3)$$

Theoretically, the smaller MSE and MD are, the better fitting is. However, we probably obtain an overfitted model if we train the SLAM model by whole training set, i.e., the MSE_j and MD_j are all equal to 0. Zhang et al. (2020) used the k -fold cross-validated MSE (CV MSE) and k -fold cross-validated MD (CV MD) to measure MSE_j and MD_j to avoid obtaining an overfitted trained model. Namely, the training set is randomly divided into k subsets, where k is set to be 10 in this work. The $f_j(\vec{\theta}_i)$ is predicted by the model which trained by the other $k - 1$ subsets of the training set. After looping through all the hyperparameters sets specified, the best set of hyperparameters can be determined for the j th pixel by searching for the lowest CV MSE_j . The best model can be obtained for each pixel by doing pixel-to-pixel. In this work, we train the SLAM model with 3670 low-resolution spectra from LAMOST of the training sample and their corresponding stellar labels from APOGEE.

3.2.2. Prediction of Stellar Labels

Using the Bayesian formula, the posterior probability density function of stellar label vector for a given observed spectrum is displayed in Equation (4).

$$p(\vec{\theta}|\vec{f}_{\text{obs}}) \propto p(\vec{\theta}) \prod_{j=1}^n p(f_{j,\text{obs}}|\vec{\theta}) \quad (4)$$

where $\vec{\theta}$ is the stellar label vector, \vec{f}_{obs} and $f_{j,\text{obs}}$ represent the normalized observed spectrum vector and the normalized flux of the j th pixel of the observed spectrum, respectively. $p(\vec{\theta})$ is the prior of stellar label vector $\vec{\theta}$, $p(f_{j,\text{obs}}|\vec{\theta})$ is the likelihood of observed spectrum flux of the j th pixel with a given stellar label vector $\vec{\theta}$. By maximizing the posterior probability $p(\vec{\theta}|\vec{f}_{\text{obs}})$, the stellar labels can be easily measured with a Gaussian likelihood adopted. Then the logarithmic form of likelihood is described

by the following equation,

$$\begin{aligned} \ln p(\vec{\theta}|\vec{f}_{\text{obs}}) = & -\frac{1}{2} \sum_{j=1}^n \times \frac{[f_{j,\text{obs}} - f_j(\vec{\theta})]^2}{\sigma_{j,\text{obs}}^2 + \sigma_j(\vec{\theta})^2} \\ & - \frac{1}{2} \sum_{j=1}^n \times \ln [2\pi(\sigma_{j,\text{obs}}^2 + \sigma_j(\vec{\theta})^2)] \end{aligned} \quad (5)$$

where $f_j(\vec{\theta})$ and $\sigma_j(\vec{\theta})$ are the model output spectrum and the uncertainty of the j th pixel corresponding to stellar label $\vec{\theta}$, respectively. $f_{j,\text{obs}}$ is the j th pixel of the normalized observed spectrum, and $\sigma_{j,\text{obs}}$ is the uncertainty of the j th pixel of normalized observed spectrum.

The trained model as described in Section 3.2.1 is applied on the spectra of all the M giant stars to obtain their atmospheric parameters. The right panel of Figure 2 illustrates the distribution of the effective temperature T_{eff} versus surface gravity $\log g$ of the prediction M giant stars with $\text{SNR} > 50$, accounting for 80% of the whole M giant stars. The color is coded by the metallicity $[M/H]$. It exhibits a similar pattern with that of the training sample as showed in the left panel of Figure 2. The same isochrones are also represented by the dashed lines with those in the left panel of Figure 2.

3.2.3. Stellar Label Uncertainty

Similar to the CV MSE and CV MD of spectrum as described in Section 3.2.1, Equations (6) and (7) describe the cross-validate scatter (CV_scatter) and cross-validate bias (CV_bias) of stellar labels, respectively. Obviously, if the predicted stars have the known stellar label, the CV_scatter and CV_bias can be measured. They can be regarded as the standard deviation and average deviation of the stellar labels, respectively, to describe the precision of the stellar parameters determined from the SLAM model. In principle, the smaller CV_bias and CV_scatter indicate a better trained model.

$$\text{CV_bias} = \frac{1}{m} \sum_{i=1}^m (\vec{\theta}_{i,\text{SLAM}} - \vec{\theta}_i) \quad (6)$$

$$\text{CV_scatter} = \frac{1}{m} \sqrt{\sum_{i=1}^m (\vec{\theta}_{i,\text{SLAM}} - \vec{\theta}_i)^2} \quad (7)$$

where $\vec{\theta}_{i,\text{SLAM}}$ is the stellar label vector from the model prediction, and $\vec{\theta}_i$ is the corresponding true stellar label vector.

Figure 4 shows the distributions of CV bias (blue) and scatter (red) of the four parameters versus the SNR of spectra, i.e., the metallicity $[M/H]$, the effective temperature T_{eff} , the surface gravity $\log g$ and the alpha abundance $[\alpha/M]$. It is obvious that the CV scatters of these four parameters decrease with increasing SNR, e.g., from 0.27 dex, 110 K, 0.39 dex, 0.12 dex at $\text{SNR} = 17$ to 0.16 dex, 57 K, 0.25 dex and 0.06 dex at $\text{SNR}=100$ for $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$, respectively. In other words, the precision of these four parameters determined from the SLAM model can reach to 0.16 dex, 57 K, 0.25 dex

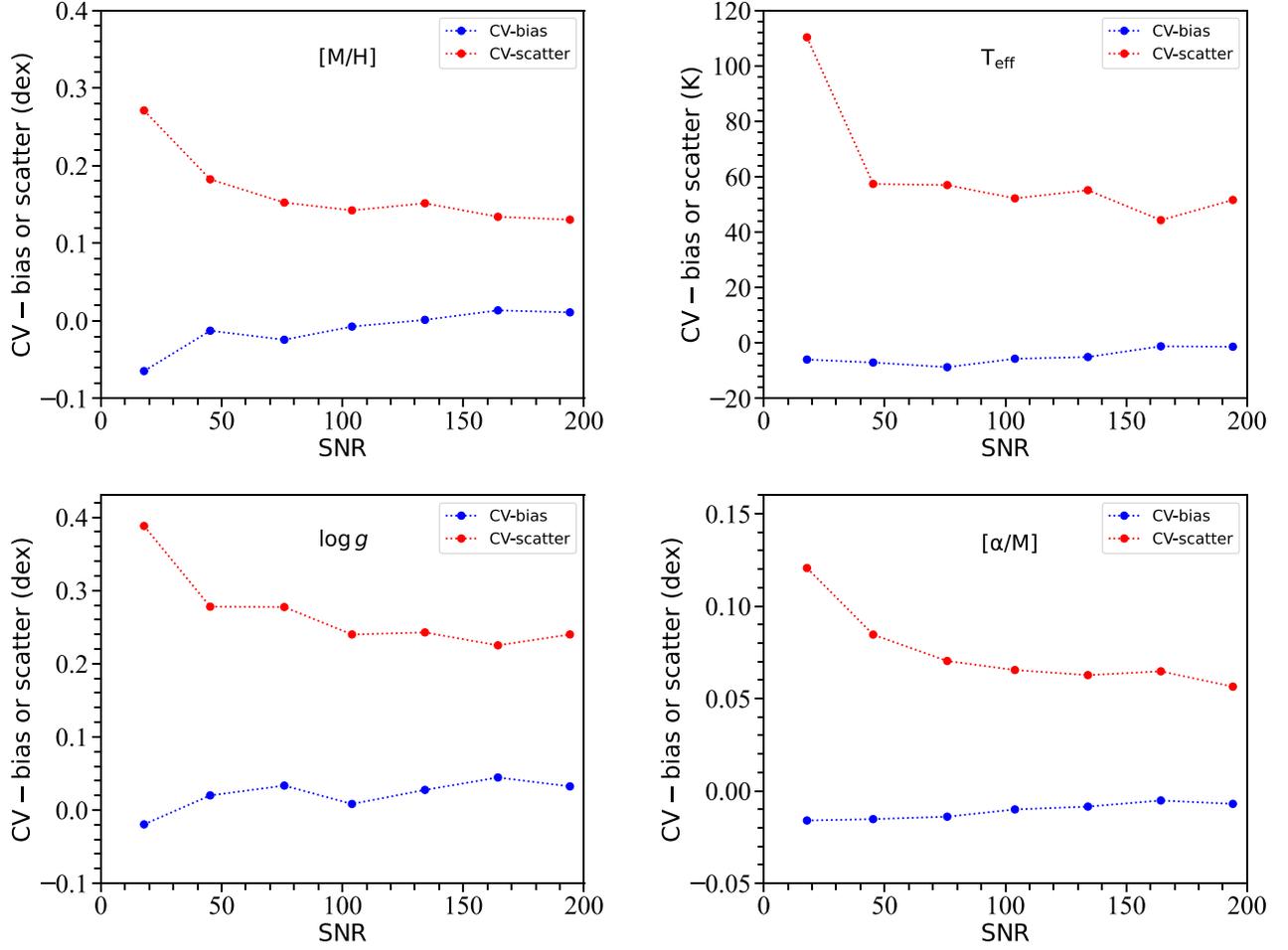


Figure 4. Four panels exhibit the CV scatter and bias as a function of SNR for $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$, respectively. The red and blue dotted lines indicate the CV scatter and CV bias in different SNR bins, respectively.

and 0.06 dex at $\text{SNR} > 100$, respectively. The mean values of CV bias are -0.01 dex, -5 K, 0.02 dex and -0.01 dex for these four parameters, respectively, which means that the predicted parameters by SLAM are in good agreement with the true stellar labels given by APOGEE.

3.2.4. Stellar Labels Self-consistent

In order to verify the self consistency of stellar labels determined from the SLAM model. The training set is randomly split into two subsets: a training set consists of 2770 M giants, the remaining 900 M giants are used as the test set. Besides, the 803 M giant stars mentioned in Section 2.2 are also used as the test set. Figure 5 displayed the comparison of four stellar labels between the true values and the model prediction values. In the left four panels, the black and red dots display the X_{AP} versus X_{SLAM} of 900 stars with $\text{SNR} > 50$ and 803 stars with $\text{SNR} < 50$ in the test set, respectively, where X can be $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$, respectively. It is

obviously that the model prediction stellar labels are agreement with the true stellar labels. However, the distribution between X_{AP} and X_{SLAM} of stars with low SNR (red dots) is more dispersed than that of stars with high SNR (black dots). As shown in the four right panels, the red and black histograms exhibit the distribution of $\Delta X = X_{\text{AP}} - X_{\text{SLAM}}$ of stars with $\text{SNR} > 50$ and stars with $\text{SNR} < 50$. The mean values of ΔX of two subsamples are close to 0, but the scatter values of stars with low SNR is larger than that of stars with higher SNR. This is because the spectra with low signal-to-noise ratio are more difficult to obtain high-precision stellar parameters.

3.2.5. Stellar Labels Validation

We take a comparison of four stellar parameters with Li et al. (2022). They designed a deep convolution neural network with training stellar labels from APOGEE DR17 to derived T_{eff} , $\log g$ and other 12 chemical abundances of 1,210,145 LAMOST DR8 giants with low resolution spectra (R

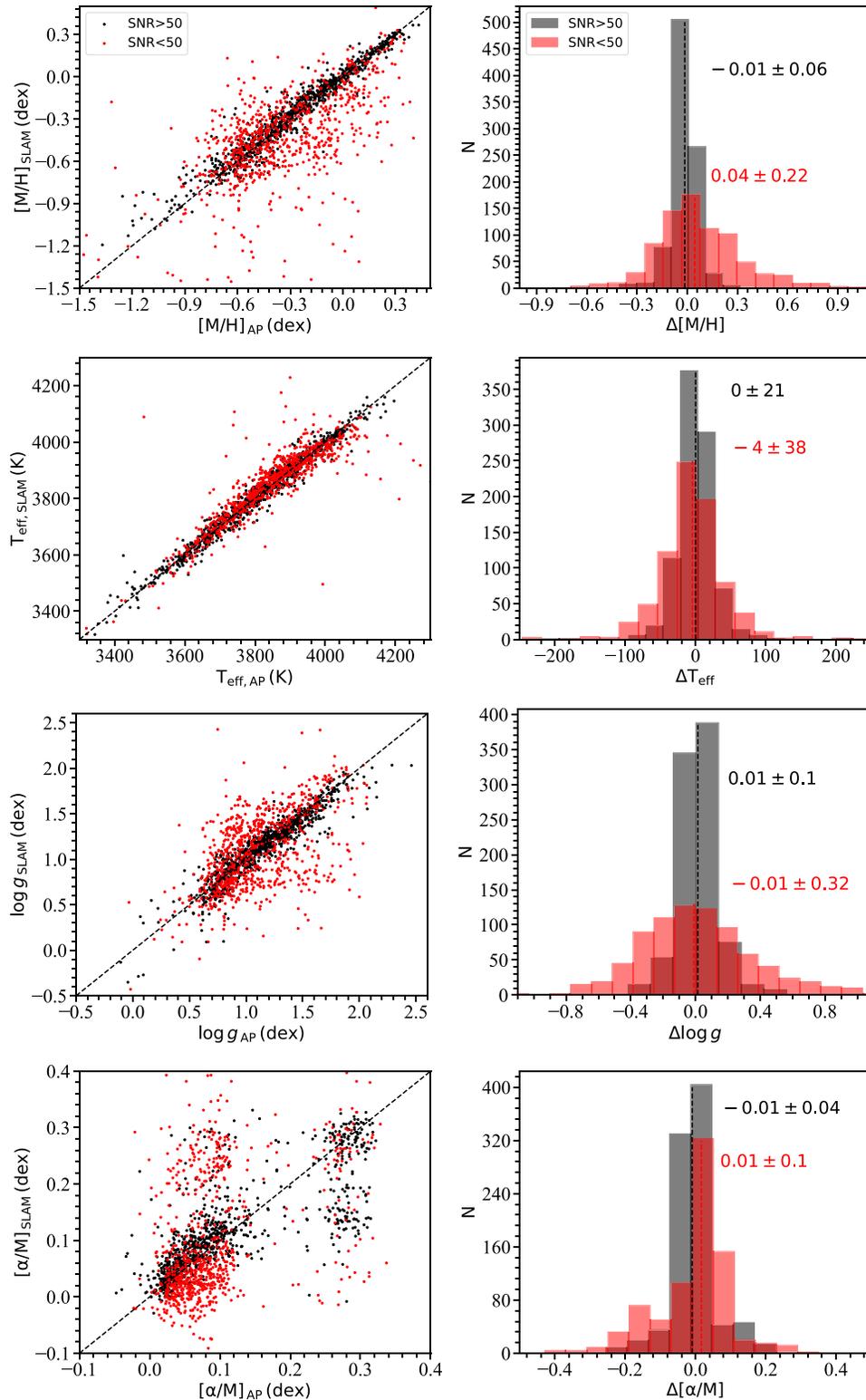


Figure 5. Four panels on the left display the comparison of stellar labels between the true values X_{AP} and model prediction values X_{SLAM} , where X can be $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$. The black and red dots represent the stars with $\text{SNR} > 50$ and $\text{SNR} < 50$, respectively. The black dashed line is the one to one line in these four panels. The histograms of $\Delta X (= X_{AP} - X_{SLAM})$ as exhibited in the right four panels. The black and red histograms show the ΔX of stars with $\text{SNR} > 50$ and stars with $\text{SNR} < 50$, respectively. The mean values of ΔX of two subsamples are marked by the black and red dashed lines.

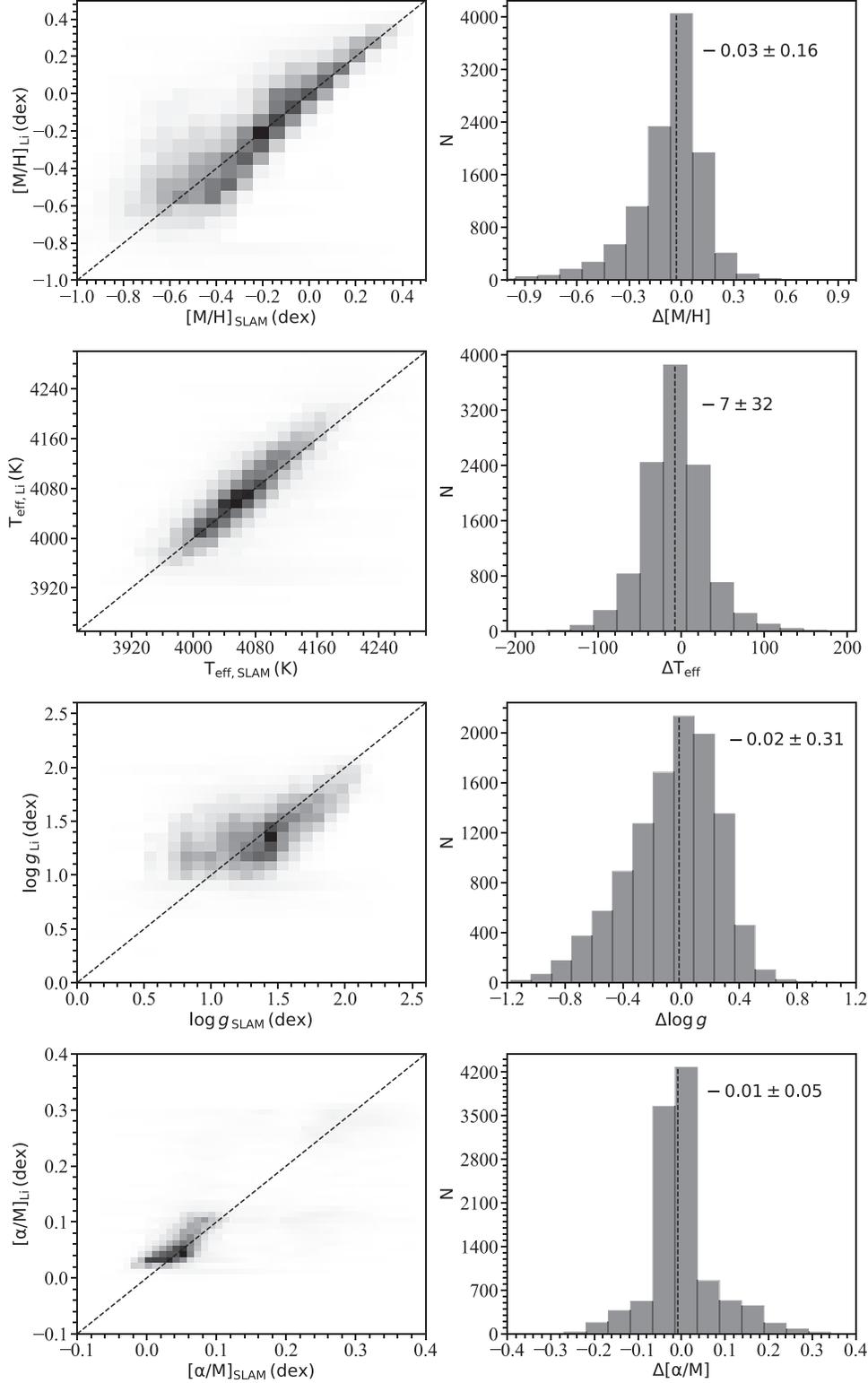


Figure 6. The comparison of $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$ of 11,132 common stars in this work and Li et al. (2022). The left four panels from top to the bottom display the distributions of X_{SLAM} and X_{Li} , where X can be $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$, respectively. The subscript “SLAM” illustrates the parameters from this work and subscript “Li” represents that from Li et al. (2022). The black dotted lines in the left four panels are the one-to-one line. The corresponding histograms of the parameter difference $\Delta X = X_{\text{SLAM}} - X_{\text{Li}}$ as shown in the right four panels. The mean values of ΔX are marked by the black dotted lines in the right panels.

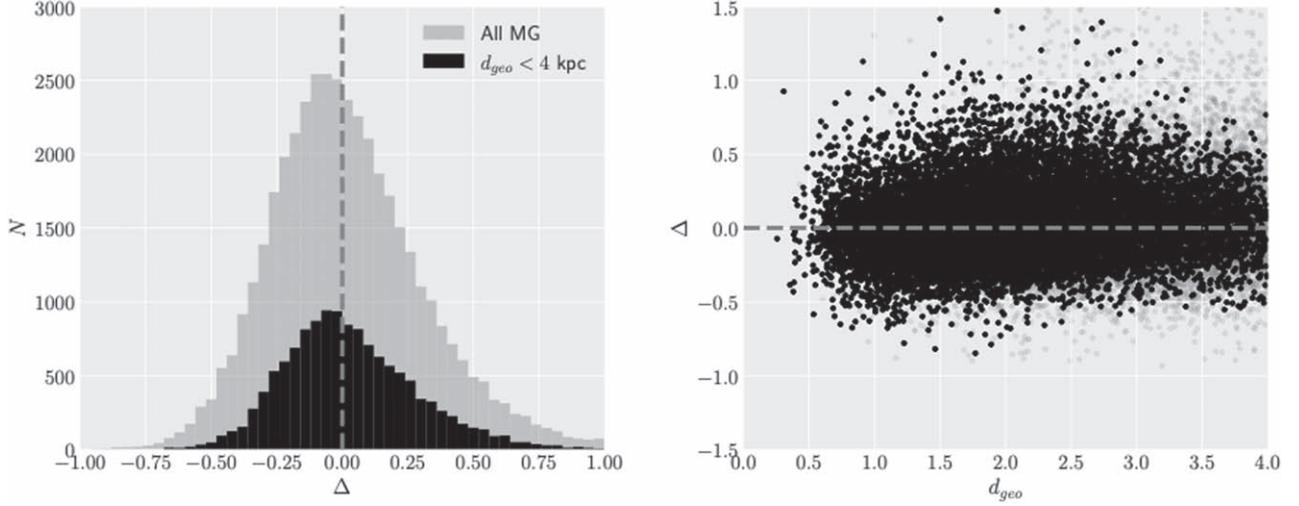


Figure 7. The distance comparison with r_{geo} from Bailer-Jones et al. (2021) is shown. The histogram distribution of the relative difference $\Delta = (d_{\text{MCM}} - d_{\text{geo}})/d_{\text{geo}}$ is shown in the left panel. The distribution of relative difference vs. the distance is shown in the right panel. In both panels, the gray symbols represent the distribution of all the samples with predicted metallicity between -0.9 and 0.5 . While the black ones represent the samples with distance $d_{\text{geo}} < 4$ kpc and its signal to noise ratio larger than 20.

~ 1800). 11,132 common stars were obtained in our samples and Li et al. (2022). The comparison of $[M/H]$, T_{eff} , $\log g$ and $[\alpha/M]$ are displayed in the left four panels in Figure 6. It obviously shows that the parameters in this work are consistent with those in Li et al. (2022). The corresponding histograms of $\Delta X = X_{\text{SLAM}} - X_{\text{Li}}$ are illustrated in the right four panels of Figure 6. The systematic offset between $T_{\text{eff,SLAM}}$ and $T_{\text{eff,Li}}$ is 7 K with a scatter of 32 K. For $[M/H]$, $\log g$ and $[\alpha/M]$, there are very small biases (0.01 0.03 dex) between this work and Li et al. (2022) with scatters of 0.16 dex, 0.31 dex and 0.05 dex, respectively. It demonstrates that these four parameters in this work are consistent with those of Li et al. (2022).

3.3. Distance

In the selected M giant stars, most of them are the red giant branch stars and a few asymptotic giant branch stars. In order to calculate the distance, we first constrain the absolute magnitude for each M giant star. Here we assume the absolute magnitude K_{abs} is a function D of the intrinsic color index c_0 and metallicity M , i.e., $K_{\text{abs}} = D(c_0, M)$. As shown by Majewski et al. (2003), Li et al. (2016) and Figure 2, the M giant stars are mainly tracing the metal-rich components, such as the disk or the Sagittarius Stream, the J and K - band magnitudes are adopted to determine the distance to reduce the extinction effect from dust.

Assuming that the i th star has the apparent magnitudes J_i and K_i and the distance d_i , and suffers the extinction $A_i^K(d_i)$, where $A^K(d_i)$ is a function of distance d_i , then we have the following relation where both sides are the absolute magnitude K_{abs} ,

$$K_i - A_i^K(d_i) - 5 \times \log(100 \times d_i) = D(c_0, M) \quad (8)$$

where c_0 is the intrinsic color index, which can be calculated by

$$c_0 = (J - K) - (A^J(d) - A^K(d)) \quad (9)$$

Then we can find that only the true distance d satisfies Equations (8) and (9). Here the 3D dust map from Green et al. (2019) is adopted, which has a median uncertainty around 30%.

Following Tian et al. (2020), we show the distributions of the comparison of the predicted distance d_{MCM} with that obtained from parallax (Bailer-Jones et al. 2021), d_{geo} . The subsample with reliable distances are selected with $d_{\text{geo}} < 4$ kpc and the $d_{\text{geo}}/\sigma_{d_{\text{geo}}} > 20$. Meanwhile the metallicity is also limited to be between -0.9 and 0.5 . Then the distributions of the relative distance difference $\Delta = (d_{\text{MCM}} - d_{\text{geo}})/d_{\text{geo}}$ of the subsample are shown in Figure 7. The comparison shows a median difference 0.2% and the 16% and 84% percentage values of 21.2% and 27.5%, which indicate a very small system offset and statistical dispersion smaller than 30%. In the right panel, the relative distance difference is represented versus the distance d_{geo} . There is not significant relation between the system offset versus the distance.

4. Discussion

With the distances and radial velocity combining the astrometric information from Gaia, we are able to calculate the full 6D information, positions (X, Y, Z) and velocities (U, V, W) . Figure 8 shows the space distributions of the M giant stars. The location of the Sun is represented by the dashed lines. We can find that most of the M giant stars are located in the disk, $|Z| < 5$ kpc, meanwhile, there are also few M giant stars of larger heights, which are possible the member stars of the Sagittarius Stream.

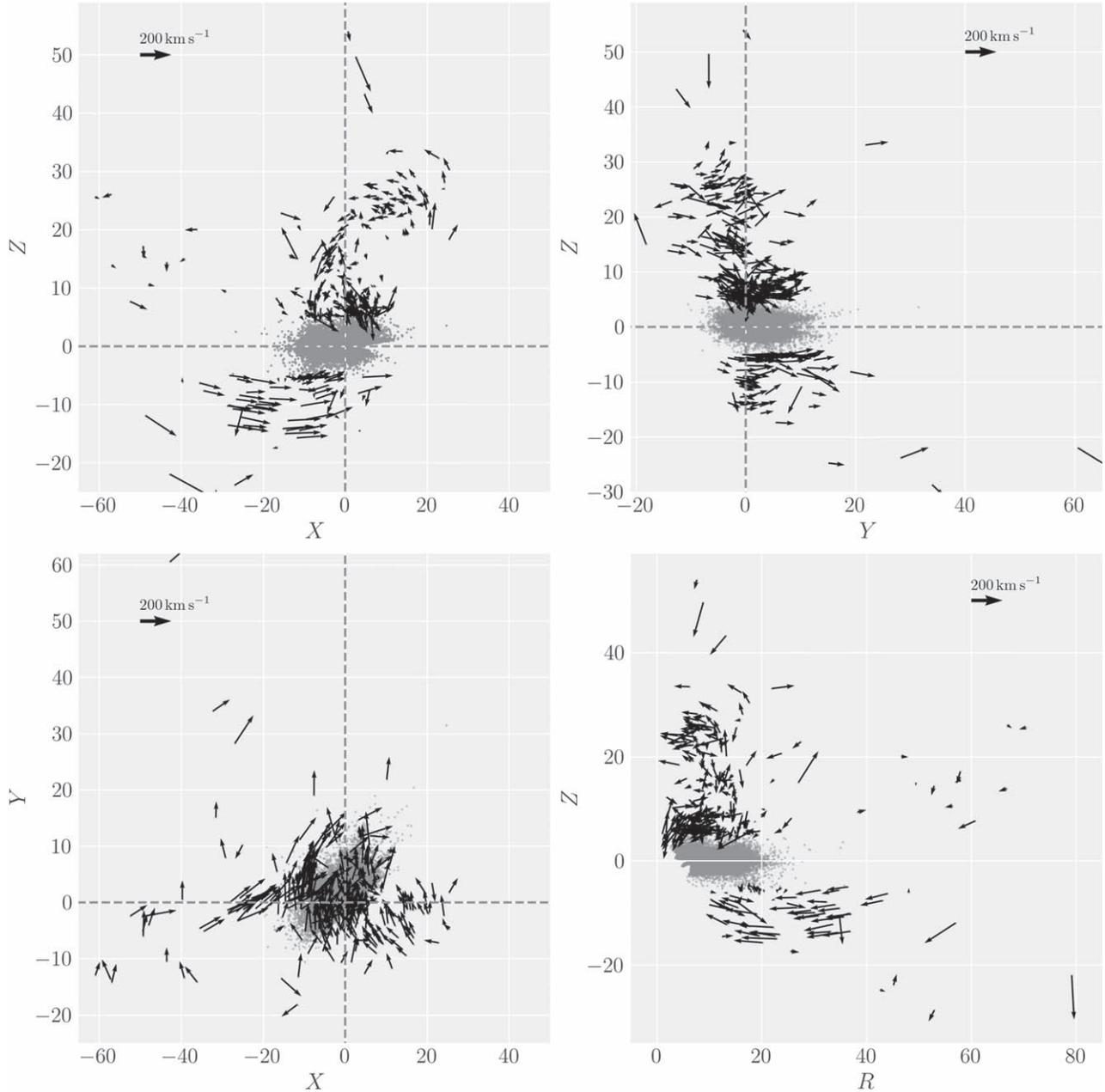


Figure 8. The space distributions are shown in different spaces. The velocities of the samples with larger height $|Z| > 5$ kpc in the corresponding space are represented by arrows. The Sun is located at $(X, Y, Z) = (0, 0, 0)$ kpc and $(R, Z) = (8.34, 0)$ kpc.

4.1. Disks

As shown in Figure 8, majority of our sample are the disk stars, which are located with height smaller than 5 kpc. Figure 9 shows the movements of disk traced by the M giant stars with different heights. The five known spiral arms are represented by the dashed curves from Chen et al. (2019). The distributions of the line integral convolution show clear streaming movements which are the rotation of the disk,

especially the thin disk represented by the subsample with $|Z| < 0.5$ kpc.

Figure 10 shows the phase space distribution of the M giant stars, J_ϕ versus J_{tot} . The samples are color-coded by the height Z to the disk plane and the action J_Z in the left and right panels, respectively. The action distributions also prove that the majority of the M giant stars are concentrated in the thin cyan and blue belt in the left and right panels, respectively, which are the disk with low height $|Z|$ and action J_Z .

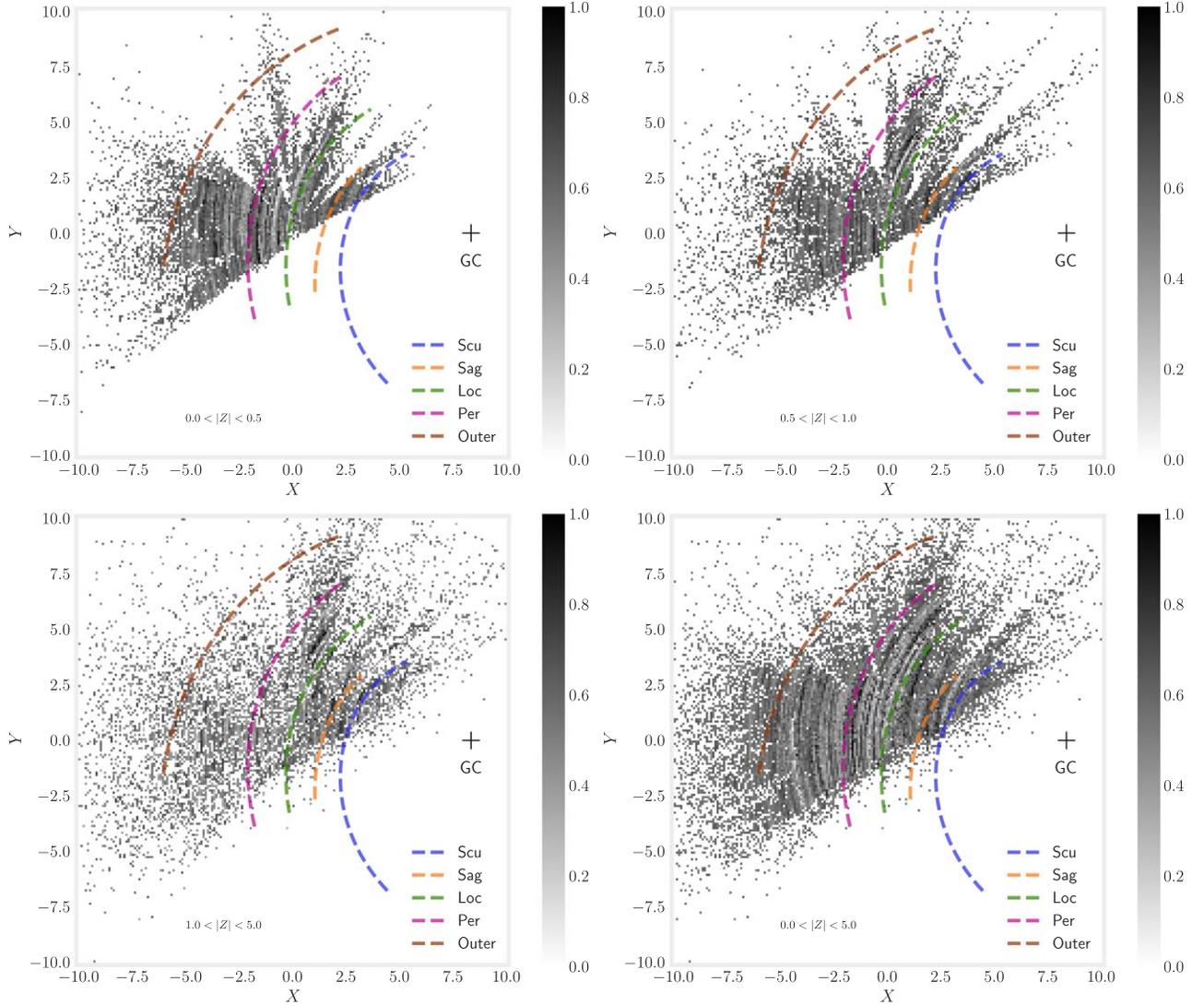


Figure 9. Disk rotational velocity distribution after line integral convolution with different heights.

4.2. Sagittarius Stream

Another application for the M giant sample is to trace the Sagittarius Stream (Majewski et al. 2003; Li et al. 2016). From Figures 8 and 10, there are also a few stars with larger heights $|Z| > 5$ kpc, whose movements are also represented by arrows. From previous studies (Majewski et al. 2003; Li et al. 2016), the Sagittarius Stream has a significant contribution to the M giant stars with high latitude. In Figure 8, those M giant stars with larger heights show significant bulk motions, especially those stars with Z around 25 kpc and -10 kpc. That is more clear in Figure 10, where those samples are of larger total actions J_{tot} and small angular momentum J_{ϕ} , which are completely different with those belong to the disk. To further illustration, we convert the coordinate (α, δ) to that based on

the Sagittarius Stream plane (Λ, B) , where the north pole $B = 90^\circ$ is set to the same direction of the normal of the orbit plane of the Sagittarius Stream. Then Figure 11 shows the distance variance versus the longitude Λ color-coded by the latitude $|B|$. The mock data for the Sagittarius Stream from Law & Majewski (2010) are also shown with gray dots. We find that the distance distribution of those M giant stars fits the model well. There are also few nearby stars offset the model which are possible the flared disk stars with $|Z| > 5$ kpc. That can be proved by the action distributions as shown in Figure 10.

5. Summary

With similar method with Zhong et al. (2015, 2019), Li et al. (2023) select M giant stars from LAMOST DR9 with a very

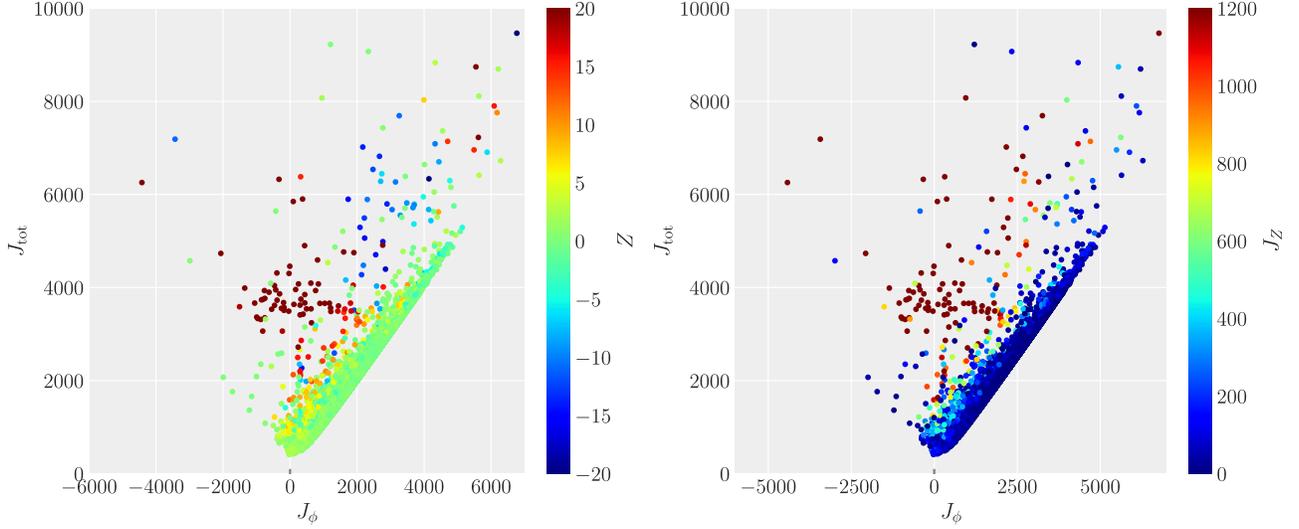


Figure 10. The distribution of the M giant stars in the action space J_ϕ vs. J_{tot} are showed, which are color-coded by the height Z and action J_z in the left and right panels, respectively.

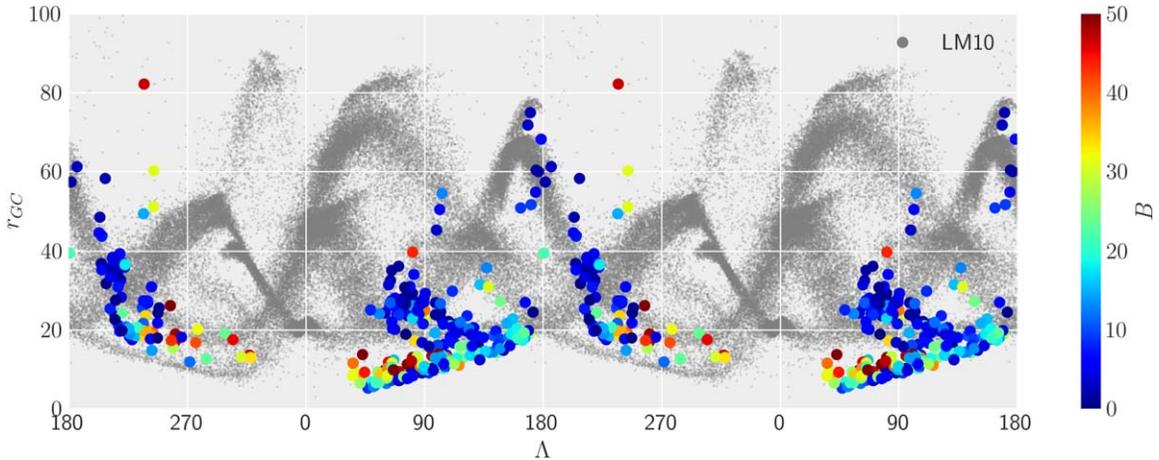


Figure 11. The distance distribution vs. longitude Λ of the M giant stars with height larger than 5 kpc is shown where the color is coded by the latitude B . The samples of the simulated Sagittarius Stream from LM10 are showed with gray symbols.

high purity. Many of those low temperature M giant stars are not given the stellar parameters and the radial velocities in the official catalog. In this work, we revise the spectra of those M giant stars and constrain the stellar parameters (T_{eff} , $\log g$, $[M/H]$), the chemical abundance $[\alpha/M]$ and the radial velocities with uncertainties of (57 K, 0.25 dex, 0.16 dex, 0.06 dex) and 4.6 km s^{-1} , respectively. With the information, we are able to calculate the full 6D information of M giants, and further to study the Milky Way disk and the Sagittarius Stream. Combining the geometric and phase space distributions, the disk and the Sagittarius Stream can be well separated. This value added sample will provide a pure sample for the chemical and kinematic studies for the disk and the Sagittarius Stream.

Acknowledgments

H.T. is supported by Beijing Natural Science Foundation with Grant No. 1214028 and the National Natural Science Foundation of China (NSFC) under grant 12103062. J.L. would like to acknowledge the NSFC under grant 12273027 and China West Normal University grant 17YC507. C.L. thanks the NSFC with grant No.11835057 and the National Key R&D Program of China No. 2019YFA0405501. J.R.S. is supported by the National Key R&D Program of China No. 2019YFA0405502 and the NSFC under Grant Nos. 12090040, 12090044, and 11833006. M.Y. gratefully acknowledges support from the NSFC (Grant No.12133002). 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 the National Astronomical Observatories, Chinese Academy of Sciences.

Appendix Catalog Sample

Table A1 lists the information in our catalog, including the observational ID (obsid) from LAMOST, the source id from Gaia (gaia_source_id), the coordinates (ra, dec) from LAMOST, the mean signal to noise ratio (SNR) of LAMOST spectra, the stellar parameters, the radial velocities and their uncertainties, the distance of M giants are also provided. The whole catalog is available to download through the link <https://nadc.china-vo.org/res/r101196/>.

Table A1
Catalog Description of $\sim 43,000$ M Giants

Column	Units	Description
obsid		LAMOST observe id
gaia_source_id		Gaia source id
ra	deg	R.A. from LAMOST
dec	deg	decl. from LAMOST
SNR		mean signal to noise ratio of LAMOST spectra
[M/H]	dex	[M/H] from SLAM
[M/H] _{err}	dex	[M/H] uncertainty
T_{eff}	K	T_{eff} from SLAM
$T_{\text{eff err}}$	K	T_{eff} uncertainty
log g	dex	log g from SLAM
log g_{err}	dex	log g uncertainty
$[\alpha/M]$	dex	$[\alpha/M]$ from SLAM
$[\alpha/M]_{\text{err}}$	dex	$[\alpha/M]$ uncertainty
rv	km s ⁻¹	radial velocity
rv_{err}	km s ⁻¹	radial velocity uncertainty
Distance	kpc	Distance of M giants

ORCID iDs

Hao Tian  <https://orcid.org/0000-0003-3347-7596>

Bo Zhang  <https://orcid.org/0000-0002-6434-7201>

References

- Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, 259, 35
- Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, *ApJS*, 249, 3
- Allende Prieto, C., Koesterke, L., Hubeny, I., et al. 2018, *A&A*, 618, A25
- Antoniadis-Karnavas, A., Sousa, S. G., Delgado-Mena, E., et al. 2020, *A&A*, 636, A9
- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, *AJ*, 161, 147
- Belokurov, V., Koposov, S. E., Evans, N. W., et al. 2014, *MNRAS*, 437, 116
- Bizyaev, D., Smith, V. V., Arenas, J., et al. 2006, *AJ*, 131, 1784
- Bizyaev, D., Smith, V. V., & Cunha, K. 2010, *AJ*, 140, 1911
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Bu, Y., & Pan, J. 2015, *MNRAS*, 447, 256
- Carlin, J. L., Sheffield, A. A., Cunha, K., & Smith, V. V. 2018, *ApJL*, 859, L10
- Castelli, F., & Kurucz, R. L. 2003, in *Modelling of Stellar Atmospheres*, ed. N. Piskunov, W. W. Weiss, & D. F. Gray, Vol. 210 (IAU: Astronomical Society of the Pacific), A20
- Chen, B. Q., Huang, Y., Hou, L. G., et al. 2019, *MNRAS*, 487, 1400
- Chene, A.-N., Padzer, J., Barrick, G., et al. 2014, *Proc. SPIE*, 9151, 915147
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA*, 12, 1197
- Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA*, 12, 735
- Ding, M.-Y., Shi, J.-R., Wu, Y., et al. 2022, *ApJS*, 260, 45
- Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, *ApJS*, 178, 89
- Gaia Collaboration 2022, *yCat*, I/355
- Galgano, B., Stassun, K., & Rojas-Ayala, B. 2020, *AJ*, 159, 193
- Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, *ApJ*, 887, 93
- Guo, Y., Zhang, B., Liu, C., et al. 2021, *ApJS*, 257, 54
- Howard, E. M. 2017, in *ASP Conf. Ser. 512, Astronomical Data Analysis Software and Systems XXV*, ed. N. P. F. Lorente, K. Shorridge, & R. Wayth (San Francisco, CA: ASP), 245
- Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, *Natur*, 370, 194
- Katz, D., Sartoretti, P., Guerrier, A., et al. 2022, arXiv:2206.05902
- Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, *A&A*, 501, 1269
- Koposov, S. E., Belokurov, V., Zucker, D. B., et al. 2015, *MNRAS*, 446, 3110
- Kurucz, R. 1993, *ATLAS9 Stellar Atmosphere Programs and 2 km/s Grid*. Kurucz CD-ROM No. 13 (Cambridge: SAO), 13
- Law, D. R., & Majewski, S. R. 2010, *ApJ*, 714, 229
- Li, J., FELLOW, L., Liu, C., et al. 2019, *ApJ*, 874, 138
- Li, J., Liu, C., Carlin, J. L., et al. 2016b, *RAA*, 16, 125
- Li, J., Liu, C., Zhang, B., et al. 2021, *ApJS*, 253, 45
- Li, J., Smith, M. C., Zhong, J., et al. 2016, *ApJ*, 823, 59
- Li, J., Long, L., Zhong, J., et al. 2023, arXiv:2304.00308
- Li, X., Wu, Q. M. J., Luo, A., et al. 2014, *ApJ*, 790, 105
- Li, Z., Zhao, G., Chen, Y., Liang, X., & Zhao, J. 2022, *MNRAS*, 517, 4875
- Liu, C., Bailer-Jones, C. A. L., Sordo, R., et al. 2012, *MNRAS*, 426, 2463
- Liu, C., Fang, M., Wu, Y., et al. 2015, *ApJ*, 807, 4
- Liu, C., Fu, J., Shi, J., et al. 2020, arXiv:2005.07210
- Liu, C., Xu, Y., Wan, J.-C., et al. 2017, *RAA*, 17, 906
- Liu, C.-X., Zhang, P.-A., & Lu, Y. 2014, *RAA*, 14, 423
- Luo, A. L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, *RAA*, 12, 1243
- Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, *RAA*, 15, 1095
- Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, *ApJ*, 599, 1082
- Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, *ApJ*, 569, 245
- Su, D.-Q., & Cui, X.-Q. 2004, *ChJAA*, 4, 1
- Tian, H., Liu, C., Wang, Y., et al. 2020, *ApJ*, 899, 110
- Tian, H., Liu, C., Xu, Y., & Xue, X. 2019, *ApJ*, 871, 184
- Tollestrup, E. V., Pazder, J., Barrick, G., et al. 2012, *Proc. SPIE*, 8446, 84462A
- Wang, R., Luo, A. L., Chen, J.-J., et al. 2020, *ApJ*, 891, 23
- Wang, S.-G., Su, D.-Q., Chu, Y.-Q., Cui, X., & Wang, Y.-N. 1996, *ApOpt*, 35, 5155
- Wu, Y., Du, B., Luo, A., Zhao, Y., & Yuan, H. 2014, in *Statistical Challenges in 21st Century Cosmology*, ed. A. Heavens, J.-L. Starck, & A. Krone-Martins, Vol. 306 (Cambridge: Cambridge Univ. Press), 340
- Xu, Y., Liu, C., Tian, H., et al. 2020, *ApJ*, 905, 6
- Xu, Y., Liu, C., Xue, X.-X., et al. 2018, *MNRAS*, 473, 1244
- Yan, H., Li, H., Wang, S., et al. 2022, *The Innovation*, 3, 100224
- Zhang, B., Li, J., Yang, F., et al. 2021, *ApJS*, 256, 14
- Zhang, B., Liu, C., & Deng, L.-C. 2020, *ApJS*, 246, 9
- Zhao, G., Zhao, Y., Chu, Y., Jing, Y., & Deng, L. 2012, arXiv:1206.3569
- Zhong, J., Lépine, S., Li, J., et al. 2015, *RAA*, 15, 1154
- Zhong, J., Li, J., Carlin, J. L., et al. 2019, *ApJS*, 244, 8