



Data-driven Spectroscopic Estimates of Absolute Magnitude, Distance, and Binarity: Method and Catalog of 16,002 O- and B-type Stars from LAMOST

Maosheng Xiang^{1,9}, Hans-Walter Rix¹, Yuan-Sen Ting^{2,3,4,5,10}, Eleonora Zari¹, Kareem El-Badry^{1,6}, Hai-Bo Yuan⁷, and Wen-Yuan Cui⁸

¹ Max-Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany; mxiang@mpia.de

² Institute for Advanced Study, Princeton, NJ 08540, USA

³ Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

⁴ Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA

⁵ Research School of Astronomy & Astrophysics, Australian National University, Canberra, ACT 2611, Australia

⁶ Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA

⁷ Department of Astronomy, Beijing Normal University, Beijing 100875, People's Republic of China

⁸ Department of Physics, Hebei Normal University, Shijiazhuang 050024, People's Republic of China

Received 2020 August 23; revised 2020 December 9; accepted 2020 December 11; published 2021 March 10

Abstract

We present a data-driven method to estimate absolute magnitudes for O- and B-type stars from the LAMOST spectra, which we combine with Gaia DR2 parallaxes to infer distance and binarity. The method applies a neural network model trained on stars with precise Gaia parallax to the spectra and predicts K_s -band absolute magnitudes M_{K_s} with a precision of 0.25 mag, which corresponds to a precision of 12% in spectroscopic distance. For distant stars (e.g., >5 kpc), the inclusion of constraints from spectroscopic M_{K_s} significantly improves the distance estimates compared to inferences from Gaia parallax alone. Our method accommodates for emission-line stars by first identifying them via principal component analysis reconstructions and then treating them separately for the M_{K_s} estimation. We also take into account unresolved binary/multiple stars, which we identify through deviations in the spectroscopic M_{K_s} from the geometric M_{K_s} inferred from Gaia parallax. This method of binary identification is particularly efficient for unresolved binaries with near equal-mass components and thus provides a useful supplementary way to identify unresolved binary or multiple-star systems. We present a catalog of spectroscopic M_{K_s} , extinction, distance, flags for emission lines, and binary classification for 16,002 OB stars from LAMOST DR5. As an illustration, we investigate the M_{K_s} of the enigmatic LB-1 system, which Liu et al. had argued consists of a B star and a massive stellar-mass black hole. Our results suggest that LB-1 is a binary system that contains two luminous stars with comparable brightness, and the result is further supported by parallax from the Gaia eDR3.

Unified Astronomy Thesaurus concepts: OB stars (1141); Early-type stars (430); Emission line stars (460); Stellar distance (1595); Binary stars (154); Spectroscopic binary stars (1557); Absolute magnitude (10); Astronomical methods (1043); Astronomy data modeling (1859); Principal component analysis (1944); Neural networks (1933); Catalogs (205)

Supporting material: machine-readable table

1. Introduction

O-type and B-type (OB) stars constitute the population of massive, young, and luminous stars in a galaxy. They play a significant role in many aspects of astrophysics. They are important factories for element production (Thielemann & Arnett 1985; Timmes et al. 1995; Woosley & Weaver 1995; Chieffi & Limongi 2004; Nomoto et al. 2006) and act as major sources of ionization and energetic feedback to the interstellar medium and intergalactic medium (Freyer et al. 2003, 2006; Hopkins et al. 2014; Mackey et al. 2015; Struck 2020). They often form in binaries and are candidate companions for, or precursors of, black holes and associated gravitational wave events (Abbott et al. 2016a, 2016b, 2017; Belczynski et al. 2016; Liu et al. 2019a). In our Galaxy and elsewhere, they

serve as signposts of star formation and diagnostics of the initial mass function (e.g., Lequeux 1979; Humphreys & McElroy 1984; Reed 2005; Bartko et al. 2010). They also serve as tracers of the structure and dynamics of the Galactic disk, including features like spiral arms (e.g., Torra et al. 2000; Shu 2016; Xu et al. 2018; Chen et al. 2019a; Cheng et al. 2019; Li et al. 2019; Wang et al. 2020). Knowledge of the luminosities and distances of the OB stars in the Milky Way is fundamental to all such analyses.

For OB stars close to the Sun, high-precision distance estimates are obtainable using parallaxes from Gaia DR2 (Gaia Collaboration et al. 2016, 2018a; Lindegren et al. 2018), provided that the stars are not binaries. But for distances larger than ~ 1.5 kpc, the parallax-based distance estimates for OB stars might become suboptimal (Shull & Danforth 2019). This calls for developing alternative ways to accurately determine the distances to such stars and detect binarity systematically.

Photometric distance estimation has a long and successful history: for stars on the lower main sequence, e.g., G and K dwarfs, unreddened broadband colors alone serve as excellent luminosity predictors, as long as the metallicity is known to within $\lesssim 0.5$ dex (e.g., Ivezić et al. 2008; Jurić et al. 2008). For

⁹ Corresponding author.

¹⁰ Hubble fellow.



OB stars, on the other hand, the time spent on or near their zero-age main sequence is so short that, at a given color, their luminosities can range over an order of magnitude. Distant young OB stars also tend to be in directions of high dust extinction with ongoing star formation, further limiting the accuracy of the derived dereddened colors. On top of that, there is also a lack of color variation beyond the Rayleigh–Jeans tail of the OB stars at ~ 4000 Å. This combination of factors makes photometric luminosity and distance estimates for hot stars particularly challenging.

The spectra of OB stars contain much more information than photometric colors and can yield powerful constraints on luminosities and distances. There are two conceptually different approaches to infer spectroscopic distances. The first is to derive the basic stellar parameters, T_{eff} , $\log g$ and $[\text{Fe}/\text{H}]$, and then, for example, employ the “flux-weighted gravity luminosity” relationship (FGLR; Kudritzki et al. 2003), which relates the $g_f \equiv g/T_{\text{eff}}^4$ of a star to its bolometric luminosity (Kudritzki et al. 2020). With this relationship, the T_{eff} and $\log g$ derived from the high-resolution spectra of single stars yield luminosities and distances with precisions of $<20\%$ and 10% , respectively, for luminous hot stars. A similar method is from Shull & Danforth (2019), who estimated the spectrophotometric distance of 139 O-type dwarfs using absolute magnitudes inferred by empirical relation with spectral types from the Galactic O-star Spectroscopic Survey (GOSSS; Maíz-Apellániz et al. 2004). Stellar isochrones have also been used to infer stellar absolute magnitudes and distances from stellar parameters T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$, as has been widely implemented on spectroscopic survey data sets (e.g., Carlin et al. 2015; Yuan et al. 2015b; Wang et al. 2016; Xiang et al. 2017a; Coronado et al. 2018a; Queiroz et al. 2020; Green et al. 2021). So far the application has been mostly focused on FGK stars, but it should provide decent distance estimates akin to the FGLR method for hot luminous stars.

The second approach is to learn the luminosity and distance directly from the data, without the detour of first deriving the stellar parameters. This can be done if numerous examples of the spectra of stars with independently known distances exist, which is the case in the age of Gaia and extensive spectroscopic surveys. Jofré et al. (2015) inferred stellar distance with spectroscopically identified twin stars that have accurate parallax measurements, and achieved a distance precision better than 10% from high-resolution spectra. Hogg et al. (2019) demonstrated that the distances of red giant branch stars with SDSS/APOGEE (Majewski et al. 2017; Abolfathi et al. 2018) can be inferred to within $<10\%$ using a data-driven model trained on Gaia parallaxes. Leung & Bovy (2019) also estimated the stellar distance from SDSS/APOGEE spectra but with a “deep-learning” method, which simultaneously generates the zero-point offset of Gaia DR2. From the LAMOST low-resolution spectra, Xiang et al. (2017b) deduced absolute magnitudes for AFGK stars by taking stars in common with the Hipparcos (Perryman et al. 1997) as the training set. Verification with Gaia parallaxes suggests that absolute magnitudes inferred from the LAMOST low-resolution spectra with such a data-driven approach can be precise to 0.26 mag, corresponding to a distance precision of 12% (Xiang et al. 2017a).

For OB stars, such a data-driven approach for spectroscopic luminosity estimation comes with two additional complications. First, a non-negligible fraction of massive stars are in close binaries, often with comparable luminosities (Sana et al. 2012, 2013). Second, the optical spectra of OB stars, such as

those from the LAMOST survey, frequently show emission lines arising either from their corona, their surrounding disks, or nearby H II regions. Eliminating these outliers is crucial for achieving robust spectroscopic luminosity estimates.

In this work, we set up a data-driven method to estimate the spectroscopic luminosity and, by implication, the distance of individual OB stars. We employ a training set built from stars with precise parallaxes from Gaia DR2 in a neural network model to map observed spectra to K_s -band absolute magnitudes M_{K_s} . The method is designed to account for issues stemming from binarity and emission lines. We combine spectroscopic M_{K_s} and Gaia parallaxes to identify binary stars as those that are overluminous compared to the single stars. We apply the method to the LAMOST low-resolution ($R \simeq 1800$) spectra for a set of 16,002 OB stars from Liu et al. (2019b), which is by far the most extensive set of spectra for luminous, hot stars. We prefer this data-driven approach over the FGLR, given the complications of validating the accuracy of the stellar parameters determined for OB stars from low-resolution spectra.

Our validation of the results suggests that the combination of the spectroscopic M_{K_s} with the Gaia DR2 parallax yields a median distance uncertainty of only 8% for the LAMOST OB star sample. These distances are presented together with a catalog of M_{K_s} , extinction, and flags for binary and emission lines for the 16,002 LAMOST OB stars. The results allow a reassessment for the nature of the binary system LB-1, which has been recently suggested to hold a $70M_{\odot}$ black hole, the most massive stellar-mass black hole ever found (Liu et al. 2019a).

This paper is laid out as follows: Section 2 gives an overview of the methods. Section 3 introduces the identification of emission lines in LAMOST OB star spectra with a principal component analysis (PCA) reconstruction method; Section 4 introduces our extinction estimation; and Section 5 presents the data-driven method for M_{K_s} estimation. Section 6 introduces our method for binary identification. Section 7 describes the inference of distance using the spectroscopic M_{K_s} together with Gaia parallaxes. Section 8 discusses our estimates for the absolute magnitude and distance to the LB-1 system. We summarize in Section 9.

2. Method Overview

We derive absolute magnitudes M_{K_s} from survey spectra with a data-driven neural network model trained on a subset of selected stars with precise parallaxes from Gaia DR2 (Section 5). In light of the non-negligible impact from emission lines on our data-driven models, we identify the spectra containing emission lines using a PCA reconstruction method (Section 3). For these stars, we derive M_{K_s} using a separate neural network model, with the emission-line wavelength regions masked. Considering the neural network model is sensitive to spectral noise, we adopt the PCA-reconstructed spectra, for both emission and non-emission stars, as an approximate of the “noiseless” spectra for M_{K_s} estimation. With the spectroscopic M_{K_s} estimates, we infer distances in combination with the apparent magnitudes and Gaia DR2 parallaxes (Section 7). Binary and multiple-star systems are discarded iteratively from the training sets used for M_{K_s} estimation. To identify these systems, which should be overluminous compared to their single-star counterparts, we measure the deviation between the spectrophotometric parallax and the Gaia astrometric parallax (Section 6). To correct for the extinction of individual stars, we use intrinsic colors estimated from

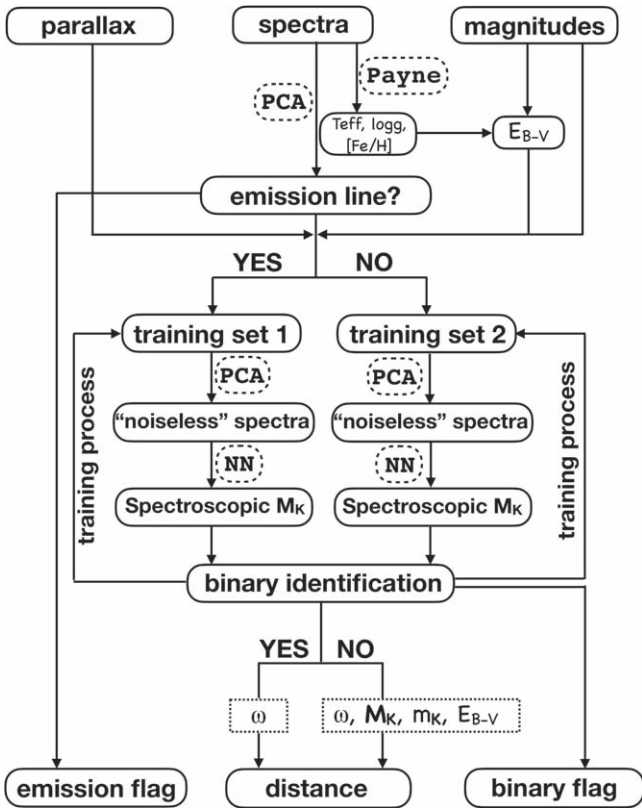


Figure 1. Schematic illustration of our data-driven approach for deriving absolute magnitudes and distances for OB stars, including the identification of stars with emission lines and stars in binary or multiple-star systems. Astrometric parallaxes, spectra, and apparent magnitudes are adopted as inputs.

synthetic photometry (Section 4). A schematic description of our method is shown in Figure 1.

Our approach leverages the information contained in spectra, astrometric parallax, and photometric magnitudes, and can be applied to a wide range of stellar types. For the present study, however, we restrict our analysis to LAMOST spectra of OB stars. We adopt the OB star catalog of Liu et al. (2019b), which contains a total of 16,032 OB stars from the fifth data release (DR5)¹¹ of the LAMOST Galactic surveys (Deng et al. 2012; Zhao et al. 2012; Liu et al. 2014). These OB stars are identified with line indices and further inspected by eye, leading to a high purity of the sample (Liu et al. 2019b). We also adopt the astrometric parallax measurements from Gaia DR2 (Gaia Collaboration et al. 2018a), following Leung & Bovy (2019) to correct for the zero-point offset as a function of G -band magnitude. For the photometric input, we use the 2MASS K_s -band measurements (Skrutskie et al. 2006).

3. Identifying Emission Lines with PCA Reconstruction

The spectra for a considerable fraction of OB stars can exhibit emission lines, either from their associated H II regions or from surrounding gas disks. Because this emission does not necessarily correlate with the properties of the star, or its M_{K_s} , these emission lines can confound our estimation of M_{K_s} . We therefore opt to remove all objects with emission lines in the spectra from our main training set and devise a separate strategy to estimate M_{K_s} for these objects, using only the part of

the spectrum free of emission lines. Almost all of the spectra with emission lines are found to exhibit H_α emission. Some of them also exhibit other hydrogen lines in the Balmer series and Paschen series, as well as some metal lines. Nonetheless, in most cases, the H_α emission line is the most prominent. As such, we adopt H_α as a key diagnostic to identify OB stellar spectra with emission lines.

To automatically identify spectra with emission lines, we adopt a PCA reconstruction method as described below. In a nutshell, we first define a set of clean wavelength windows that suffer minimal impact from emission lines with which we will determine the coefficient of the principal component of the spectrum. We then reconstruct the full spectrum using the coefficients as well as the full spectrum eigenbases derived from a sample of non-emission stars. We identify the stars with emission lines through the H_α difference between the observed spectra and the PCA reconstruction counterparts. This process is implemented iteratively to obtain a sample without emission lines. In practice, we found that only one iteration is sufficient because further iterations will not make a significant change to the results.

Given a set of M spectra x_i (with $i = 1, \dots, M$) each containing N pixels, PCA projects the spectra in spectral space to an eigenspace, where the eigenvalues represent the amount of variance held by each orthonormal eigenbasis. The eigenbases are obtained by diagonalizing the covariance matrix,

$$C = XX^T, \quad (1)$$

$$C\xi = \lambda\xi, \quad (2)$$

where X is the $N \times M$ -dimensional array of spectra and λ is the eigenvalue associated with the eigenbasis (eigenvector) ξ . Entries in X are standardized by subtracting, from each pixel, the mean flux of the training set and then normalized by the standard deviation. The eigenvalues and eigenvectors are computed using the `tried.pro` and `trigl.pro` scripts in IDL.

In practice, we first consider only pixels in the clean windows (as shown in Figure 2) and use those pixels to construct the matrix X of the training spectra. We adopt the matrix X to determine the principal components (eigenvectors/eigenbases) $\{\xi\}$ in this restricted space. For any given training spectrum, x , we calculate the principal component coefficients by projecting x onto each principal component ξ . Because the eigenvectors are by definition normalized, the projection is simply the dot product of the two vectors, $p = x \cdot \xi$. The collection of all principal coefficients p for all training spectra constitute a $M \times K$ matrix, P , where K is the number of principal components, and M is the number of training spectra. Here we choose only the top $K = 100$ principal components in order to denoise and omit irrelevant information in the spectra.

With matrix P in place, we then reconstruct the full spectra as follow. Let \tilde{X} to be the corresponding full spectra matrix for the training spectra, we search for an array B such that $PB = \tilde{X}^T$. In other words, we approximate the eigenbases for the full spectra with which \tilde{X} shares the same principal component coefficients in the full spectral space as those from X in the restricted space. Practically, the matrix B is solved by inverting P with the Gram-Schmidt orthogonalization method. Let P' be principal coefficients of the test spectra X' determined by projecting X' onto the eigenvectors $\{\xi\}$, we can then reconstruct the full test spectra via $X''^T = P'B$. Here X'' is the PCA-reconstructed spectra of the test spectra X' .

¹¹ <http://dr5.lamost.org>

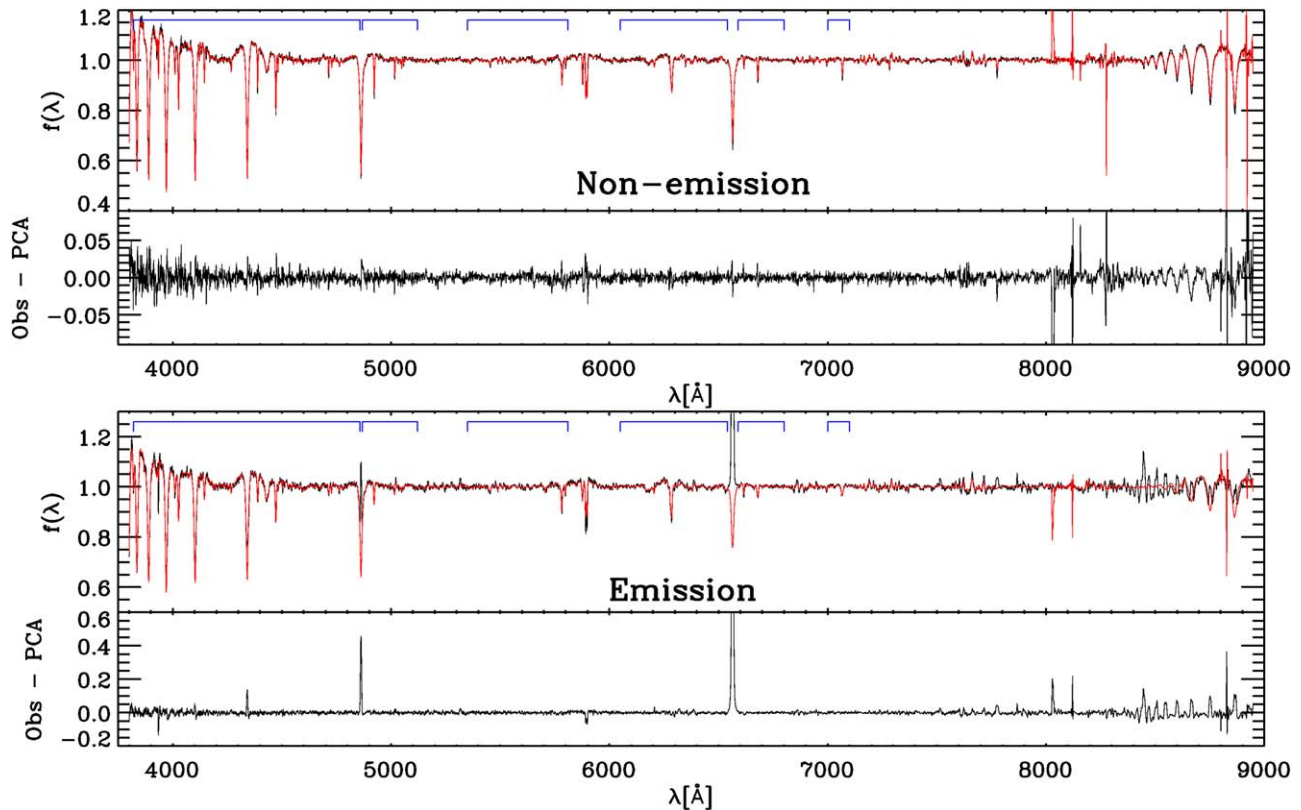


Figure 2. Two example spectra reconstructed with a PCA method for a non-emission (top) and an emission (bottom) star. In each panel, black is the LAMOST spectrum, while red is the PCA-reconstructed spectrum. Marked in blue are the clean wavelength windows with which we construct the principal component coefficients for the PCA reconstruction. Note that, throughout this work, we normalize the spectra with a pseudo-continuum derived by smoothing the spectra with a Gaussian kernel 50 \AA in width, as such some pixels in the continuum-normalized spectra can exceed unity.

Figure 2 shows that the full stellar spectra can be well reconstructed using the PCA method. Objects with emission lines are identified using residuals between the PCA-reconstructions and the LAMOST spectra. Figure 3 shows the mean H_{α} flux residual determined with our technique versus the reduced χ^2 measured around the H_{α} line. A clear branch of stars with flux excess at the position of the H_{α} line is present. We deem a star to have H_{α} emission if the flux excess exceeds the red vertical dashed line as delineated in Figure 3. Such a criterion identifies 2074 of the 16,002 unique LAMOST OB stars (13%) as emission-line stars.¹² Figure 3 also shows a small fraction of stars with large χ^2 . These stars are found to have erroneous spectra due to multiple reasons, such as data artifacts or wrong wavelength calibration. Figure 15 in Appendix A shows a few examples of such erroneous spectra.

4. Extinction

As OB stars are mostly located in the Galactic disk with ongoing star formation, they can suffer from serious interstellar extinction. Accurate correction for extinction is thus necessary to obtain accurate geometric M_{K_s} . Throughout this study, we refer to geometric M_{K_s} as the “apparent”, distance-corrected, M_{K_s} inferred using Gaia parallax and 2MASS K_s apparent magnitudes. We note that extinction is an issue even we work with the infrared K_s band. A reddening E_{B-V} of 1 mag, which is not uncommon for distant OB stars, can cause a ~ 0.3 mag

extinction in the K_s band (e.g., Yuan et al. 2013; Wang & Chen 2019), leading to a distance bias of 13%.

There are a variety of possible ways that we could obtain extinction for our OB star sample: e.g., through direct use of an existing 3D reddening map (e.g., Green et al. 2019; Chen et al. 2019b), the application of the Rayleigh–Jeans Color Excess method with infrared colors (RJCE; Majewski et al. 2011), or deducing from the intrinsic colors of OB stars either empirically (e.g., Deng et al. 2020) or theoretically from synthetic models. We adopt the latter in this study.

Deriving the intrinsic color requires a robust estimation of the stellar parameters. To achieve that, we leverage the stellar parameters derived for our OB star sample via the implementation of the spectral fitting codes THE PAYNE (Ting et al. 2019) in the hot star regime (M. Xiang et al. 2021, in preparation). Adopting T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ derived in this companion work, we estimate the intrinsic colors of the stars with the MIST isochrones (Choi et al. 2016). We have tested that using the PARSEC isochrones (Bressan et al. 2012) or the empirical T_{eff} –color relation of Deng et al. (2020) only incurs a difference of <0.03 mag for the E_{B-V} estimate, which is negligible for the purposes of this study.

We estimate E_{B-V} through the observed color excesses in the 2MASS J and H magnitudes (Skrutskie et al. 2006), Gaia DR2 BP/RP (Evans et al. 2018), and the g , r , i magnitudes from the XSTPS-GAC survey (Zhang et al. 2014; Liu et al. 2014). For bright ($r < 13$ mag) stars that the XSTPS-GAC photometry saturates, we adopt the APASS B , V , g , r , i photometric magnitudes (Henden et al. 2012; Munari et al. 2014) instead. To estimate E_{B-V} (and subsequently A_{K_s}), we opt to avoid

¹² We note that there could be multiple visits for the same star in the LAMOST database. For those stars, we adopt the results from the visit with the highest signal-to-noise ratio (S/N).

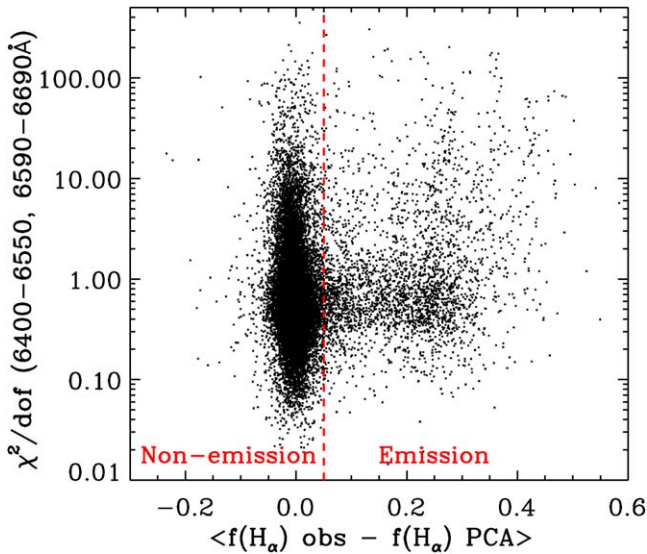


Figure 3. Observed H_α flux excess for emission spectra with respect to the PCA reconstruction. The horizontal axis shows the differences in mean H_α fluxes in $\lambda 6571\text{--}6557\text{ \AA}$ between the observed LAMOST spectra and the reconstructed PCA spectra. The vertical axis shows the reduced χ^2 across the wavelength range that encapsulates the H_α features ($\lambda 6400\text{--}6550\text{ \AA}$, $\lambda 6590\text{--}6690\text{ \AA}$). Spectra with emission lines exhibit a large difference in the H_α line between the LAMOST and PCA-reconstructed spectra and are clearly separated from the majority of stars. The vertical dashed line delineates our criterion to select OB spectra with emission lines.

using K_s -band photometry as it can be contaminated by emission from surrounding gas and dust disk. With all these N photometry bands, we construct $N - 1$ colors using only the photometry of adjacent bands. When compared to the intrinsic color, each observed color gives an estimate of E_{B-V} . We take the average of these E_{B-V} estimates weighted by uncertainties of the colors, which are assumed to be the quadratic sum of the uncertainties of the two photometric bands. On top of that, the uncertainties in E_{B-V} estimates are further derived through the propagation of uncertainties in photometric colors. Note that we have assigned a minimal uncertainty of 0.014 mag in all the colors, including the Gaia $BP - RP$.

To convert color excesses into E_{B-V} , we adopt extinction coefficients determined by convolving the public Kurucz model spectra (Castelli & Kurucz 2003; Kurucz 2005) with the Fitzpatrick (1999) extinction curve, assuming $R_V = 3.1$. We note that the R_V value may vary among different environments, from about 2 to 5 (e.g., Fitzpatrick & Massa 2007). In the case of $R_V = 5$, there will be a 0.2 mag difference in the extinction estimate for a star with $E_{B-V} = 1$, an extreme case for our sample stars. This will lead to a 9% bias in the distance estimate. First, we note that, a constant shift of R_V will not change the precision of our distance estimate. Furthermore, we expect such cases are rare for our sample stars, both because the $R_V = 3.1$ has shown to be a good approximation in most cases (for diffuse interstellar medium) (e.g., Fitzpatrick & Massa 2007; Schlafly et al. 2016) and that the E_{B-V} values of our sample stars are moderate as they are mostly in the Galactic anticenter direction. In addition, the extinction coefficient can vary among stars with different stellar parameters according to the SEDs (e.g., Green et al. 2021). However, considering we only focus on the hot stars, this effect is negligible because all the SEDs for hot stars are similar in the K_s band (the Rayleigh–Jeans tail). With these taken into account, we argue that the

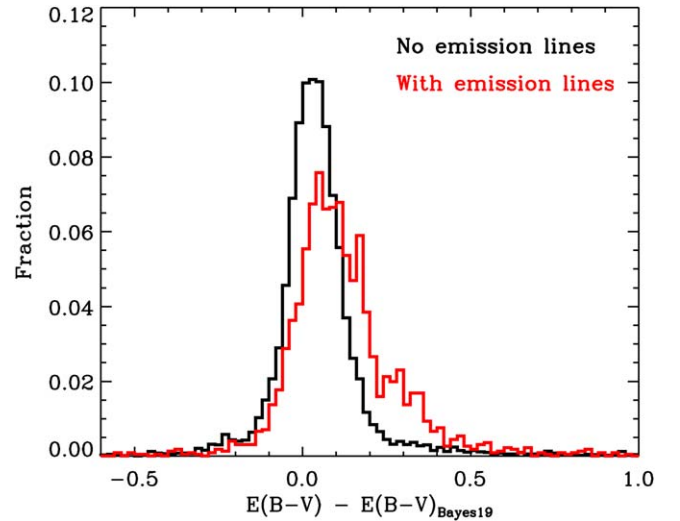


Figure 4. Comparison of the E_{B-V} derived in this work and the E_{B-V} from the 3D map of Green et al. (2019). We only consider stars with reliable Gaia parallax measurements ($\varpi/\sigma_\varpi > 5$). For stars without emission lines, the mean and dispersion of the E_{B-V} difference are 0.04 mag and 0.09 mag, respectively, whereas for stars with emission lines in the spectra we find a shifted mean of 0.11 mag and a standard deviation of 0.13 mag. The larger E_{B-V} for stars with emission lines indicates that these stars are more extinguished by the associated H II regions or by their surrounding gas and dust.

effect of R_V variation is likely to be subdominant, and we choose to derive the extinction coefficients using a fixed Kurucz spectrum with $T_{\text{eff}} = 12,000\text{ K}$, $\log g = 4.5$, and $[\text{Fe}/\text{H}] = 0$.

Figure 4 shows a comparison of the derived E_{B-V} with E_{B-V} interpolated from the 3D map of Green et al. (2019), using the Gaia distance from Bailer-Jones et al. (2018). The differences for the majority of stars are smaller than 0.1 mag, which corresponds to a negligible extinction difference of $\lesssim 0.03$ mag in the K_s band. There are a small fraction of stars with large E_{B-V} differences. Many of them turn out to be stars with emission lines. The large discrepancy could be caused by the fact that these stars are distant objects that are outside the feasible range of the Bailer-Jones et al. (2018) distance and/or the Green et al. (2019) reddening map. Alternatively, especially for stars with emission lines, they could suffer from additional extinction from the dense H II regions or surrounding disks. Due to the possibility of contamination in the K_s -band flux by the surrounding environment, we caution about distance estimates from stars with an emission-line flag in this study.

5. Data-driven M_{K_s} Estimation

The absolute magnitude (luminosity) is an intrinsic astrophysical property of a star that is derivable from a stellar spectrum. It is related to the stellar parameters via the Stefan–Boltzmann equation

$$L = 4\pi R^2 T_{\text{eff}}^4, \quad (3)$$

as well as the gravity equation

$$g = \frac{GM}{R^2}, \quad (4)$$

where T_{eff} , R , M , and g are, respectively, the effective temperature, radius, mass, and surface gravity of the star.

Equations (3) and (4) together yield

$$\log \frac{L}{L_{\odot}} = \log \frac{M}{M_{\odot}} + 4 \log \frac{T_{\text{eff}}}{T_{\text{eff},\odot}} - \log \frac{g}{g_{\odot}}. \quad (5)$$

Recall that T_{eff} and $\log g$ are basic stellar parameters that are derivable from stellar spectra. Furthermore, the stellar mass itself also implicitly depends on T_{eff} , $\log g$, abundance $[X/H]$ and rotation velocities, all of which are stellar properties that are readily measurable from stellar spectra. Consequently, one would expect that there exists an empirical relation that connects stellar spectra to the absolute magnitude of the stars. In this study, we will model this relation with a neural network.

5.1. The Neural Network Model

We consider a feed-forward multilayer perceptron neural network model that maps the LAMOST spectra to the absolute magnitude M_{K_s} of the stars. Adopting the Einstein sum notation, our neural network contains two-layers and can be succinctly written as

$$M_{K_s} = w \cdot \sigma(\tilde{w}_i \sigma(w_{\lambda i} f_{\lambda} + b_i) + \tilde{b}), \quad (6)$$

where σ is the Sigmoid activation function, w and b are weights and biases of the network to be optimized, the index i denotes the neurons, and λ denotes the wavelength pixels. We adopt 100 neurons for both layers. The training process is carried out with the Pytorch package in Python. To reduce overfitting in the training process, we also employ the dropout method (Srivastava et al. 2014) with a dropout parameter of 0.2.

Considering the impact of emission lines, we set up two neural network models. One neural network is constructed for stars without emission lines, using the whole wavelength range except for $\lambda 5720\text{--}6050 \text{ \AA}$, $\lambda 6270\text{--}6380 \text{ \AA}$, $\lambda 6800\text{--}6990 \text{ \AA}$, $\lambda 7100\text{--}7320 \text{ \AA}$, $\lambda 7520\text{--}7740 \text{ \AA}$, and $\lambda 8050\text{--}8350 \text{ \AA}$; these wavelength windows contain prominent absorption bands of the Earth atmospheres and strong Na I D absorption lines from the interstellar medium. The other neural network is for stars with emission lines, using only wavelength windows that are devoid of strong emission lines as demonstrated in Figure 2. Finally, considering that the neural network model is sensitive to spectral noise, we attempted to denoise spectra through the PCA reconstruction, using only the first 100 PCs. For non-emission stars, the PCA-reconstructed spectra are reconstructed using the full wavelength range, while for emission stars, the PCA-reconstructed spectra are reconstructed using only the clean wavelength windows.

5.2. The Training and Test Set

We define two training sets that correspond to the two neural network models set up above, for stars with or without emission lines. We also define a test set that we use to verify the M_{K_s} estimation.

The training and test sets adopt stars with good parallax measurements ($\varpi/\sigma_{\varpi} > 10$). To derive a more robust empirical relation, we require the training stars to have a spectral S/N (per pixel) higher than 50. Stars that meet these requirements are divided into two groups: four-fifths of them are adopted as the training set to train the neural network model, while the remaining one-fifth constitute the test set, in

combination with stars with good parallaxes but lower spectral S/N. Roughly 200 stars that are not in the Liu et al. (2019b) sample but have $M_{K_s} < -1.5$ mag are also added to the training set. The inclusion of these stars is to enlarge the sample size at the brighter end where there is a limited number of training stars. Although not in the original LAMOST OB stars catalog, all of these additional training stars have $T_{\text{eff}} > 7000$ K according to the LAMOST DR5 stellar parameter catalog of Xiang et al. (2019). Their spectra are further manually inspected to ensure they are early-type stars. About half of them are OB stars that exhibit He absorption lines, while the others are likely late-B or early A-type stars. In total, we obtain 6861 stars for the training set for stars without emission lines and 7161 stars for training set for stars with emission lines.

Because the geometric “apparent” M_{K_s} for binaries are biased, the training process is iterated to exclude the binaries. Binaries are singled out based on significant deviation between the predicted M_{K_s} and the geometric M_{K_s} (see Section 6 for details). In practice, two iterations are implemented, as we find negligible change in the resultant M_{K_s} estimates after these iterations.

Figure 5 shows the comparison of the resultant spectroscopic M_{K_s} estimates with the geometric M_{K_s} for the test stars with $S/N > 20$. The figure shows good overall consistency between the two sets of M_{K_s} estimates across the range roughly from -4 mag to 1.5 mag. Below $M_{K_s} \sim 0$ mag, more stars have fainter spectroscopic M_{K_s} than the geometric M_{K_s} . This is due to a contribution to the geometric M_{K_s} from binaries at these magnitudes. Recall that the geometric M_{K_s} is derived from the apparent magnitudes, where both stars contribute.

We note that there are some subdwarfs and white dwarfs in the LAMOST OB star sample. They typically have geometric M_{K_s} fainter than 2 mag and are not shown in the figure. For these stars, our spectroscopic M_{K_s} estimates, which are trained primarily on normal OB star spectra, can be problematic. The spectroscopic M_{K_s} for stars fainter than 2 mag should therefore be used with caution.

Figure 5 also demonstrates that the scatter between the spectroscopic M_{K_s} and the geometric M_{K_s} is larger for stars with emission lines than for the non-emission stars. Particularly, there are a number of emission-line stars that exhibit spectroscopic M_{K_s} much brighter than their geometric M_{K_s} . Peculiarly, we found that both spectroscopic M_{K_s} and geometric M_{K_s} of these stars appear to be robust. On the one hand, as illustrated in Figure 6, the spectroscopic M_{K_s} for these stars is consistent with the luminosity predicted by the temperature-weighted gravity (Kudritzki et al. 2020). On the other hand, for about half of these stars, their Gaia renormalized unit weight error (RUWE)¹³ values are around 1.0, suggesting that at least half of these outliers have decent astrometry measurements. Investigating the nature of these stars is beyond the scope of this study, but one possibility is that they might be stripped stars as a consequence of binary evolution (e.g., Götberg et al. 2018). As a result of the stripping, they are in reality fainter (as probed by the geometric M_{K_s}) while exhibiting similar spectra to a main-sequence or subgiant star, and hence a brighter spectroscopic M_{K_s} (see also Section 8).

¹³ A detailed explanation on the RUWE can be found in the public DPAC document from L. Lindegren, titled “Re-normalizing the astrometric chi-square in Gaia DR2”, via <https://www.cosmos.esa.int/web/gaia/public-dpac-documents> (at the bottom of the page).

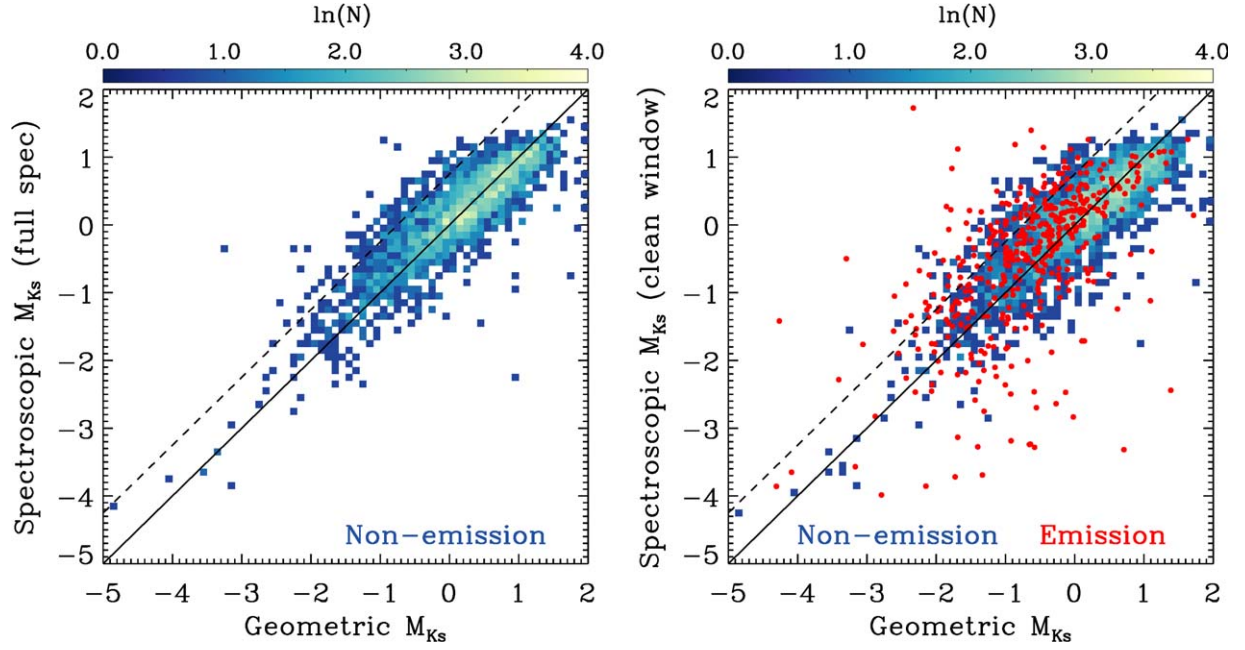


Figure 5. Comparison between the geometric M_{Ks} inferred from Gaia parallaxes and the spectroscopic M_{Ks} derived from LAMOST spectra for test stars. We only show results for test stars that have precise Gaia parallax measurements ($\varpi/\sigma_\varpi > 10$). The left panel demonstrates the spectroscopic M_{Ks} derived from the full spectrum for stars without emission lines. The right panel illustrates the spectroscopic M_{Ks} derived using only wavelength pixels that are devoid of emission lines for both non-emission stars (blue/green background) and emission-line stars (red dots). In both panels, the solid line delineates the one-to-one line. The dashed line shows an offset of 0.75 mag from the one-to-one line, the offset one would expect from equal-mass binaries. Stars that are close to the dashed line have geometric M_{Ks} brighter than the spectroscopic M_{Ks} , signaling the possibility of binaries or multiple systems.

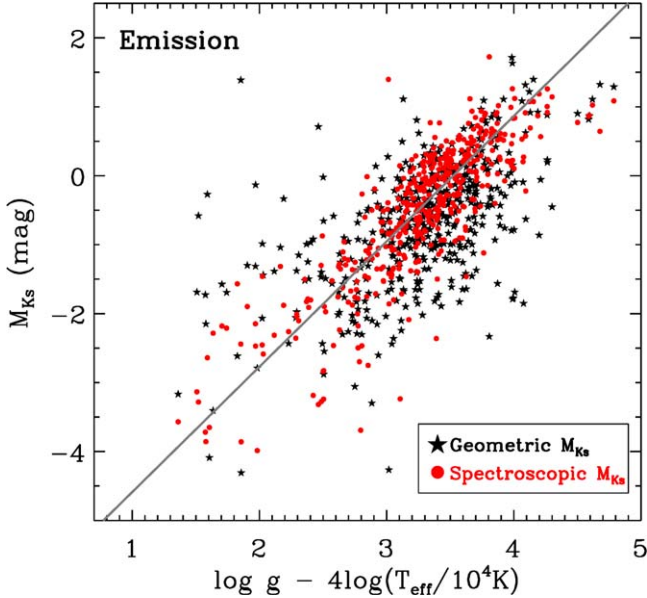


Figure 6. Temperature-weighted gravity vs. M_{Ks} for test stars with emission lines. The temperature-weighted gravity of a star is an indicator of its bolometric luminosity. Spectroscopic M_{Ks} measurements are shown with dot symbols while geometric M_{Ks} are shown with star symbols. The solid line indicates the best-fit linear relation from the sample of nonemission stars.

Figure 7 illustrates the difference between the spectroscopic M_{Ks} and the geometric M_{Ks} for the test stars as a function of stellar parameters T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$. The stellar parameters were derived from LAMOST spectra in a parallel work (M. Xiang et al. 2021, in preparation). Only results from single stars without emission lines are shown. This figure demonstrates that our spectroscopic M_{Ks} estimates do not exhibit bias with respect to stellar parameters.

5.3. Measurement Uncertainty and Intrinsic Uncertainty

In this section, we evaluate the quality of our spectroscopic M_{Ks} estimates. The nature of the uncertainty can be aleatoric or epistemic. For the former, the uncertainty in M_{Ks} is caused by uncertainties in the spectra. The latter can be caused by multiple sources. For example, the spectra simply do not contain the full information of the luminosity of the stars. On top of that, our data-driven models might also be suboptimal in extracting such information. In the following, we will quantify both the aleatoric and epistemic uncertainties using the test set.

To make a complete accounting of the uncertainties in our M_{Ks} estimates requires careful characterization of both the contribution from uncertainties in the geometric M_{Ks} and the additional scatter raised as a result of unresolved binaries. In order to minimize the impact of binary stars, we calculate the measurement uncertainty and intrinsic uncertainty with an iterative approach. In each iteration, we identify and discard all likely binaries that have more than 2σ difference between their spectroscopic M_{Ks} and geometric M_{Ks} .

The measurement uncertainty and intrinsic uncertainty are estimated in a Bayesian framework. In the following, we will denote the ground-truth geometric absolute magnitude of each star as M^g and the ground-truth spectroscopic absolute magnitude as M^s . We assume that there is a linear relation between the expected mean spectroscopic absolute magnitude \bar{M}^s and M^g , with a slope close to 1, with a correction term ε , and an intercept of δ ,

$$\bar{M}^s = (1 + \varepsilon)M^g + \delta. \quad (7)$$

We further assume that, for a given \bar{M}^s , due to epistemic uncertainty, the M^s of the individual stars are distributed as a

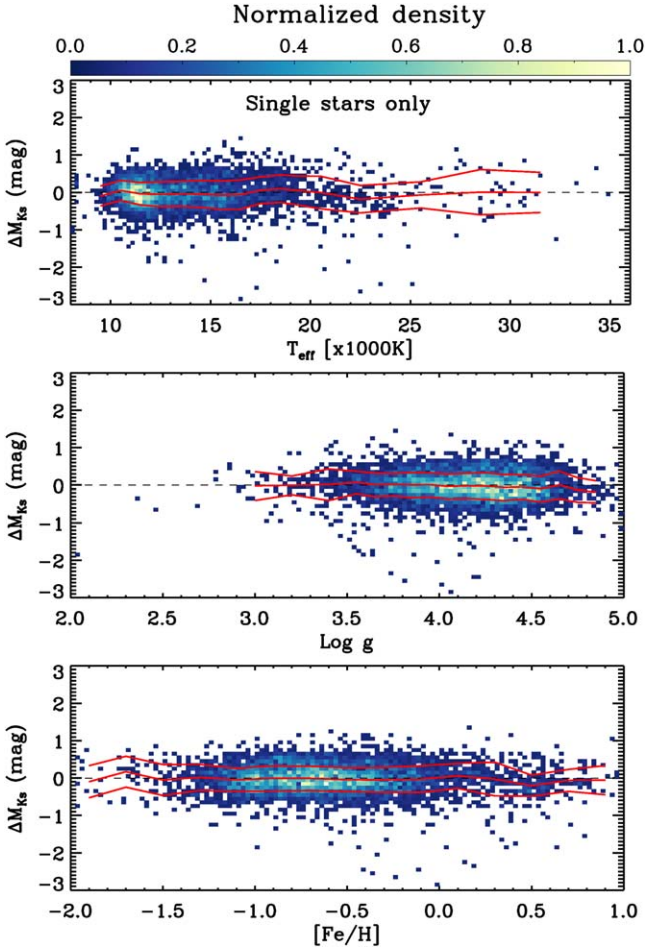


Figure 7. Differences between spectroscopic M_{K_s} and geometric M_{K_s} as a function of stellar parameters. Only results for single stars are shown. The solid lines delineate the median and standard deviation as a function of stellar parameters.

Gaussian with an intrinsic uncertainty of σ_{int} , i.e.,

$$P(M^s | \bar{M}^s, \sigma_{\text{int}}) = \mathcal{G}(\bar{M}^s, \sigma_{\text{int}}). \quad (8)$$

The estimated absolute magnitudes M_o^s are themselves assumed to distribute around a given M^s as a Gaussian distribution with width set by the aleatoric measurement uncertainty σ_s ,

$$P(M_o^s | M^s, \sigma_s) = \mathcal{G}(M^s, \sigma_s). \quad (9)$$

For simplicity, we assume that the aleatoric measurement uncertainty depends only on, and scales linearly with, the spectral S/N. In particular, we have

$$\sigma_s = \sigma_s^{100} / [(S/N)/100], \quad (10)$$

where σ_s^{100} is the uncertainty for spectra with S/N = 100.

Finally, for the geometric absolute magnitude, we assume a flat prior on the M^s . We can deduce that

$$P(M^s | m_0, \delta_{m_0}, \varpi, \delta\varpi) = \mathcal{G}(m_0 + 5 \log \varpi - 10 - M^s, \sigma_{\varpi, m_0}), \quad (11)$$

$$\sigma_{\varpi, m_0} \equiv \sqrt{\left(\frac{5}{\ln 10} \frac{\delta\varpi}{\varpi}\right)^2 + \delta_{m_0}^2}, \quad (12)$$

where ϖ and $\delta\varpi$ are the Gaia parallax and its uncertainty in units of milliarcseconds. m_0 and δm_0 are the dereddened apparent magnitude and its uncertainty, respectively. The latter is computed for each star as the quadratic sum of the uncertainties in the photometric magnitude and the extinction estimate.

Combining all these ingredients, we arrived at the final posterior probability distribution for the parameters ε , δ , σ_{int} , σ_s^{100} ,

$$\begin{aligned} &P(\varepsilon, \delta, \sigma_{\text{int}}, \sigma_s^{100} | \{M_{o,i}^s, (S/N)_i, m_{0,i}, \varpi_i, \delta_{i,m_0}, \delta\varpi_i\}) \\ &= \left(\prod_{i=1}^N P(M_{o,i}^s, (S/N)_i, m_{0,i}, \varpi_i, |\varepsilon, \delta, \sigma_{\text{int}}, \sigma_s^{100}, \delta_{i,m_0}, \delta\varpi_i) \right) \\ &P_{\text{prior}}(\varepsilon, \delta, \sigma_{\text{int}}, \sigma_s^{100}), \end{aligned} \quad (13)$$

where N is the number of stars, and the likelihood reads

$$\begin{aligned} &P(M_{o,i}^s, (S/N)_i, m_{0,i}, \varpi_i, |\varepsilon, \delta, \sigma_{\text{int}}, \sigma_s^{100}, \delta_{i,m_0}, \delta\varpi_i) \\ &= \mathcal{G}(M_{o,i}^s - [(1 + \varepsilon)(m_{0,i} + 5 \log \varpi_i - 10) + \delta], \\ &\sqrt{\sigma_{\text{int}}^2 + \left(\frac{\sigma_s^{100}}{(S/N)_i/100}\right)^2 + \left(\frac{5}{\ln 10} \frac{\delta\varpi_i}{\varpi_i}\right)^2 + \delta_{i,m_0}^2}) \end{aligned} \quad (14)$$

In this work, we adopt a flat prior for all the parameters and sample the posterior with Markov Chain Monte Carlo (MCMC).

Figure 8 displays the results of the MCMC fitting to non-emission single stars. The mean of posterior suggests an aleatoric measurement uncertainty of $\sigma_s^{100} = 0.10$ mag for spectra with S/N = 100 and an epistemic intrinsic uncertainty $\sigma_{\text{int}} = 0.25$ mag. We find a moderate slope correction of $\varepsilon = -0.18$, which is likely due to the lingering presence of binaries, considering that our 2σ cut can leave a considerable number of binaries with M_{K_s} excess smaller than 2σ ($\lesssim 0.5$ mag) in the sample. The presence of lingering binaries also implies that the intrinsic uncertainty from the MCMC posterior is a conservative estimate as binaries can contribute to part of the scatter.

6. Binary Identification

Binaries are ubiquitous and play important roles in astrophysics (Abt 1983; Duchêne & Kraus 2013; Moe & Di Stefano 2017). In star clusters, where all member stars share the same distance and the same age, binary stars in the lower main sequence are recognizable as they are brighter than all other single stars that distribute along a well-described locus in the color–magnitude diagram (e.g., Hurley & Tout 1998; Kouwenhoven et al. 2005; Li et al. 2013). The identification of binary stars in the field is more complicated due to the mixture of multiple stellar populations. There are a number of tailored approaches that have been implemented for binary identification in the field, such as interferometry (e.g., Raghavan et al. 2010), eclipsing transits (e.g., Qian et al. 2017; Zhang et al. 2017; Liu et al. 2018; Yang et al. 2020), color displacements (e.g., Pourbaix et al. 2004; Yuan et al. 2015a), astrometric noise excess (e.g., Kervella et al. 2019; Penoyre et al. 2020; Belokurov et al. 2020), common phase space motion (e.g., Andrews et al. 2017; Oh et al. 2017; Coronado et al. 2018b; El-Badry & Rix 2018; Hollands et al. 2018), radial velocity

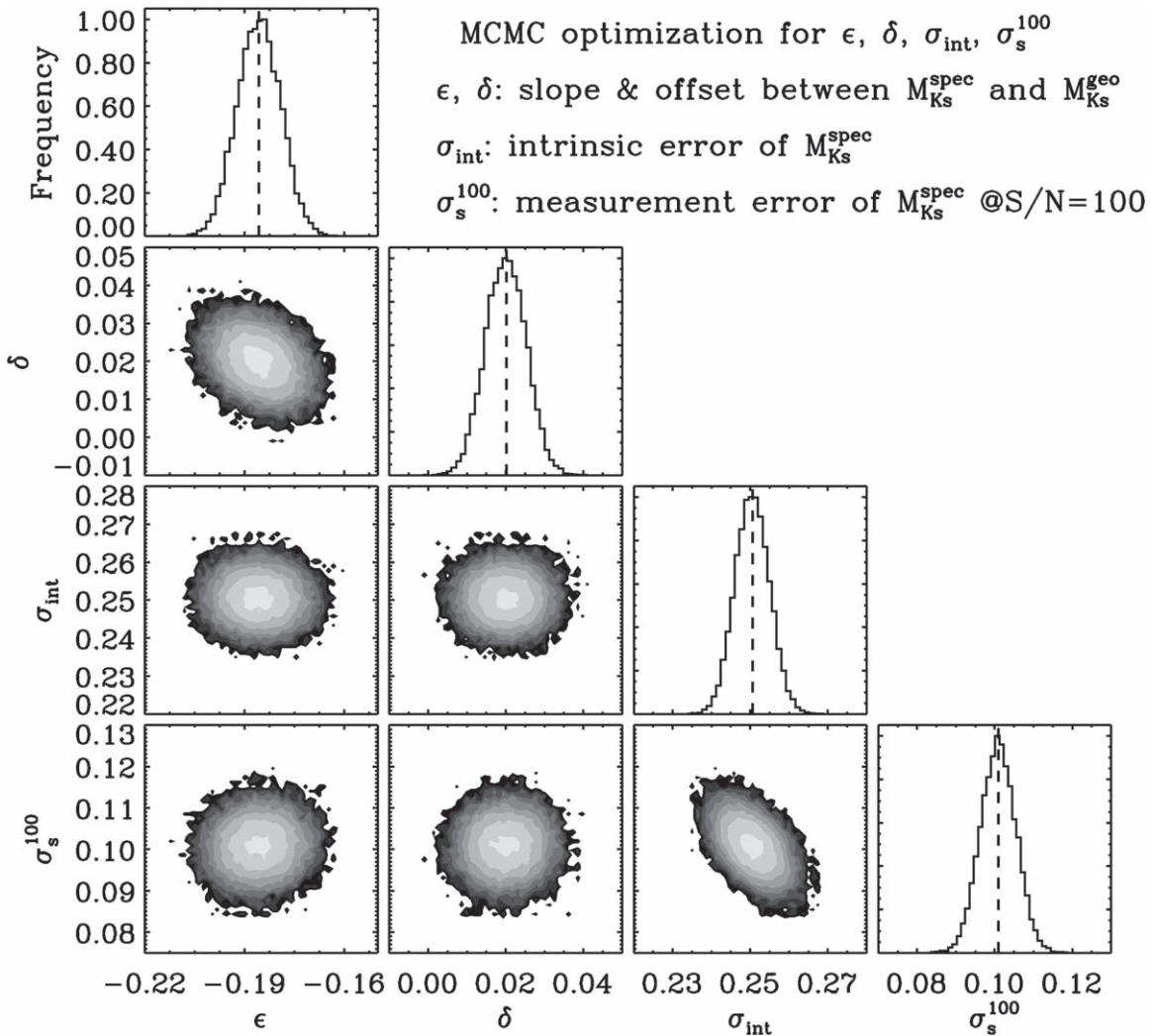


Figure 8. MCMC fitting of the intrinsic uncertainty and measurement uncertainty in our M_{K_s} estimates. Binaries were excluded iteratively. The intrinsic uncertainty σ_{int} indicates the epistemic uncertainty in M_{K_s} estimates, whereas the measurement uncertainty σ_s quantifies the aleatoric uncertainty induced by spectral noise. The MCMC results show an intrinsic uncertainty of $\sigma_{\text{int}} = 0.25$ mag and a typical measurement uncertainty of $\sigma_s = 0.10$ mag at S/N = 100.

variations (e.g., Matijević et al. 2011; Gao et al. 2014, 2017; Price-Whelan et al. 2017; Badenes et al. 2018; Tian et al. 2018, 2020), spectroscopic binaries with double lines (e.g., Fernandez et al. 2017; Merle et al. 2017; Traven et al. 2017; Skinner et al. 2018; Traven et al. 2020), and full spectral fitting (El-Badry et al. 2018b; Traven et al. 2020).

The various methods listed above have been extensively employed to identify and characterize binaries from large surveys. Belokurov et al. (2020) demonstrated that Gaia RUWE is an efficient tool for identifying unresolved, short-period binaries with low-to-intermediate-mass ratios. Short-period binaries have also been characterized through their double-line spectra from high-resolution spectroscopic surveys (e.g., Fernandez et al. 2017; Merle et al. 2017; Traven et al. 2020), or through radial velocity variations with both high- and low-resolution surveys (e.g., Gao et al. 2017; Badenes et al. 2018; Tian et al. 2018). Unresolved binaries with longer period, which exhibit single lines in their spectra, are generally harder to identify, but not impossible. Based on the full spectral fitting technique, El-Badry et al. (2018b) have characterized thousands of main-sequence binaries from the APOGEE spectra, many of them long-period binaries.

Nonetheless, for long-period binaries, the method presented in El-Badry et al. (2018b) is only mostly effective for systems with intermediate mass ratios ($0.4 \lesssim q \lesssim 0.85$; where $q = m_2/m_1$). It remains a challenge to identify unresolved single-line binaries with higher mass ratios ($q \gtrsim 0.85$). In this regime, identifying binaries via the binary sequence in the H-R diagram for cool stars with $T_{\text{eff}} \lesssim 5200$ K is possible (e.g., Gaia Collaboration et al. 2018b; Coronado et al. 2018a; Liu 2019), while this method is not as applicable to the hotter ($\gtrsim 5200$ K) stars or giants due to the larger intrinsic variation of luminosity (at a given T_{eff}).

Here we present a method of binary identification that tackles this challenging regime (long-period binaries with hot stars), leveraging differences between spectroscopic M_{K_s} and geometric M_{K_s} , or analogously, differences between the spectrophotometric parallaxes deduced from the spectroscopic M_{K_s} and the parallaxes determined with Gaia. The method is particularly efficient for identifying single-line binaries with large mass ratios, e.g., binaries with equal-mass components, thus serving as a complement to the aforementioned approaches. A brief introduction on the philosophy of the method has been laid out and applied to LAMOST AFGK stars

in Xiang et al. (2019). Here we present a more detailed description based on the application to LAMOST OB stars.

The basic idea is that, because the observed apparent magnitude of an unresolved binary/multiple-star system is brighter than any individual star in the system¹⁴, we should expect the geometric M_{K_s} for binary systems to be brighter than the spectroscopic M_{K_s} . This is possible because, while the geometric M_{K_s} reflects the contributions from both stars faithfully, the spectroscopic M_{K_s} mostly reflects the dominant star in the system. To demonstrate the latter, we build an empirical library of mock binary spectra using the LAMOST spectra for single stars. The mock test allows us to compare, and measure the differences between the spectroscopic M_{K_s} derived from composites and that from the individual components.

To generate the composite spectra, we scale the fluxes of the LAMOST spectra to the same distance. We restrict our spectral library to stars with robust Gaia parallax ($\varpi/\sigma_\varpi > 10$) and spectral S/N > 100, to ensure high data quality. Accurate spectral flux calibration is necessary to generate realistic composite spectra. For this purpose, we adopt the LAMOST spectra deduced with the flux calibration method of Xiang et al. (2015). The calibrated spectral SEDs have a relative precision of $\sim 10\%$ in the wavelength range of $\lambda 4000\text{--}9000\text{ \AA}$. We ensure that, for spectra in our library, the scaled fluxes are consistent with the photometry in all individual passbands, including the Gaia G -band, the XSTPS-GAC, and APASS g , r , and i band. Each of the spectra are further dereddened using the reddening estimates in Section 4, assuming the extinction curve from Fitzpatrick (1999). We assemble a set of binaries by taking an OB star for the primary star and a star of any (O/B/A/F/G/K) spectral type for the secondary. Figure 16 in Appendix B shows a few examples of the composite spectra.

Figure 9 shows the difference between the M_{K_s} derived from the composite binary spectra by simply treating them as single-star spectra and the M_{K_s} derived from the spectra of the primaries in these composites. In most cases, the spectroscopic M_{K_s} for the binary is comparable to that of the single, primary component. As expected, the spectroscopic M_{K_s} for equal-mass binary systems are identical to those of the primary component; the normalized spectra of the two component stars are identical. A similar result also applies to binary systems with small mass ratios. In this case, the secondary contributes minimally to the spectrum. Our mock test suggests that their spectroscopic M_{K_s} are fainter than that of the primary (by ~ 0.2 mag on average). Note that this is consistent with the findings of El-Badry et al. (2018a), which suggest that, for AFGK stars in binaries, the binary spectrum yields a larger $\log g$ than the single star. In any case, this effect facilitates the identification of binaries, as the difference between the geometric M_{K_s} and spectroscopic M_{K_s} would be even larger. In short, our experiment concludes that, unlike geometric M_{K_s} , spectroscopic M_{K_s} mostly reflects the contribution from the dominant stars, regardless of the mass ratio of the binary systems.

Nonetheless, we note that for some binary systems composed of an OB-type primary star with $M_{K_s} \gtrsim 0.5$ mag and an A/F/G/K-type secondary star, our spectroscopic M_{K_s} estimates from the composite spectra could be brighter than the primary by more than 0.2 mag. This is particularly common for

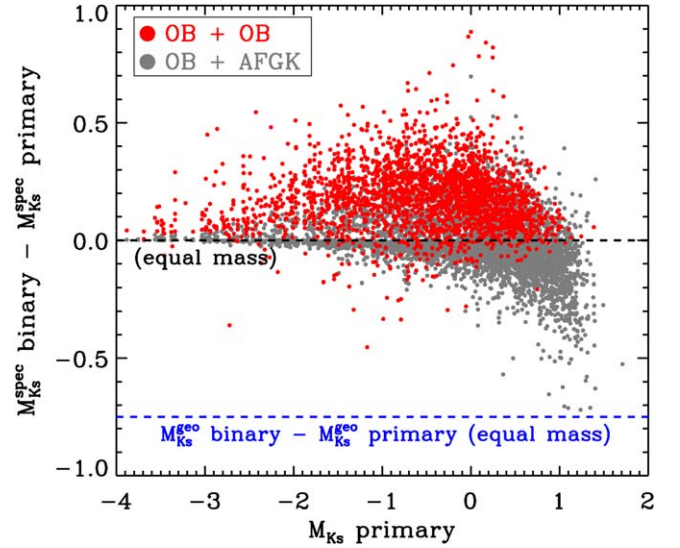


Figure 9. Differences between the spectroscopic M_{K_s} estimates derived from mock composite binary spectra vs. those for the primaries of the composites. Red symbols show OB+OB binary systems and the gray symbols are binary systems composed of an OB star and a companion with another (AFGK) spectral type. In most cases, the composite spectra cause the inferred spectroscopic M_{K_s} to be slightly fainter than the primary, further facilitating the identification of binaries through the difference between geometric M_{K_s} and spectroscopic M_{K_s} . For example, the geometric M_{K_s} for equal-mass binaries are 0.75 mag brighter than their primaries (dashed line in blue), due to the contribution from the secondary. We note that for some binaries composed of a late-B type primary with $M_{K_s} \gtrsim 0.5$ mag and an AFGK-type secondary, the spectroscopic M_{K_s} could be brighter than that of the primary. As such, the binary identification in this regime is less efficient (see text for details).

systems with a late-B-type primary. For these systems, the Balmer lines of the composite are shallower than the primary (Figure 16). The neural network model predicts a brighter M_{K_s} because the model is trained on single OB stars, for which the strength of Balmer lines decreases with increasing temperature, and hence a brighter M_{K_s} .

To identify binary stars, we compute the spectrophotometric parallax (in milliarcseconds),

$$\varpi_s = 10^{-0.2(m_{K_s} - M_{K_s} - 10 - A_{K_s})}, \quad (15)$$

and derive the S/N of the parallax excess of ϖ_s with respect to the Gaia astrometric parallax ϖ ,

$$S/N_{\Delta\varpi} = \frac{\varpi_s - \varpi}{\sqrt{\delta\varpi_s^2 + \delta\varpi^2}}, \quad (16)$$

where $\delta\varpi_s$ and $\delta\varpi$ are the measurement uncertainty of the spectrophotometric parallax and the Gaia parallax, respectively. The $\delta\varpi_s$ is defined via

$$\frac{\delta\varpi_s}{\varpi_s} = 0.2 \ln 10 \sqrt{\sigma_s^2 + \sigma_{\text{int}}^2 + \sigma_{m_{K_s}}^2 + \sigma_{A_{K_s}}^2}, \quad (17)$$

where σ_s and σ_{int}^2 are the (S/N-dependent) measurement uncertainty and the intrinsic uncertainty, derived in Section 5.3; $\sigma_{m_{K_s}}$ is the photometric uncertainty for the 2MASS K_s magnitude, and $\sigma_{A_{K_s}}$ the uncertainty of the extinction estimate. We adopt a 2σ criterion, i.e., we assign a star to be a binary if $S/N_{\Delta\varpi} > 2$. In total, 1597 of the 16,002 LAMOST OB stars in our sample (10.0%) are identified as binary stars, 13,257 (82.8%) are marked as single stars, and 1148 (7.2%) stars are

¹⁴ In general, the light centroids of binaries exhibit only little variation, resulting in only minor systematics in the astrometry, especially for binaries with similar mass companions.

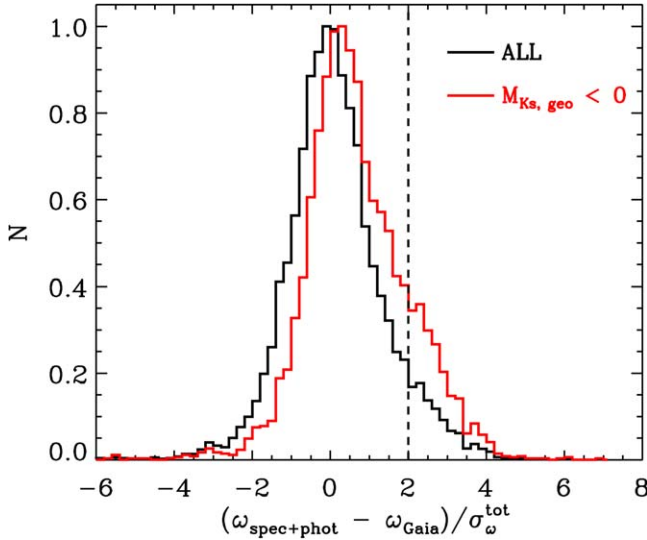


Figure 10. The normalized differences between the inferred spectrophotometric parallax and the Gaia parallax for individual LAMOST OB stars. Only the results from stars with robust Gaia parallaxes ($\varpi/\sigma_\varpi > 10$) and decent spectral quality ($S/N > 50$) are shown. The histogram in black shows results from stars of all geometric M_{K_s} , while the red one shows only results for stars with geometric $M_{K_s} < 0$. Our method is more effective for finding binaries for the latter. The vertical dashed line delineates the criterion adopted for binary identification.

unclassified due to a lack of either a Gaia parallax or 2MASS photometric magnitudes.

Figure 10 shows the differences between the spectrophotometric parallax and Gaia parallax for individual LAMOST OB stars. We show results from stars with robust Gaia parallaxes ($\varpi/\sigma_\varpi > 10$) and decent spectral quality ($S/N > 50$). Especially among stars with $M_{K_s} < 0$ mag, there is a clear positive tail, contributed by binary/multiple-star systems. These stars are overluminous compared to their single star counterparts.

Most of the binary stars selected with our method have a small RUWE value (~ 1). This partly reflects that our method is efficient for identifying binaries with large mass ratios, especially equal-mass binaries, for which the Gaia RUWE value is small due to negligible wobbles of the light centroids. Viewed in this way, our method complements approaches that identify binaries through large astrometric wobbles quantified by the Gaia RUWE value (e.g., Belokurov et al. 2020).

Nonetheless, a few caveats apply. As discussed above, many binaries with late-B-type primaries (with $M_{K_s} \gtrsim 0.5$ mag) can be missed. This is also illustrated in Figure 10, where the positive tail diminishes as we consider the full sample. The quality of the Gaia parallax is another limit to the effectiveness of this method. Because our method relies on the comparison between the spectrophotometric parallax and the Gaia astrometric parallax, the results are less robust in recognizing distant binary systems that have larger Gaia parallax uncertainty.

7. Distance

The distance to the star can be estimated by combining its Gaia parallax with the spectrophotometric distance derived from the spectroscopic M_{K_s} . We estimate the distance using the Bayesian scheme presented below.

In terms of M_{K_s} , m_{K_s} , extinction A_{K_s} , and Gaia parallax ϖ , the probability distribution function of distance d is

$$P(d|\varpi, M_{K_s}, m_{K_s}, A_{K_s}) = P(d|\varpi)P(d|M_{K_s}, m_{K_s}, A_{K_s}), \quad (18)$$

where

$$P(d|\varpi) = P(\varpi|d)P(d) \quad (19)$$

and

$$\begin{aligned} P(d|M_{K_s}, m_{K_s}, A_{K_s}) \\ = P(M_{K_s}|d, m_{K_s}, A_{K_s})P(m_{K_s}, A_{K_s}|d)P(d). \end{aligned} \quad (20)$$

The likelihood function can be written as

$$P(\varpi|d) = \mathcal{G}(1/d, \delta\varpi), \quad (21)$$

$$\begin{aligned} P(M_{K_s}|d, m_{K_s}, A_{K_s}) \\ = \mathcal{G}(M_{K_s} - (m_{K_s} - 5 \log d + 5 - A_{K_s}), \sigma_{\text{tot}}), \end{aligned} \quad (22)$$

$$\sigma_{\text{tot}} = \sqrt{\delta M_{K_s}^2 + \delta m_{K_s}^2 + \delta A_{K_s}^2}. \quad (23)$$

The extinction A_{K_s} is derived by

$$A_{K_s} = R_{K_s} \times E_{B-V}, \quad (24)$$

and R_{K_s} is the extinction coefficient in the 2MASS K_s passband. We adopt a flat prior $P(d)$, and $P(m, A|d)$. Note that, for binary systems, we adopt the Gaia parallax alone for distance estimation, as their spectrophotometric distances might be biased.

We sample the posterior distribution function (PDF) for individual stars with a fine distance step of 0.1 pc and adopt the mode of the PDF as the distance estimate and the 16th and 84th percentile as the 1σ estimates. We also sample the PDF in logarithmic distance scale. The PDFs in logarithmic distance are close to Gaussian, and we thus adopt the PDF-weighted mean and standard deviation as the estimates of the logarithmic distance and its uncertainty, respectively. The left panel of Figure 11 shows the distribution of distances to our sample of LAMOST OB stars. While the majority of stars are located within 3 kpc from the Sun, a number of them could lie beyond 10 kpc. Figure 11 illustrates the relative distance uncertainty as a function of distance. The median distance uncertainty of the sample is 8%, and the distance uncertainty only increases moderately with distance; the distance uncertainty is about $\sim 14\%$ at 15 kpc. The figure also shows the distance uncertainty when only the Gaia parallax is adopted. It illustrates that the inclusion of the spectroscopic M_{K_s} outperforms Gaia distance estimates for stars farther than ~ 1.5 kpc from the Sun, and the improvement becomes critical for stars more distant than about 5 kpc. This is consistent with Shull & Danforth (2019), who found substantial differences between spectrophotometric and parallax distances at $d > 1.5$ kpc. Although not shown, we have also checked the distance of Bailer-Jones et al. (2018) for our sample stars and found good consistency for stars with $d \lesssim 2$ kpc, a regime where their distance estimates are robust and are not dominated by the priors imposed in their studies. Nonetheless, we note that for binaries that are not identified with our method due to large uncertainties in either the geometric and/or spectroscopic M_{K_s} , the distance estimates may suffer large systematics, which can reach 35% in the case of unidentified equal-mass binaries when only spectroscopic M_{K_s} are available.

Figure 12 shows the LAMOST OB star sample in the X - Y plane in Galactic Cartesian coordinates (X , Y , Z). The Sun is assumed to be located at position $X = -8.1$ kpc, $Y = 0$ and $Z = 0$. The figure highlights the wide spatial range $-16 < X - 5$ kpc, $-4 < Y < 5$ kpc covered by the sample. A small number of stars outside these ranges are not shown here.

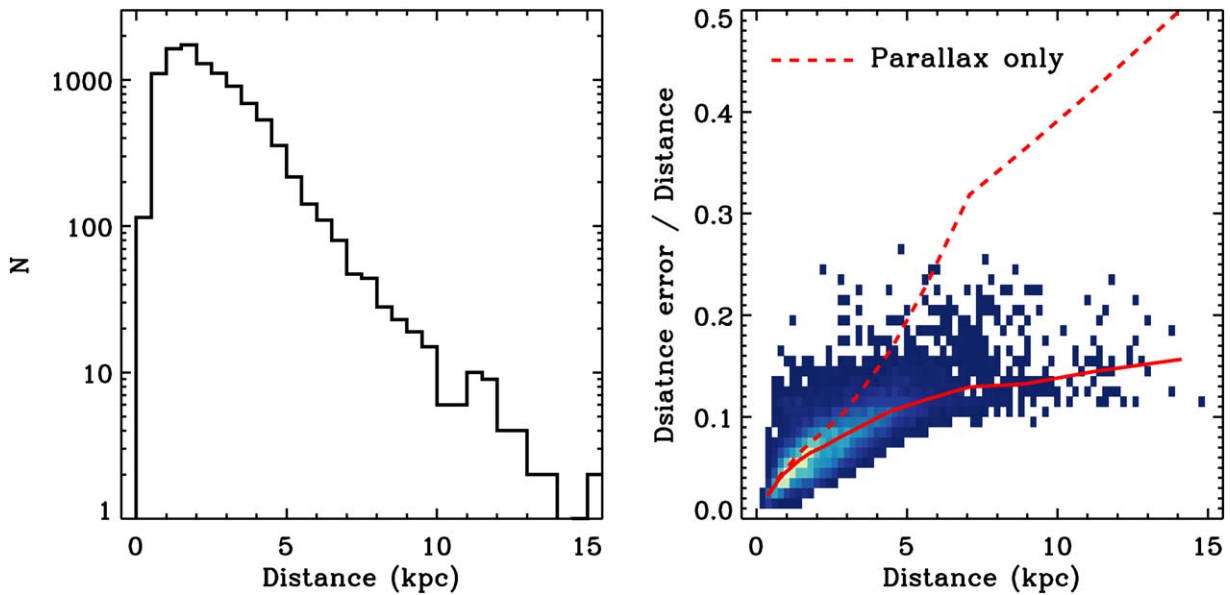


Figure 11. Left: distribution of heliocentric distance of the LAMOST OB star sample. Right: Relative distance uncertainty ($\sigma_d/d = \ln 10 \times \sigma_{\log d}$) as a function of distance. Only results for single stars are shown. The solid line delineates the median value of the relative distance uncertainty at various distances. The dashed line delineates the relative distance uncertainty in the case where only the Gaia parallax is adopted to infer the distance. The improvement to distance estimates using the spectroscopic M_{K_s} is clearly visible for distant stars.

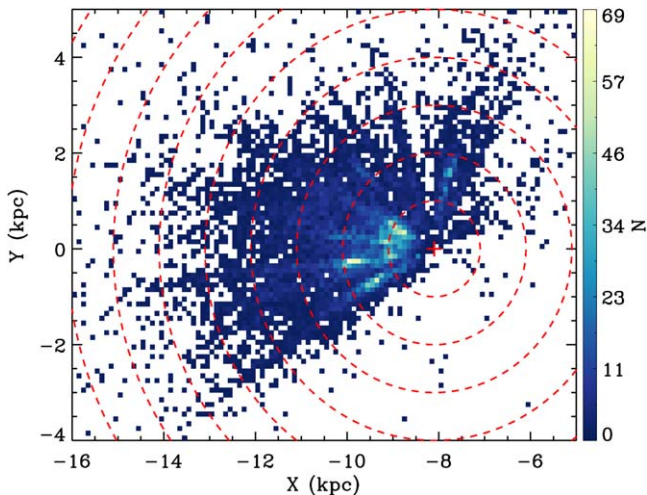


Figure 12. Spatial distribution of the LAMOST OB star sample in the disk X - Y plane in Galactic Cartesian coordinates. The plus symbol designates the position of the Sun ($X = -8.1$ kpc, $Y = 0$ kpc). The dashed rings delineate constant distances from the Sun in step of 1 kpc.

The data exhibit overdensities at approximately the distance of Perseus arm (e.g., at $X = -10$ kpc and $Y \simeq -0.3$ kpc), which is about 2 kpc from the Sun (e.g., Xu et al. 2006).

Finally, our estimates of M_{K_s} , extinction, and distance, and flags for binary and emission lines for the 16,002 LAMOST OB stars are made publicly available.¹⁵ Table 1 presents a summary of the catalog.

8. The Distance to the LB-1 B-star System

LB-1 (LS V +22 25; R.A. = 92°95450, decl. = 22°82575) is a binary system discovered by Liu et al. (2019a). It was

purported to consist of a B-type star that exhibits periodic radial velocity variations with an amplitude of around 50 km s^{-1} and a period of 78.9 days; the system also exhibits broad emission hydrogen lines that show periodic radial velocity variations of $\sim 10 \text{ km s}^{-1}$ (Liu et al. 2019a, 2020). Liu et al. (2019a) interpreted LB-1 as a B star orbiting a $68^{+11}_{-13} M_{\odot}$ black hole as the unseen primary companion, which, if true, would be the most massive stellar-mass black hole ever found. Since its discovery, there have been various controversies about the origin of this system (Abdul-Masih et al. 2020; El-Badry & Quataert 2020, 2021; Eldridge et al. 2020; Irrgang et al. 2020; Liu et al. 2020; Rivinius et al. 2020; Shenar et al. 2020; Simón-Díaz et al. 2020; Yungelson et al. 2020). Alternative explanations have been proposed. In particular, spectral disentangling (Shenar et al. 2020) has all but demonstrated that LB-1 is an SB2 binary with two luminous components: it shows a Be star, with its broad absorption lines and a surrounding emission-line disk, and a luminous hot star with low $\log g$, presumably recently stripped to its current mass of $\sim 1 M_{\odot}$. This has led Shenar et al. (2020) and El-Badry & Quataert (2020) to conclude that the low mass and high luminosity of the stripped star leads to the high radial velocity variations in the combined spectrum. There is no need, and presumably no room, for a black hole.

Here we investigate the M_{K_s} estimates of LB-1. Our catalog contains spectroscopic M_{K_s} and distance measurements from 12 individual LAMOST spectra for the LB-1 system, all with $S/N > 300$. From these, we obtain a spectroscopic M_{K_s} of -0.89 ± 0.30 mag, which is fainter than the geometric M_{K_s} based on Gaia parallax and 2MASS photometry (Figure 13). Particularly, Gaia eDR3 (Brown et al. 2020) yields a parallax of 0.359 ± 0.030 mas for LB-1. The zero-point offset correction of Lindegren (2020) suggests a zero-point offset of -0.051 mas for LB-1. These lead to a geometric M_{K_s} of $-1.67^{+0.15}_{-0.16}$ mag, with a significantly better precision compared to the Gaia DR2 value ($-1.39^{+0.39}_{-0.47}$ mag).

¹⁵ The catalog is published online as a machine-readable version of Table 1. It can also be accessed via a temporary link at <https://keeper.mpd.mpg.de/f/56d86145cfb0417eb8a8/?dl=1>.

Table 1Descriptions for the Distance Catalog of 16,002 OB Stars in LAMOST DR5^a

Field	Description
specid	LAMOST spectra ID in the format of “date-planid-spid-fiberid”
fitsname	Name of the LAMOST spectral .FITS file
ra	R.A. from the LAMOST DR5 catalog (J2000; deg)
dec	Decl. from the LAMOST DR5 catalog (J2000; deg)
uniqflag	Flag to indicate repeat visits; uqflag = 1 means unique star, uqflag = 2, 3, ..., n indicates the n th repeat visit For stars with repeat visits, the uniqflag is sorted by the spectral S/N, with uqflag = 1 having the highest S/N
star_id	A unique ID for each unique star based on its RA and Dec, in the format of “Sdddmmss±ddmmss”
snr_g	Spectral signal-to-noise ratio per pixel in SDSS g-band
rv	Radial velocity from LAMOST (km s^{-1})
rv_err	Uncertainty in radial velocity (km s^{-1})
M_{K_s}	M_{K_s} estimated from LAMOST spectra
$M_{K_s_err}$	Uncertainty in M_{K_s}
$M_{K_s_geo}$	Geometric M_{K_s} inferred from Gaia parallaxes and 2MASS apparent magnitudes
$M_{K_s_geo_err}$	Uncertainty in $M_{K_s_geo}$
dis	Distance at the mode of the distance probability density function (PDF)
dis_low	Distance at the 16th percentile of the cumulative probability distribution function
dis_high	Distance at the 84th percentile of the cumulative probability distribution function
logdis	PDF-weighted mean logarithmic distance
logdis_err	Uncertainty in logdis
ebv	Reddening estimated in this work
ebv_err	Uncertainty in ebv
snr_dparallax	Excess in spectrophotometric parallax with respect to the Gaia astrometric parallax
binary_flag	Flag of binarity; 1 = binary ($\text{snr_dparallax} \geq 2$), 0 = single ($\text{snr_dparallax} < 2$), -9 = unknown
em_flag	Flag of emission lines; 1 = with emission lines, 0 = no emission lines
gaia_id	Gaia DR2 Source ID
parallax	Gaia DR2 parallax (mas)
parallax_error	Uncertainty in gaia_parallax (mas)
parallax_offset	Offset of Gaia parallax according to the offset— G magnitude relation of Leung & Bovy (2019)
pmra	Gaia DR2 proper motion in R.A. direction
pmra_error	Uncertainty in pmra
pmdec	Gaia DR2 proper motion in decl. direction
pmdec_error	Uncertainty in pmdec
ruwe	Gaia DR2 RUWE
J	2MASS J -band magnitude
J_err	Uncertainty in J
H	2MASS H -band magnitude
H_err	Uncertainty in H
K_s	2MASS K_s -band magnitude
K_s_err	Uncertainty in K_s
X/Y/Z	3D position in the Galactic Cartesian coordinates (kpc)

Note.

^a Due to multiple visits of common stars, the catalog contains 27,784 entries for 16,002 unique stars. This is slightly different from the original catalog of Liu et al. (2019b), which contains 22,901 spectra entries for 16,032 stars. We have increased the number of spectra by including all repeat visits in the LAMOST DR5 database. All of the spectra can be found on the LAMOST DR5 website.

(This table is available in its entirety in machine-readable form.)

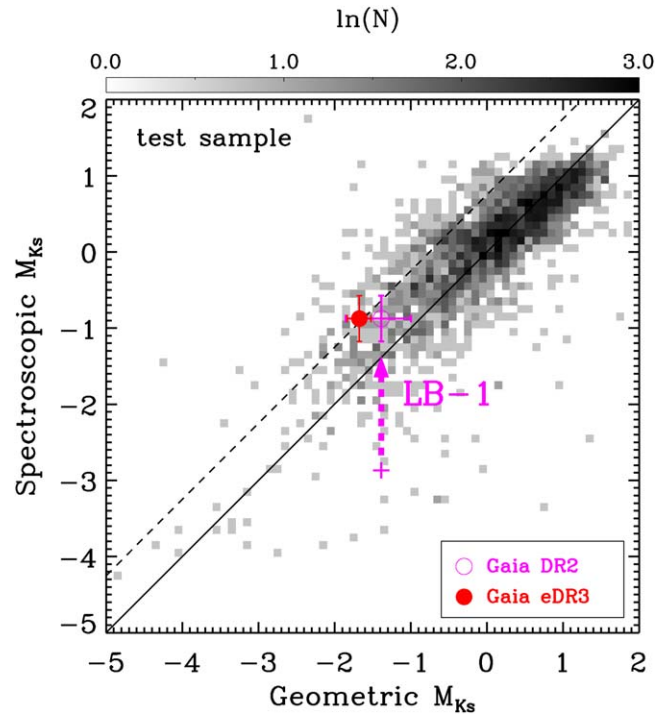


Figure 13. Comparison of spectroscopic M_{K_s} with geometric M_{K_s} for a test star sample with precise Gaia parallax ($\tilde{\omega}/\sigma_{\tilde{\omega}} > 10$). The solid line delineates the one-to-one line, while the dashed line delineates an offset of 0.75 mag to the one-to-one line. The inferred spectroscopic M_{K_s} and geometric M_{K_s} of the LB-1 system are highlighted by the pink circle with error bars. The plus symbol in pink shows the M_{K_s} (-2.9 mag) inferred from the distance in Liu et al. (2019a), which is at odds with our spectroscopic M_{K_s} . The red dot with error bars shows the LB-1 geometric M_{K_s} based on the Gaia eDR3 parallax, which has a significantly smaller error bar compared to the Gaia DR2 result. It shows that LB-1 is a binary system, for which the component stars are comparably bright.

These results suggest that the geometric M_{K_s} of LB-1 is significantly brighter than the spectroscopic M_{K_s} by more than 0.6 mag, implying the LB-1 is a binary system that contains two luminous stars. This is largely consistent with the conclusion of Shenar et al. (2020) and El-Badry & Quataert (2020) and is in tension with the conclusion of Liu et al. (2019a) that LB-1 is a system composed of one luminous B star and a massive stellar-mass black hole.

As an independent check, we adopt stellar parameters for LB-1 derived from the LAMOST spectra by fitting the Kurucz ATLAS12 model spectra (Kurucz 1970, 1993) with THE PAYNE (M. Xiang et al. 2021, in preparation) and evaluate the temperature-weighted gravity (Kudritzki et al. 2020) as a luminosity indicator. Note that, for hot stars, the rotation may play a significant role in determining the luminosity. While here we have ignored the rotation effect due to the lack of accurate rotation measurements for our sample stars. Nonetheless, Figure 14 illustrates that there is a good relation between the inferred spectroscopic M_{K_s} and the T_{eff} -weighted gravity. The LB-1 stellar parameters are in line with such a relation, leading credence to our spectroscopic M_{K_s} estimate of LB-1. In short, our exploration suggests that both the geometric M_{K_s} from the Gaia parallax and the spectroscopic M_{K_s} from the LAMOST spectra for LB-1 are likely to be robust. The intrinsic difference between the geometric M_{K_s} and the spectroscopic

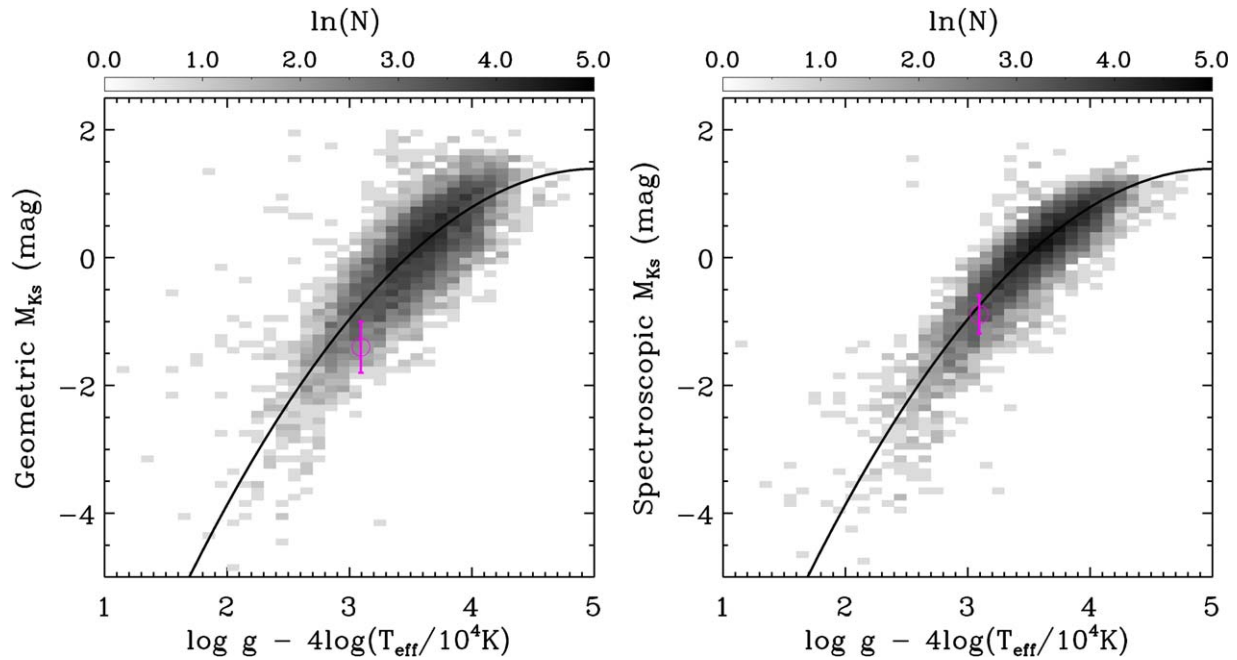


Figure 14. The temperature-weighted gravity vs. the geometric M_{K_s} (the left panel) and the spectroscopic M_{K_s} (the right panel). The temperature-weighted gravity is adopted as a “ground-truth” luminosity indicator. Only stars with precise Gaia parallax ($\bar{\varpi}/\sigma_{\bar{\varpi}} > 10$) are shown. The solid line is a second-order polynomial fit to the geometric M_{K_s} as a function of the temperature-weighted gravity. The circle with error bar in pink highlights the LB-1 B-star companion. The spectroscopic M_{K_s} estimate for LB-1 is perfectly consistent with the temperature-weighted gravity. The geometric M_{K_s} is brighter than the prediction from the gravity.

M_{K_s} could be explained if LB-1 is a binary system with two luminous component stars.

9. Summary

In this study, we have presented a data-driven approach for deriving K_s -band absolute magnitudes M_{K_s} for OB stars from low-resolution ($R \simeq 1800$) LAMOST spectra. Our method uses a neural network model trained on a set of stars with good parallaxes from Gaia DR2. Applying to a test data set, we find that the neural network is capable of delivering M_{K_s} with 0.25 mag precision from the LAMOST OB star spectra. We have also applied the method separately to stars with emission lines in their spectra. The emission-line spectra are identified through comparing the observed spectra and the PCA reconstruction of the spectra.

We verify that the M_{K_s} estimated from the composite spectrum of a binary system is comparable to, or slightly fainter than, the M_{K_s} of the primary star. This is in contrast to the geometric M_{K_s} calculated from Gaia parallaxes, as both components of the binary contribute to the geometric M_{K_s} . We propose a new method of binary identification, leveraging differences between the spectroscopic M_{K_s} and the geometric M_{K_s} . The method is particularly effective for identifying equal-mass binaries or multiple-star systems, because the geometric M_{K_s} of these systems are much brighter than their primaries. Our method is generic and can be applied to any combined astrometric and spectroscopic data beyond this study.

With the spectroscopic M_{K_s} determinations, we derive accurate distances to 16,002 OB stars from the LAMOST sample of Liu et al. (2019b). The median distance uncertainty for our sample stars is 8%, and the distance uncertainty for the most distant stars at more than 10 kpc away is about 14%. We present a value-added catalog of OB stars for future studies of the structure and dynamics of the Galactic disk. Besides

absolute magnitudes and distances, the catalog presents also emission-line flags and binary flags for the LAMOST OB stars, significantly expanding the number of known emission-line objects and binaries for massive stars, especially those with mass ratios close to unity. Our method yields a spectral M_{K_s} of 0.89 ± 0.30 mag from the LAMOST spectra of LB-1. However, the geometric M_{K_s} of LB-1 derived from the Gaia parallax, both Gaia DR2 and Gaia eDR3, is significantly brighter than the spectroscopic M_{K_s} , suggesting that LB-1 is likely a binary system that contains two luminous stars with comparable brightness. This supports the previous conclusion of Shenar et al. (2020) and El-Badry & Quataert (2020), and contradicts Liu et al. (2019a), who argued that LB-1 is a system composed of one luminous B star and a massive stellar-mass black hole.

H.-W.R. acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 138713538—SFB 881 (“The Milky Way System”, subproject A03). M.X. is grateful for Dr. Bodem for the successful dental surgery and the attentive care from him during recovery. Y.S.T. is grateful to be supported by the NASA Hubble Fellowship grant HST-HF2-51425.001 awarded by the Space Telescope Science Institute.

This work has made use of data acquired through the Guoshoujing Telescope. Guoshoujing 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.

This work has also made use of data from the European Space Agency (ESA) mission Gaia, processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

Appendix A Examples of Problematic LAMOST Spectra

As shown in Figure 2 in the main text, there are some stars with large χ^2 in the residuals between the LAMOST spectra and the PCA reconstruction. We find that, for these stars, the LAMOST spectra are often problematic due to various reasons,

including instrument problems, erroneous wavelength calibration, and other reasons. Figure 15 shows a few examples of the erroneous LAMOST spectra.

Appendix B Examples of Mock Binary Spectra

As discussed in Section 6, in order to study the spectroscopic M_{K_s} derived from composite spectra, we build an empirical library of binary spectra using the LAMOST spectra of single stars. Figure 16 shows a few examples of our mock composite spectra as well as their inferred spectroscopic M_{K_s} . The spectroscopic M_{K_s} estimates of the binary spectra typically agree with those from the spectra of the primary stars, with a difference of $\lesssim 0.2$ mag.

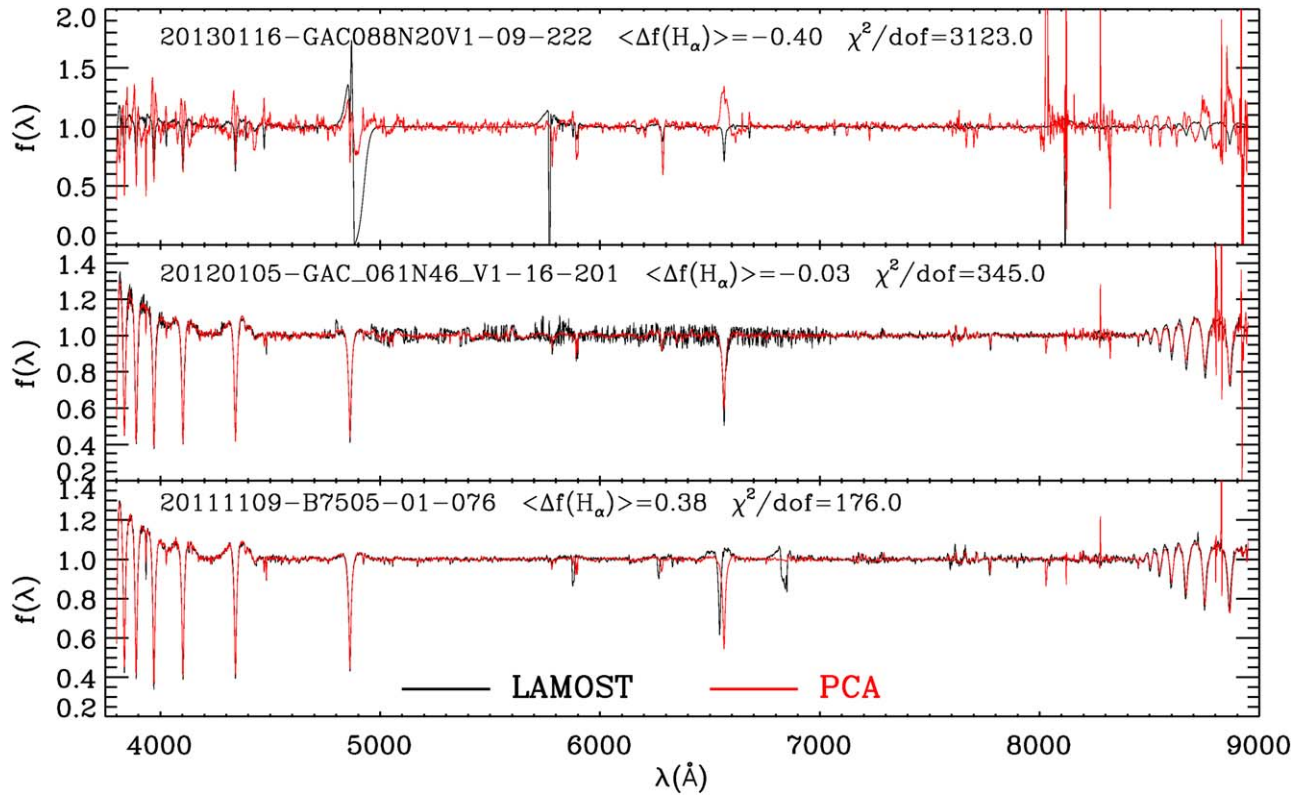


Figure 15. A few typical examples that exhibit abnormal residuals between the LAMOST spectrum and the PCA reconstruction. The top panel shows a spectrum with problematic fluxes over the wavelength range of $\lambda 4700\text{--}5000$ Å, leading to poor PCA reconstruction. The middle panel shows a spectrum with artifacts in the LAMOST spectra across $\lambda 5000\text{--}7000$ Å. The PCA reconstruction is visually reasonable, but the χ^2 between the LAMOST and PCA-reconstructed spectra is suboptimal. The bottom panel shows a spectrum that has a problematic wavelength calibration from the LAMOST pipeline in the wavelength range of $\lambda 5800\text{--}7000$ Å.

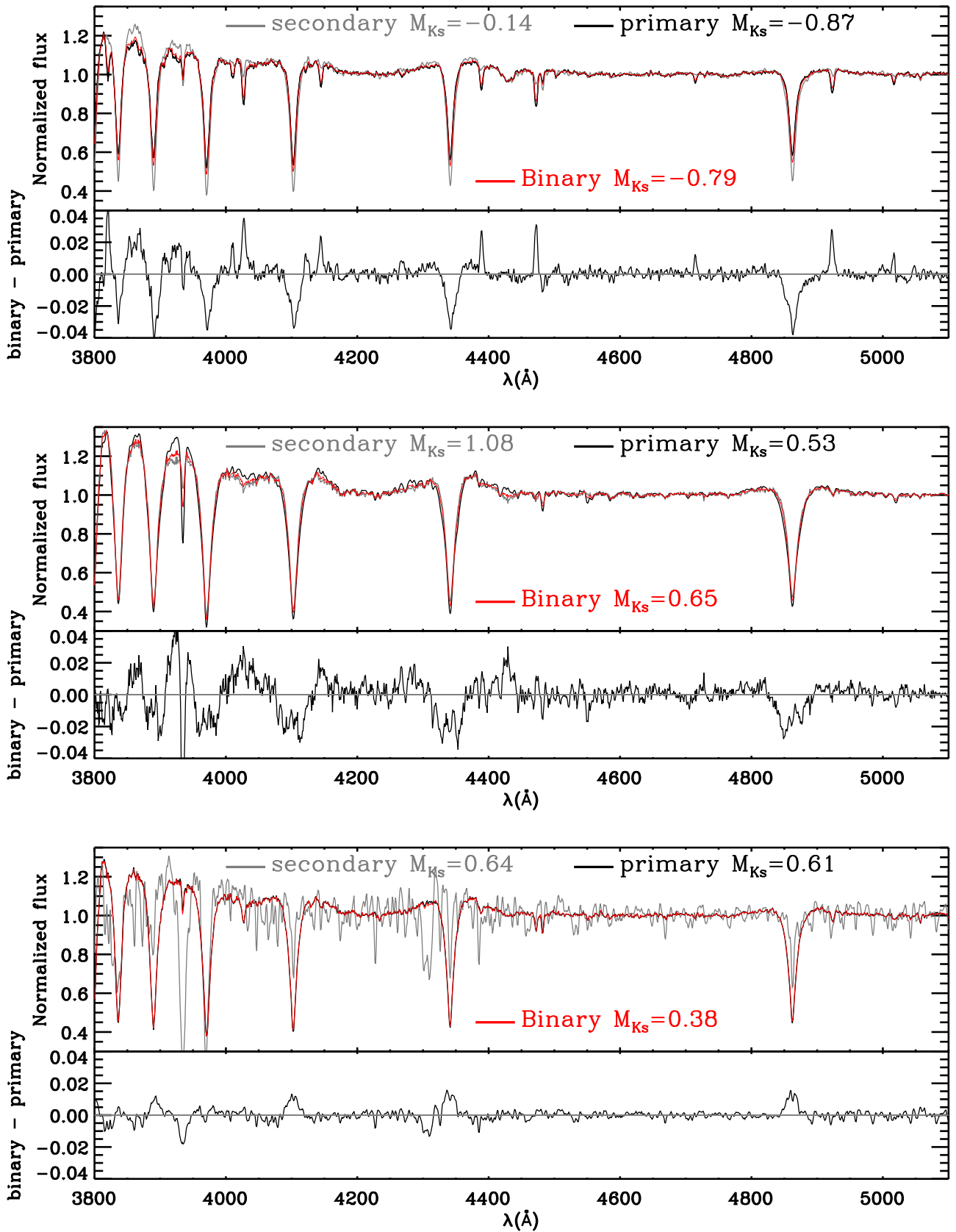


Figure 16. A few typical examples of mock composite spectra. We focus on the wavelength range of λ 3800–5100 Å. The spectra of the primary, secondary, and binary are shown in black, gray, and red, respectively. For all these cases, the primary is a B-type star. The secondary are B-, A- and G-type stars from the top to bottom, respectively. The spectroscopic M_{Ks} estimates applied to the primary, secondary, and composite binary spectra are also shown in the figure. The spectroscopic M_{Ks} estimate of the binary spectra typically agrees with the one from the primary spectra, with a difference $\lesssim 0.2$ mag.

ORCID iDs

Maosheng Xiang  <https://orcid.org/0000-0002-5818-8769>
 Hans-Walter Rix  <https://orcid.org/0000-0003-4996-9069>
 Yuan-Sen Ting  <https://orcid.org/0000-0001-5082-9536>
 Kareem El-Badry  <https://orcid.org/0000-0002-6871-1752>
 Hai-Bo Yuan  <https://orcid.org/0000-0003-2471-2363>
 Wen-Yuan Cui  <https://orcid.org/0000-0003-1359-9908>

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, *PhRvL*, **116**, 061102
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, *PhRvL*, **116**, 241103
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *PhRvL*, **119**, 141101
 Abdul-Masih, M., Banyard, G., Bodensteiner, J., et al. 2020, *Natur*, **580**, E11
 Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, **235**, 42
 Abt, H. A. 1983, *ARA&A*, **21**, 343
 Andrews, J. J., Chanamé, J., & Agüeros, M. A. 2017, *MNRAS*, **472**, 675
 Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, *ApJ*, **854**, 147
 Bailor-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, *AJ*, **156**, 58
 Bartko, H., Martins, F., Trippe, S., et al. 2010, *ApJ*, **708**, 834
 Belczynski, K., Holz, D. E., Bulik, T., & O'Shaughnessy, R. 2016, *Natur*, **534**, 512
 Belokurov, V., Penoyre, Z., Oh, S., et al. 2020, *MNRAS*, **496**, 1922
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
 Brown, A. G. A., Vallenari, A., Prusti, T., et al. 2020, *A&A*, in press
 Carlin, J. L., Liu, C., Newberg, H. J., et al. 2015, *AJ*, **150**, 4
 Castelli, F., & Kurucz, R. L. 2003, in *IAU Symp. 210, Modelling of Stellar Atmospheres*, ed. N. Piskunov, W. W. Weiss, & D. F. Gray (Cambridge: Cambridge Univ. Press), A20
 Chen, B. Q., Huang, Y., Hou, L. G., et al. 2019a, *MNRAS*, **487**, 1400
 Chen, B. Q., Huang, Y., Yuan, H. B., et al. 2019b, *MNRAS*, **483**, 4277
 Cheng, X., Liu, C., Mao, S., & Cui, W. 2019, *ApJL*, **872**, L1
 Chieffi, A., & Limongi, M. 2004, *ApJ*, **608**, 405
 Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
 Coronado, J., Rix, H.-W., & Trick, W. H. 2018a, *MNRAS*, **481**, 2970
 Coronado, J., Sepúlveda, M. P., Gould, A., & Chanamé, J. 2018b, *MNRAS*, **480**, 4302
 Deng, D., Sun, Y., Jian, M., Jiang, B., & Yuan, H. 2020, *AJ*, **159**, 208
 Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA*, **12**, 735
 Duchêne, G., & Kraus, A. 2013, *ARA&A*, **51**, 269
 El-Badry, K., & Quataert, E. 2020, *MNRAS*, **493**, L22
 El-Badry, K., & Quataert, E. 2021, *MNRAS*, **502**, 3436
 El-Badry, K., & Rix, H.-W. 2018, *MNRAS*, **480**, 4884
 El-Badry, K., Rix, H.-W., Ting, Y.-S., et al. 2018a, *MNRAS*, **473**, 5043
 El-Badry, K., Ting, Y.-S., Rix, H.-W., et al. 2018b, *MNRAS*, **476**, 528
 Eldridge, J. J., Stanway, E. R., Breivik, K., et al. 2020, *MNRAS*, **495**, 2786
 Evans, D. W., Riello, M., De Angeli, F., et al. 2018, *A&A*, **616**, A4
 Fernandez, M. A., Covey, K. R., De Lee, N., et al. 2017, *PASP*, **129**, 084201
 Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
 Fitzpatrick, E. L., & Massa, D. 2007, *ApJ*, **663**, 320
 Freyer, T., Hensler, G., & Yorke, H. W. 2003, *ApJ*, **594**, 888
 Freyer, T., Hensler, G., & Yorke, H. W. 2006, *ApJ*, **638**, 262
 Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, **595**, A1
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, *A&A*, **616**, A1
 Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018b, *A&A*, **616**, A10
 Gao, S., Liu, C., Zhang, X., et al. 2014, *ApJL*, **788**, L37
 Gao, S., Zhao, H., Yang, H., & Gao, R. 2017, *MNRAS*, **469**, L68
 Götzberg, Y., de Mink, S. E., Groh, J. H., et al. 2018, *A&A*, **615**, A78
 Green, G. M., Rix, H.-W., Tschesche, L., et al. 2021, *ApJ*, **907**, 57
 Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, *ApJ*, **887**, 93
 Henden, A. A., Levine, S. E., Terrell, D., Smith, T. C., & Welch, D. 2012, *JAVSO*, **40**, 430
 Hogg, D. W., Eilers, A.-C., & Rix, H.-W. 2019, *AJ*, **158**, 147
 Hollands, M. A., Tremblay, P. E., Gänsicke, B. T., Gentile-Fusillo, N. P., & Toonen, S. 2018, *MNRAS*, **480**, 3942
 Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, *MNRAS*, **445**, 581
 Humphreys, R. M., & McElroy, D. B. 1984, *ApJ*, **284**, 565
 Hurley, J., & Tout, C. A. 1998, *MNRAS*, **300**, 977
 Irrgang, A., Geier, S., Kreuzer, S., Pelisoli, I., & Heber, U. 2020, *A&A*, **633**, L5
 Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, *ApJ*, **684**, 287
 Jofré, P., Mädler, T., Gilmore, G., et al. 2015, *MNRAS*, **453**, 1428
 Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, *ApJ*, **673**, 864
 Kervella, P., Arenou, F., Mignard, F., & Thévenin, F. 2019, *A&A*, **623**, A72
 Kouwenhoven, M. B. N., Brown, A. G. A., Zinnecker, H., Kaper, L., & Portegies Zwart, S. F. 2005, *A&A*, **430**, 137
 Kudritzki, R. P., Bresolin, F., & Przybilla, N. 2003, *ApJL*, **582**, L83
 Kudritzki, R.-P., Urbaneja, M. A., & Rix, H.-W. 2020, *ApJ*, **890**, 28
 Kurucz, R. L. 1970, *SAOSR*, **309**
 Kurucz, R. L. 1993, *SYNTHE Spectrum Synthesis Programs and Line Data* (Cambridge, MA: Smithsonian Astrophysical Observatory)
 Kurucz, R. L. 2005, *MSAIS*, **8**, 14
 Lequeux, J. 1979, *A&A*, **80**, 35
 Leung, H. W., & Bovy, J. 2019, *MNRAS*, **489**, 2079
 Li, C., de Grijs, R., & Deng, L. 2013, *MNRAS*, **436**, 1497
 Li, C., Zhao, G., Jia, Y., et al. 2019, *ApJ*, **871**, 208
 Lindegren, L. 2020, *A&A*, **633**, A1
 Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, **616**, A2
 Liu, C. 2019, *MNRAS*, **490**, 550
 Liu, J., Zhang, H., Howard, A. W., et al. 2019a, *Natur*, **575**, 618
 Liu, J., Zheng, Z., Soria, R., et al. 2020, *ApJ*, **900**, 42
 Liu, N., Fu, J. N., Zong, W., et al. 2018, *AJ*, **155**, 168
 Liu, X. W., Yuan, H. B., Huo, Z. Y., et al. 2014, in *IAU Symp. 298, Setting the Scene for Gaia and LAMOST*, ed. S. Feltzing et al. (Cambridge: Cambridge Univ. Press), 310
 Liu, Z., Cui, W., Liu, C., et al. 2019b, *ApJS*, **241**, 32
 Mackey, J., Gvaramadze, V. V., Mohamed, S., & Langer, N. 2015, *A&A*, **573**, A10
 Maíz-Apellániz, J., Walborn, N. R., Galué, H. Á., & Wei, L. H. 2004, *ApJS*, **151**, 103
 Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, **154**, 94
 Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, **739**, 25
 Matijević, G., Zwitter, T., Bienaymé, O., et al. 2011, *AJ*, **141**, 200
 Merle, T., Van Eck, S., Jorissen, A., et al. 2017, *A&A*, **608**, A95
 Moe, M., & Di Stefano, R. 2017, *ApJS*, **230**, 15
 Munari, U., Henden, A., Frigo, A., et al. 2014, *AJ*, **148**, 81
 Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C., & Maeda, K. 2006, *NuPhA*, **777**, 424
 Oh, S., Price-Whelan, A. M., Hogg, D. W., Morton, T. D., & Spergel, D. N. 2017, *AJ*, **153**, 257
 Penoyre, Z., Belokurov, V., Wyn Evans, N., Everall, A., & Koposov, S. E. 2020, *MNRAS*, **495**, 321
 Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, **500**, 501
 Pourbaix, D., Ivezić, Ž., Knapp, G. R., Gunn, J. E., & Lupton, R. H. 2004, *A&A*, **423**, 755
 Price-Whelan, A. M., Hogg, D. W., Foreman-Mackey, D., & Rix, H.-W. 2017, *ApJ*, **837**, 20
 Qian, S.-B., He, J.-J., Zhang, J., et al. 2017, *RAA*, **17**, 087
 Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, *A&A*, **638**, A76
 Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, *ApJS*, **190**, 1
 Reed, B. C. 2005, *AJ*, **130**, 1652
 Rivinius, T., Baade, D., Hadrava, P., Heida, M., & Klement, R. 2020, *A&A*, **637**, L3
 Sana, H., de Koter, A., de Mink, S. E., et al. 2013, *A&A*, **550**, A107
 Sana, H., de Mink, S. E., de Koter, A., et al. 2012, *Sci*, **337**, 444
 Schlafly, E. F., Meisner, A. M., Stutz, A. M., et al. 2016, *ApJ*, **821**, 78
 Shenar, T., Bodensteiner, J., Abdul-Masih, M., et al. 2020, *A&A*, **639**, L6
 Shu, F. H. 2016, *ARA&A*, **54**, 667
 Shull, J. M., & Danforth, C. W. 2019, *ApJ*, **882**, 180
 Simón-Díaz, S., Maíz Apellániz, J., Lennon, D. J., et al. 2020, *A&A*, **634**, L7
 Skinner, J., Covey, K. R., Bender, C. F., et al. 2018, *AJ*, **156**, 45
 Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *J. Mach. Learn. Res.*, **15**, 1929
 Struck, C. 2020, *MNRAS*, **494**, 1838
 Thielemann, F. K., & Arnett, W. D. 1985, *ApJ*, **295**, 604
 Tian, Z., Liu, X., Yuan, H., et al. 2020, *ApJS*, **249**, 22
 Tian, Z.-J., Liu, X.-W., Yuan, H.-B., et al. 2018, *RAA*, **18**, 052
 Timmes, F. X., Woosley, S. E., & Weaver, T. A. 1995, *ApJS*, **98**, 617
 Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, *ApJ*, **879**, 69
 Torra, J., Fernández, D., & Figueras, F. 2000, *A&A*, **359**, 82
 Traven, G., Feltzing, S., Merle, T., et al. 2020, *A&A*, **638**, A145
 Traven, G., Matijević, G., Zwitter, T., et al. 2017, *ApJS*, **228**, 24

- Wang, H. F., Huang, Y., Zhang, H. W., et al. 2020, *ApJ*, 902, 70
- Wang, J., Shi, J., Zhao, Y., et al. 2016, *MNRAS*, 456, 672
- Wang, S., & Chen, X. 2019, *ApJ*, 877, 116
- Woosley, S. E., & Weaver, T. A. 1995, *ApJS*, 101, 181
- Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, *ApJS*, 245, 34
- Xiang, M. S., Liu, X. W., Shi, J. R., et al. 2017b, *MNRAS*, 464, 3657
- Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, *MNRAS*, 448, 90
- Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2017a, *MNRAS*, 467, 1890
- Xu, Y., Bian, S. B., Reid, M. J., et al. 2018, *A&A*, 616, L15
- Xu, Y., Reid, M. J., Zheng, X. W., & Menten, K. M. 2006, *Sci*, 311, 54
- Yang, F., Long, R. J., Shan, S.-S., et al. 2020, *ApJS*, 249, 31
- Yuan, H., Liu, X., Xiang, M., et al. 2015a, *ApJ*, 799, 135
- Yuan, H. B., Liu, X. W., Huo, Z. Y., et al. 2015b, *MNRAS*, 448, 855
- Yuan, H.-B., Liu, X.-W., & Xiang, M.-S. 2013, *MNRAS*, 430, 2188
- Yungelson, L. R., Kuranov, A. G., Postnov, K. A., & Kolesnikov, D. A. 2020, *MNRAS*, 496, L6
- Zhang, H.-H., Liu, X.-W., Yuan, H.-B., et al. 2014, *RAA*, 14, 456
- Zhang, J., Qian, S.-B., & He, J.-D. 2017, *RAA*, 17, 22
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *RAA*, 12, 723