The Astrophysical Journal Supplement Series, 249:22 (13pp), 2020 August https://doi.org/10.3847/1538-4365/ab9904

A Catalog of RV Variable Star Candidates from LAMOST

Zhijia Tian[1], Xiaowei Liu[2], Haibo Yuan[3], Xuan Fang[4,5], Bingqiu Chen[2], Maosheng Xiang[6], Yang Huang[2], Shaolan Bi[3], Wuming Yang[3], Yaqian Wu[7], Chun Wang[8,10], Huawei Zhang[8], Zhiying Huo[7], Yong Yang[2], Gaochao Liu[9], Jincheng Guo[8,10], and Meng Zhang[8]

[1] Department of Astronomy, Key Laboratory of Astroparticle Physics of Yunnan Province, Yunnan University, Kunming 650200, People's Republic of China tianzhijia@ynu.edu.cn
[2] South-Western Institute for Astronomy Research, Yunnan University, Chenggong District, Kunming 650500, People's Republic of China; x.liu@pku.edu.cn
[3] Department of Astronomy, Beijing Normal University, Beijing 100875, People's Republic of China
[4] Key Laboratory of Optical Astronomy, National Astronomical Observatories of Chinese Academy of Sciences (NAOC), 20A Datun Road, Chaoyang District, Beijing 100101, People's Republic of China
[5] Department of Physics, Faculty of Science, The University of Hong Kong, Pokfulam Road, Hong Kong, People's Republic of China
[6] Max-Planck Institute for Astronomy, Heidelberg, Germany
[7] National Astronomy Observatories, Chinese Academy of Sciences, Beijing 100012, People's Republic of China
[8] Department of Astronomy, Peking University, Beijing 100871, People's Republic of China
[9] China Three Gorges University, Yichang 443002, People's Republic of China
Received 2019 February 19; revised 2020 May 28; accepted 2020 May 29; published 2020 July 28

Abstract

Radial velocity (RV) variable stars are important in astrophysics. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) spectroscopic survey has provided ∼6.5 million stellar spectra in its Data Release 4 (DR4). During the survey ∼4.7 million unique sources were targeted and ∼1 million stars observed repeatedly. The probabilities of stars being RV variables are estimated by comparing the observed RV variations with simulated ones. We build a catalog of 80,702 RV variable candidates with probability greater than 0.60 by analyzing the multi-epoch sources covered by LAMOST DR4. Simulations and cross-identifications show that the purity of the catalog is higher than 80%. The catalog consists of 77% binary systems and 7% pulsating stars as well as 16% pollution by single stars. 3138 RV variables are classified through cross-identifications with published results in literatures. By using the 3138 sources common in both LAMOST and a collection of published RV variable catalogs, we are able to analyze LAMOST's RV variable detection rate. The efficiency of the method adopted in this work relies not only on the sampling frequency of observations but also periods and amplitudes of RV variables. With the progress of LAMOST, Gaia, and other surveys, more and more RV variables will be confirmed and classified. This catalog is valuable for other large-scale surveys, especially for RV variable searches. The catalog will be released according to the LAMOST Data Policy via http://dr4.lamost.org.

Unified Astronomy Thesaurus concepts: Variable stars (1761); Catalogs (205); Radial velocity (1332); Extrinsic variable stars (514); Intrinsic variable stars (859); Binary stars (154); Pulsating variable stars (1307); Star counts (1568); Stellar astronomy (1583); Astrostatistics (1882); Spectroscopy (1558); Surveys (1671)

Supporting material: FITS file

1. Introduction

Binary stars play a crucial role in astrophysics. Statistics and identifications of binary systems are significant for several reasons, the major ones being that such basic issues as star formation and evolution, the initial mass function (IMF), and Galactic chemical evolution are all influenced by the binary properties of the stellar population. Despite the high fraction of binary stars (∼50% for main-sequence stars), our understandings of the physics of binary stars are still at a basic stage. Raghavan et al. (2010) presents the multiplicity of 454 solar-type stars within 25 pc at high completeness. They show that early-type and metal-poor stars dominate higher binary factions than late-type and metal-rich stars. The period distribution of the sample follows a log-normal distribution with a median of about 300 yr. Meanwhile, early- and late-type stars do not stem from the same parent period distribution (Kroupa & Petr-Gotzens 2011). The discrepancy is yet to be explained and could be related to the mechanism of binary formation.

A summary on empirical knowledge of stellar multiplicity for embedded protostars, pre-main-sequence stars, main-sequence stars, and brown dwarfs is performed by Duchêne & Kraus (2013). It is demonstrated that the multiplicity rate and breadth of the orbital period distribution are steep functions of the primary mass and environment. More efforts in recent years have been made in analyses of binary fractions based on large samples of survey data (e.g., Duquennoy & Mayor 1991; Gao et al. 2014, 2017; Yuan et al. 2015a; Badenes et al. 2018; Tian et al. 2018, hereafter Paper I). These works investigate the binary fractions against stellar parameters, i.e., mass, $T_{\rm eff}$, and abundance. All the researches indicate that metal-poor stars have a higher binary fraction than metal-rich stars. However, metal-rich disk stars are found to be 30% more likely to have companions with periods shorter than 12 days than metal-poor halo stars (Hettinger et al. 2015). The binary fraction is not only related to stellar parameters but also orbital periods (Maxted et al. 2001; Moe & di Stefano 2017).

[10] LAMOST Fellow.

Besides estimating binary fractions in large samples, identifications of binary systems have been carried out. The American Association of Variable Star Observers contributes to building an International Variable Star Index (VSX; Watson et al. 2006). A database of thousands of eclipsing binaries is established (Matijevič et al. 2012, and references therein) with Kepler light curves (Borucki et al. 2010; Koch et al. 2010). Drake et al. (2014) present ∼47,000 periodic variables found during the analysis of 5.4 million variable star candidates covered by the Catalina Surveys Data Release-1 (CSDR1; Drake et al. 2012) and investigate the rate of confusion between objects classified as contact binaries and type c RR Lyrae (RRc's) based on periods, amplitudes, radial velocities (RVs) and stellar parameters. The General Catalog of Variable Stars (GCVS) containing binary stars is released in the latest version (GCVS Version 5.1; Samus' et al. 2017). The Binary star DataBase collects data on physical and positional parameters of 240,000 components of 110,000 multiple-star systems (Kovaleva et al. 2015). Price-Whelan et al. (2018) make use of the multi-epoch data obtained with the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017; Abolfathi et al. 2018) and select ∼5000 evolved stars with probable companions. To build a sample of distant halo wide binaries, Coronado et al. (2018) search stellar pairs with small differences in proper motion and small projected separation on the sky as binary candidates and validate the sample through RVs from medium- and low-resolution spectra obtained with the Sloan Digital Sky Survey (York et al. 2000). Binaries and triples are identified using high-dispersion spectra, which can be much better fit with a superposition of two or three model spectra, drawn from the same isochrone, than any single-star model. El-Badry et al. (2018) apply the data-driven spectral model to APOGEE DR13 spectra of main-sequence stars and identify unresolved multiple-star systems. Gaia Data Release 2 (GDR2; Gaia Collaboration et al. 2018) enables catalogs of variable stars (Mowlavi et al. 2018; Roelens et al. 2018; Clementini et al. 2019; Rimoldini et al. 2019).

However, binary identification based on RVs derived from a low-dispersion spectroscopic spectra survey is still almost blank. Fortunately, the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) provided millions of stellar spectra, of which about 20% of the targets have been observed repeatedly. The quantity of these spectra can enhance time-domain studies of stars, stellar parameters, and their RVs and help select and confirm variable candidates. We build a catalog of RV variable candidates detected with LAMOST. In Section 2, we describe the data used in this work. The method is presented in Section 3. The results are shown in Section 4, followed by discussions and conclusions in Section 5.

2. Data

The LAMOST spectroscopic survey provides the largest database of low-resolution ($R \sim 2000$) spectra to measure stellar atmospheric parameters and RVs for millions of stars (Cui et al. 2012; Deng et al. 2012; Zhao et al. 2012; Liu et al. 2014; Yuan et al. 2015b). The survey has obtained ∼6.5 million spectra for ∼4.7 million unique stars in its DR4. In this survey, ∼1 million stars have been observed in 2 to over 40 epochs. As presented in Figure 1, the number of stars decreases with the number of epochs nearly exponentially. By comparing multi-epoch observations, especially for RVs, variable stars could be detected. We adopt the RV and stellar parameters yielded by the LAMOST stellar parameter pipeline at Peking University (LSP3[11]; Xiang et al. 2015, 2017) to select variable stars. The pipeline estimates RV through cross-correlating with an ELODIE template (Prugniel & Soubiran 2001; Prugniel et al. 2007) with close values of atmospheric parameters. For the determinations of stellar atmospheric parameters (e.g., $T_{\rm eff}$, log g, and [Fe/H]), templates from the Medium-resolution Isaac Newton Telescope library of empirical spectra (MILES; Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011), obtained with a spectral resolving power similar to that of LAMOST spectra and accurately flux calibrated, are used instead. As discussed in Xiang et al. (2015), the MILES spectra with low-resolution are wavelength calibrated to an accuracy of only approximately 10 km s^{-1}, which is not good enough for the purpose of RV determinations for the LAMOST spectra. However, the ELODIE library of high-resolution spectra is more appropriately used as RV templates. Furthermore, the LSP3 estimates an RV prior to atmospheric parameters, which avoids systemic uncertainties of RVs caused by adopting different spectral libraries in the pipeline. In the latest version of LSP3, 267 new template spectra obtained using the National Astronomical Observatories, Chinese Academy of Sciences (NAOC) 2.16 m telescope and the Yunnan Astronomical Observatory 2.4 m telescope (obtained by Wang et al. 2018) have been added to the MILES library to generate parameter estimates (Xiang et al. 2017).

The LSP3 pipeline ignores the effects of binary stars when estimating RV and other stellar parameters. Most of the stars have an RV error of a few km s^{-1}. However, some of them, mostly hot stars with low signal-to-noise ratios (S/Ns), have errors as large as 20 km s^{-1} (Xiang et al. 2015, 2017). To identify binary systems or candidates reliably, we limit the S/N of spectra greater than 10.

2.1. RVs and Their Uncertainties

As discussed in Xiang et al. (2015, 2017), the $\sigma_{\rm RV}$ is quite sensitive to S/N and depends on other stellar parameters. The LSP3 pipeline estimates $\sigma_{\rm RV}$ by comparing RVs from multi-epoch observations of similar S/Ns and spectral types, assuming that $\sigma_{\rm RV}$ is contributed from random error following a Gaussian distribution and systematic error. It considers the stars as single ones and ignores the influence of binary stars on RV and attributes the variation in RV as uncertainties and therefore overestimates $\sigma_{\rm RV}$. The $\sigma_{\rm RV}$ has been reappraised in Paper I when estimating the binary fraction ($f_{\rm B}$) of dwarfs with S/N > 50, taking into account the degeneracy between $f_{\rm B}$ and $\sigma_{\rm RV}$. A comparison of the RV uncertainties from LSP3 ($\sigma_{\rm RV-LSP3}$) and those from Paper I ($\sigma_{\rm RV-I}$) is presented in Figure 2, which shows that the LSP3 pipeline overestimates the uncertainties of RVs. The median $\sigma_{\rm RV}$ of dwarfs with S/N > 50 is around 2.9 km s^{-1}, while for the LSP3 pipeline, it is ∼1.5 times higher at 4.3 km s^{-1}. The precision of RVs with high S/N is adequate enough to detect short-period binaries.

Figure 3 presents the distribution of mean $\sigma_{\rm RV}$ in the Hess diagram, which shows that $\sigma_{\rm RV}$ of hot stars are higher than those of cooler stars. The distribution of the average number of epochs in the Hess diagram is shown in Figure 4. The distribution of the multi-epoch observations are uniform, which indicates that the $\sigma_{\rm RV}$ are not biased by selection effects of epochs.

[11] http://dr4.lamost.org/v2/doc/vac

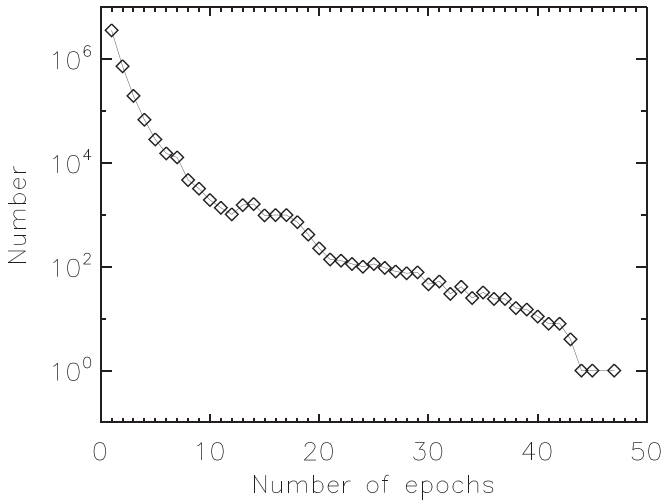

Figure 1. The number of stars against the number of epochs in LAMOST DR4.

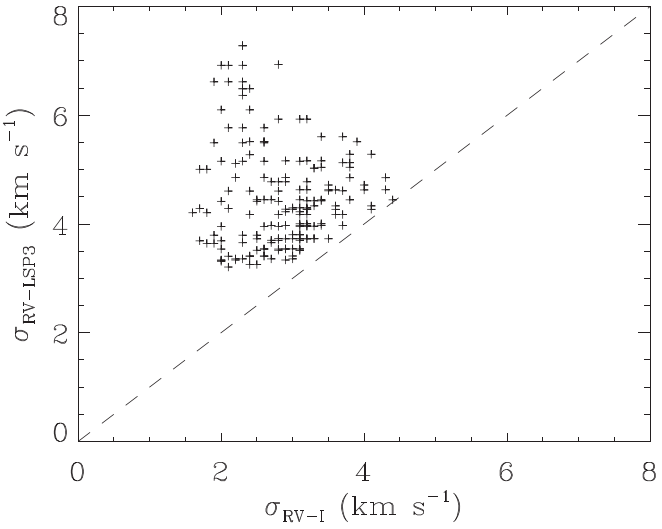

Figure 2. The comparison of $\sigma_{\rm RV}$ from LSP3 and Paper I.

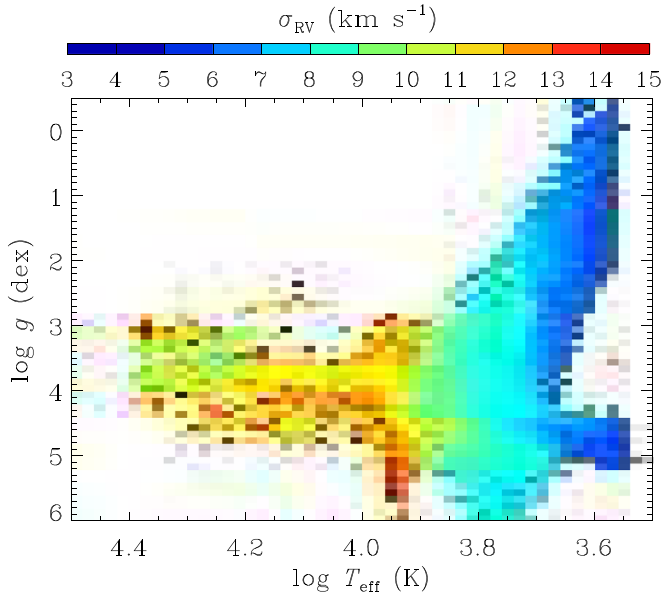

Figure 3. The mean $\sigma_{\rm RV}$ for stars in the Hess diagram.

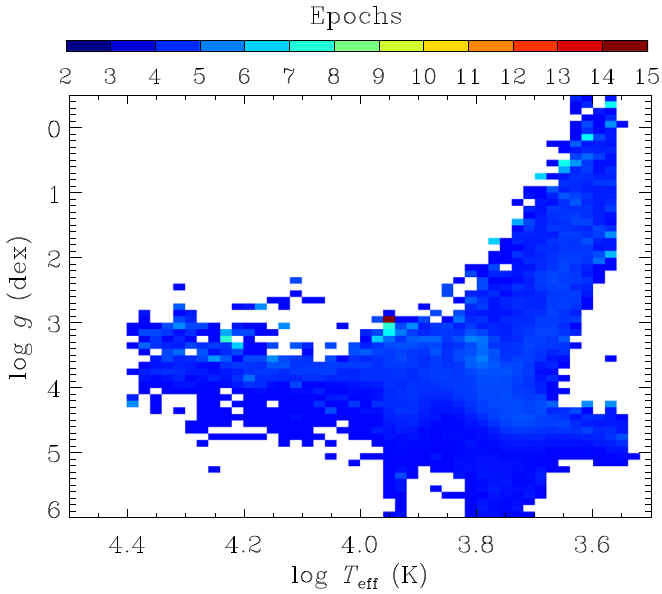

Figure 4. The mean number of epochs for stars in the Hess diagram.

2.2. Reliability of the Data

In this work, we adopt the RVs and $\sigma_{\rm RV}$ from LSP3 in our binary identification. For a single star with multiple observations in the same condition, the RVs obey a Gaussian distribution with a mean $\overline{\rm RV}$ and variance $\sigma^2_{\overline{\rm RV}}$. However, for a star observed repeatedly in different conditions, we have a sample of RVs for the star, $\rm RV_1$, $\rm RV_2$,..., $\rm RV_n$, where each RV value is from a Gaussian distribution with the same mean $\overline{\rm RV}$ but a different standard deviation $\sigma_{\rm RV_i}$. The weighting factor is the inverse of $\sigma^2_{\rm RV_i}$; thus, the weighted $\overline{\rm RV}$ is expressed as[12]

$$\overline{\rm RV} = \frac{\sum_{i=1}^{n} {\rm RV}_i / \sigma^2_{{\rm RV}_i}}{\sum_{i=1}^{n} 1/\sigma^2_{{\rm RV}_i}}, \quad (1)$$

where the error of the $\overline{\rm RV}$ is

$$\hat{\sigma}_{\overline{\rm RV}} = \sqrt{\frac{1}{\sum_{i=1}^{n} 1/\sigma^2_{{\rm RV}_i}}}, \quad (2)$$

and the weighted error is

$$\overline{\sigma_{\rm RV}} = \frac{\sum_{i=1}^{n} \sigma_{{\rm RV}_i} / \sigma^2_{{\rm RV}_i}}{\sum_{i=1}^{n} 1/\sigma^2_{{\rm RV}_i}}. \quad (3)$$

The variance of the RV is

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{({\rm RV}_i - \overline{\rm RV})^2}{\sigma^2_{{\rm RV}_i}}. \quad (4)$$

The distribution of S^2 for the sources with multiple epochs is shown in Figure 5. The S^2 converges into 1 with enough epochs, which proves the validity of RVs with errors. Although the RVs of a binary or other RV variable stars do not follow a normal distribution, we could also define their $\overline{\rm RV}$ and $\sigma_{\rm RV}$ through Equations (1) and (3).

[12] https://ned.ipac.caltech.edu/level5/Leo/Stats4_5.html

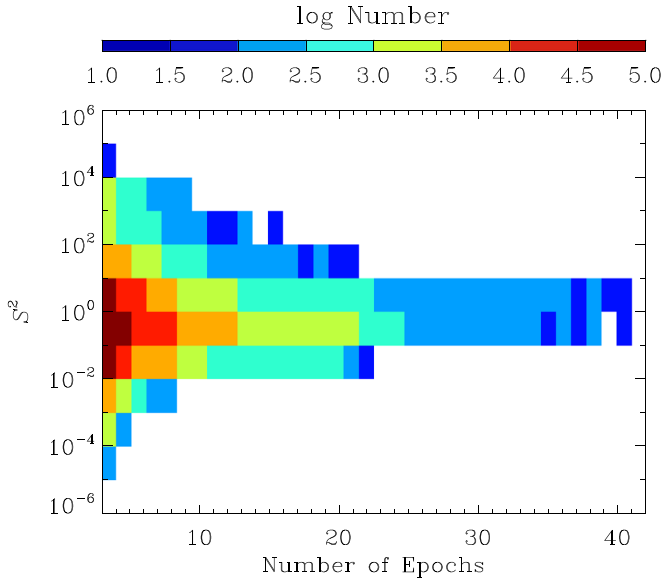

Figure 5. The distribution of the multi-epoch sources against the S^2 and the number of epochs.

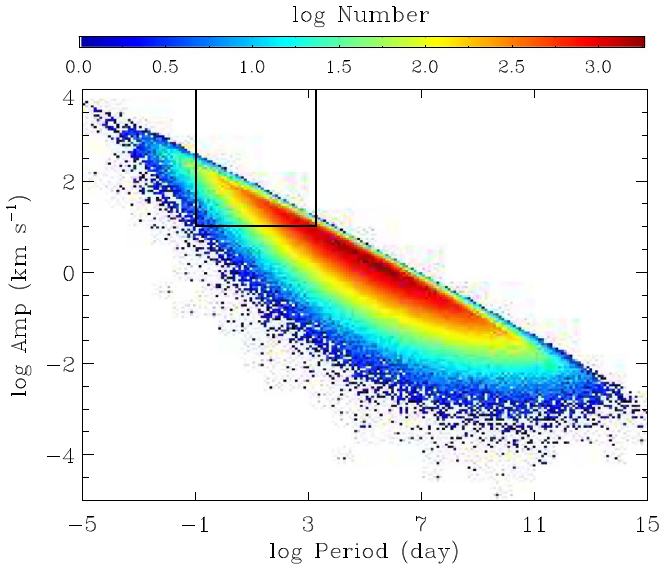

Figure 6. The joint distribution of periods and amplitudes for the mock binaries. The box marks out the detection limit based on the LAMOST's capability.

3. Method

3.1. Feasibility Analysis

In order to analyze the feasibility of detecting binaries through ΔRV_{max}, a simulation is performed. We construct a sample of 1 million binary stars and count the percentage of stars detected based on LAMOST's capability. For the binary systems M_B, we assume that: (1) the RVs are contributed by their primary stars; (2) their orbital orientations are isotropic in 3D space and initial phases follow a uniform distribution; (3) their primary masses follow the measured mass distribution of the LAMOST sample, which are determined by fitting the atmospheric parameters with the Yonsei-Yale isochrones (Demarque et al. 2004, and references therein); (4) the mass ratio q follows a power-law distribution ($f(q) \propto q^{0.3\pm0.1}$; e.g., Duchêne & Kraus 2013); and (5) for the orbital period distribution, a log-normal profile (with a mean value of $\log P = 5.03$ and a dispersion of $\sigma_{\log P} = 2.28$, where P is in units of days; see Raghavan et al. 2010) is adopted. The σ_{RVs} adopted in the simulation follow those derived from the LAMOST DR4 data. As shown in Figure 6, the amplitudes of the simulated binary stars are strongly dependent on the period distribution. Considering that the typical exposure time of each observation is about one hour and the time span of LAMOST DR4 is less than 5 yr, the detection is more efficient for binary systems with periods in the range of 0.1 day–5 yr rather than those with extremely short or long periods. We adopt 10 km s^{-1} ($\sim 3.0\sigma_{RV}$ for dwarfs with qS/Ns higher than 50) as a threshold of RV amplitude to recognize RV variable stars in the simulation. The box in the figure marks out the 12% of simulated binaries detectable with LAMOST based on these thresholds. It demonstrates that a certain proportion of binary stars are detectable based on the LAMOST observations.

3.2. Probability of Belonging to a Binary System

The binary system could be identified by comparing ΔRV_{max} with σ_{RV}, where the ΔRV_{max} presents the maximum RV difference between any two epochs for the same object (e.g., Maoz et al. 2012). In order to test the effectiveness of the method, we mock three samples and count the percentages of detected stars at different thresholds. The three samples are defined as:

a. the single stellar population (SSP) sample,
b. the binary stellar population (BSP) sample, and
c. the composite stellar population (CSP, composed of 45% single and 55% binary systems) sample.

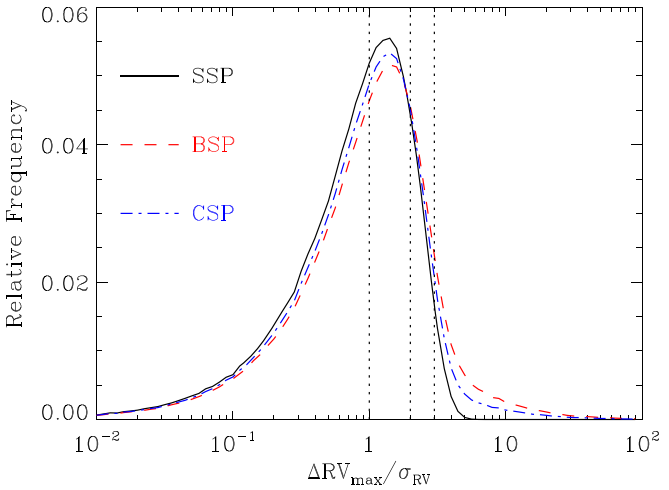

Figure 7. The histograms of $\Delta RV_{max}/\sigma_{RV}$ for the SSP, BSP, and CSP samples, respectively. The vertical dotted lines denote the thresholds of $\Delta RV_{max}/\sigma_{RV}$ equal to 1, 2, and 3, respectively.

The assumptions for the simulated samples are the same as those described in Section 3.1. The time separations of the multi-epoch observations are derived from the LAMOST DR4 data. The binary fraction of 55% adopted in the CSP is the median value derived from LAMOST (Gao et al. 2014; Yuan et al. 2015a; Tian et al. 2018). Each sample consists of 1 million stars or systems. Note that intrinsic variables, e.g., pulsating stars, are ignored in these simulations. Under these assumptions, the distributions of ΔRV_{max} for the SSP, BSP, and CSP samples are constructed and presented in Figure 7. The vertical dashed lines from left to right in the figure

Table 1
The DR, FPR, TPR, and Purity for the CSP when Adopting Different Cutoffs of $\Delta RV_{max}/\sigma_{RV}$

Rate	$\Delta RV_{max} > 1.0\sigma_{RV}$	$\Delta RV_{max} > 2.0\sigma_{RV}$	$\Delta RV_{max} > 3.0\sigma_{RV}$	$\Delta RV_{max} > 4.0\sigma_{RV}$
DR	0.515	0.206	0.076	0.036
FPR	0.479	0.157	0.034	0.005
TPR	0.545	0.246	0.110	0.062
purity	0.582	0.656	0.798	0.942

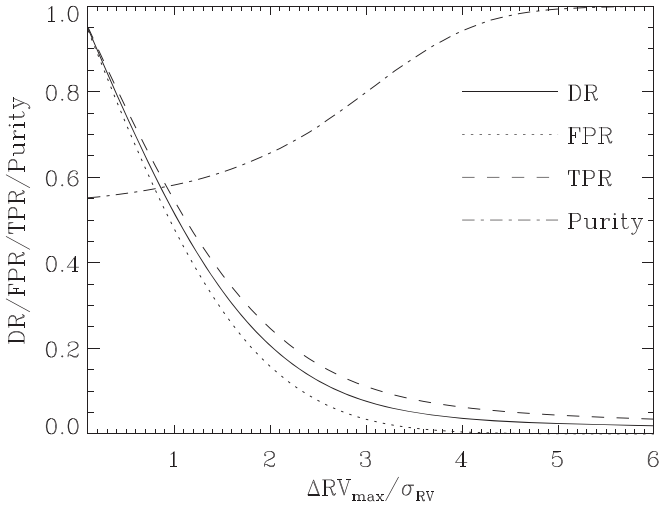

Figure 8. The DR, FPR, TPR, and purity against the cutoff of $\Delta RV_{max}/\sigma_{RV}$ for the CSP are plotted with solid, dotted, dashed, and dashed–dotted lines, respectively.

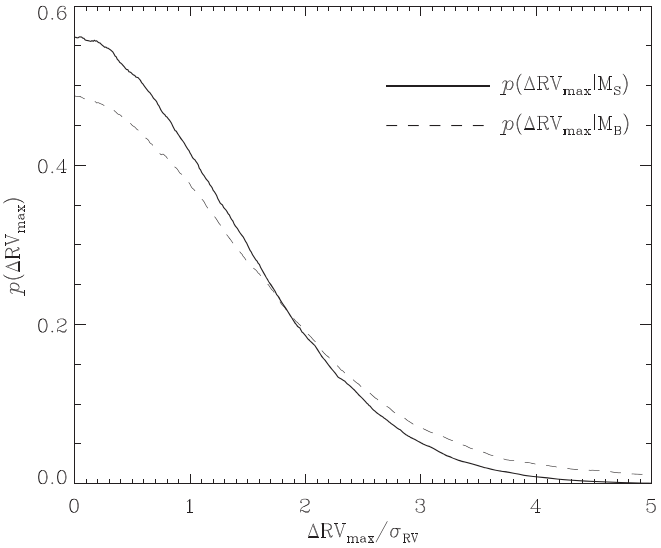

Figure 9. The probability of obtaining ΔRV_{max} based on the single (SSP) and binary (BSP) assumptions, respectively.

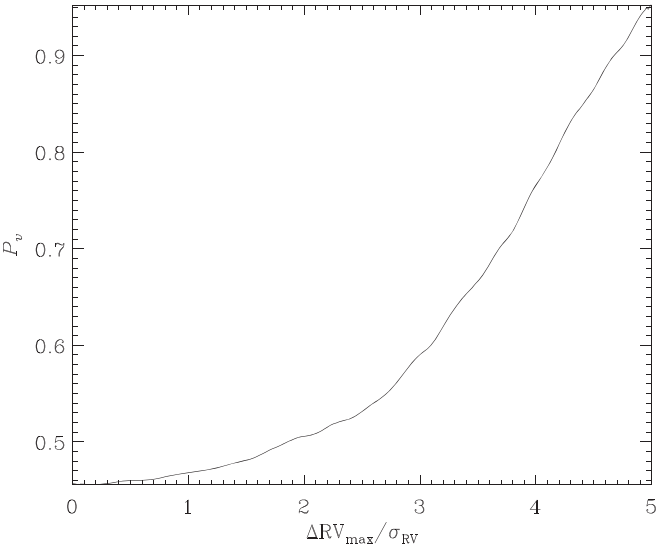

Figure 10. The estimated probability P_v of being a binary system based on the CSP with a binary fraction of 55%.

mark the cutoffs of $\Delta RV_{max}/\sigma_{RV}$ equal to 1, 2, and 3, respectively. The BSP sample has less low-value ΔRV_{max} and more high-value ΔRV_{max} than the SSP sample. The low-value ΔRV_{max} are dominated by random errors, while the high-value ΔRV_{max} are produced by variations of binary phases in the BSP (and the CSP). The detection rate (DR), false positive rate (FPR), true positive rate (TPR), and the fraction of real RV variables (purity) of the identified binaries against cutoffs of $\Delta RV_{max}/\sigma_{RV}$ are presented in Figure 8. Improving the threshold of $\Delta RV_{max}/\sigma_{RV}$ will increase the purity of the catalog but reduces the DR at the same time. The DR, FPR, TPF, and purity of the CSP with different cutoffs of $\Delta RV_{max}/\sigma RV$ are listed in Table 1. Here the threshold of $\Delta RV_{max} > 3.0\sigma_{RV}$ is adopted to identify RV variable stars. There are 3%, 11%, and 8% stars with ΔRV_{max} greater than $3.0\sigma_{RV}$ in the SSP, BSP, and CSP samples, respectively. The stars with $\Delta RV_{max}/\sigma_{RV} > 3.0$ in the CSP consist of 20% single stars and 80% binary systems. It indicates that the RV variable stars detected with the following method may be polluted by single stars.

Given the value of ΔRV_{max} from observations, the probability of the star being a binary could be calculated based on the CSP simulation using Bayes' theorem:

$$P_v = p(M_B|\Delta RV_{max}) = \frac{p(M_B \Delta RV_{max})}{p(\Delta RV_{max})} = \frac{p(\Delta RV_{max}|M_B)p(M_B)}{p(\Delta RV_{max}|M_B)p(M_B) + p(\Delta RV_{max}|M_S)p(M_S)}, \quad (5)$$

where $p(M_B)$ and $p(M_S)$ denote prior binary and single-star fractions, respectively. Here we adopt a $p(M_B)$ of 55% derived from LAMOST. The $p(\Delta RV_{max}|M_B)$ and $p(\Delta RV|M_S)$ indicate the probabilities of obtaining ΔRV_{max} based on assumptions of the BSP and SSP models, respectively. Their values as functions of $\Delta RV_{max}/\sigma_{RV}$ are shown in Figure 9. For stars with $\Delta RV/\sigma_{RV} < 1.9$, they are more likely to be a single star rather than a binary system. The probability of being a binary system P_v as a function of $\Delta RV_{max}/\sigma_{RV}$ calculated through Equation (5) is presented in Figure 10. The higher value of $\Delta RV_{max}/\sigma_{RV}$ is, the higher probability of the star belonging to a binary system.

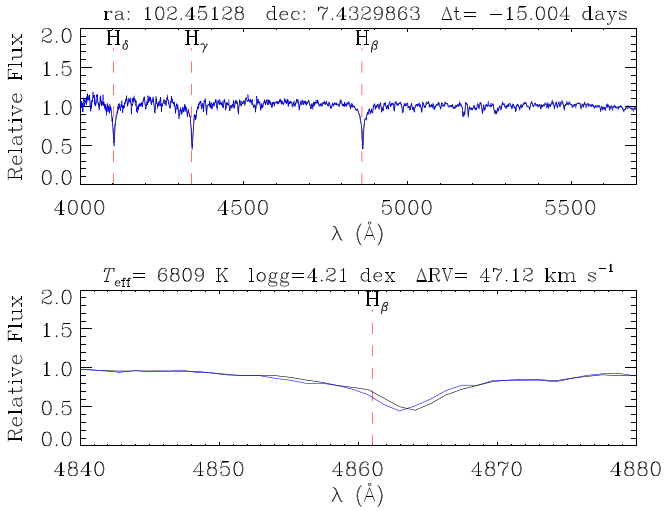

Figure 11. Normalized spectra from two epochs (black and blue lines) for a representative star. The rest wavelengths of Balmer lines are plotted with vertical dashed lines.

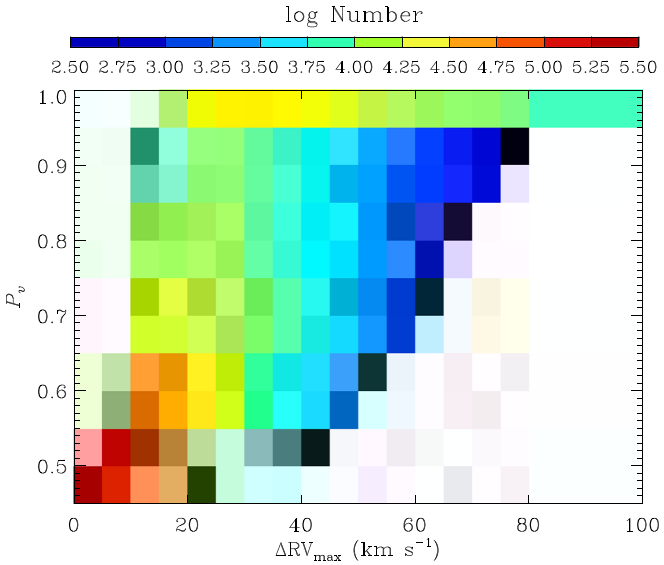

Figure 12. The distributions of stars against ΔRV_{max} and P_v.

This method is more sensitive to short-period binary stars, since their RVs vary more rapidly than long-period ones. For long-period (e.g., ∼300 yr) binary stars, the time span of the LAMOST DR4 observations (∼5 yr) is too short to produce a large enough ΔRV_{max} to test their binarity efficiently.

The Balmer lines are covered in the blue arm (3700–5900 Å) of LAMOST. Figure 11 plots the normalized LAMOST spectra for a representative star at two different epochs. The shift of H_β is clearly seen in the bottom panel, demonstrating LAMOST's capability to measure ΔRV_{max}. Here we measure the depths of H_β from the normalized spectra. In order to ensure the reliability of RV measurements, we eliminate the sources with H_β depths less than 0.3. Meanwhile, the sources with high ΔRV_{max} values are confirmed by visual inspections to identify and remove the spectra affected by cosmic rays.

4. Catalog of RV Variable Stars

We apply the method to the LAMOST (DR4) data and estimate the binary probabilities of stars. Here we adopt a threshold of $P_v > 0.6$ ($\Delta RV_{max} > 3.0\sigma_{RV}$) to identify binary stars and build a catalog of binary candidates. According to the simulation of CSP in Section 3.2, the FPR is about 3% at this threshold based on the capability of LAMOST. Since the cumulative run time of LAMOST is much less than the mean period of binary systems, the LAMOST data is not suitable for detecting long-period binaries. There are ∼120,000 stars with $P_v > 0.6$ ($\Delta RV_{max} > 3.0\sigma_{RV}$) in the LAMOST's DR4 sources with multiple epochs. After adopting the criteria of spectral depth and visual inspections, an assemblage of 80,702 RV variable star candidates remains in our final catalog as listed in Table 2. Note that in the simulation we only consider single and binary stars, but the sample observed with LAMOST includes some intrinsic variables such as pulsating stars.

The distribution of the repeatedly observed stars in two-dimensional space of ΔRV_{max} versus P_v is shown in Figure 12. The majority of the repeated targets that dominate low P_v values ($P_v < 0.6$) are single stars or unrecognized RV variables. Meanwhile, we present the fraction f_v of stars with $P_v > 0.6$ in each bin with a size of 0.02 dex by 0.2 dex for logT_{eff} and log g, respectively, in Figure 13. As shown in the figure, the extended

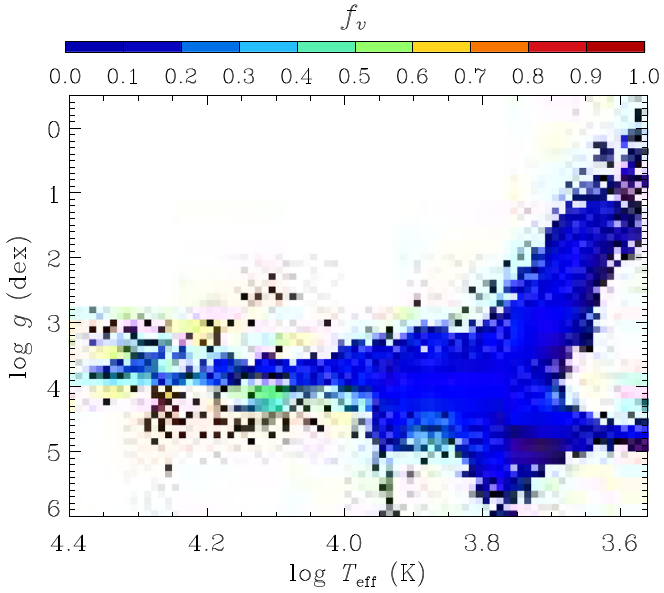

Figure 13. The fractions f_v of stars with $P_v > 0.6$ in each bin of the Hess diagram.

distribution of main-sequence stars with $P_v > 0.6$ is broader than those with $P_v < 0.6$. Stars with high P_v have higher probabilities of being binaries than those with low P_v.

4.1. The Purity of the Catalog

In order to verify the purity of the catalog and estimate pollutions by single stars, we perform a cross-identification between the LAMOST multi-epoch sources and a catalog of RV standard stars published by Huang et al. (2018) based on the APOGEE data (Majewski et al. 2017; Abolfathi et al. 2018). There are 1274 common sources between them. Among the common sources, 103 RV standard stars have $P_v > 0.6$. It means a single-star contribution of ∼8% to our catalog. The purity of our catalog is approximately 92%, which agrees with the simulation in Section 3.2. Considering the cross-identification between our catalog and Huang et al. (2018), as well as the

Table 2
Catalog of RV Variable Star Candidates

No.	R.A.	Decl.	Epochs	Time Duration	ΔRV_{max}	$\overline{\sigma_{RV}}$	$S/N_{RV_{max}}$	$S/N_{RV_{min}}$	$t_{RV_{max}}$	$t_{RV_{min}}$	P_v	Classification	Notes[a]
1	0.0170327	56.0176468	2	707.0498657	40.3	7.5	11	21	2456968.093	2456261.043	0.95		
2	0.0204036	61.5655899	2	414.8555298	27.2	8.4	77	39	2457324.099	2456909.243	0.62		
3	0.0236430	60.1553001	2	412.9315491	37.2	11.4	12	25	2457322.097	2456909.166	0.63		
4	0.0246262	62.7630920	2	412.8926392	18.1	5.4	16	32	2457322.097	2456909.204	0.65		
5	0.0312140	35.5037842	2	1029.2077637	15.2	4.4	70	183	2456262.009	2457291.217	0.65		
6	0.0470053	61.6638145	2	412.9263611	17.2	4.2	95	27	2456909.204	2457322.131	0.76		
7	0.0742439	60.3994408	2	412.9653015	41.7	7.1	16	124	2457322.131	2456909.166	0.95		
8	0.0804520	37.1159744	2	674.0565796	25.5	4.3	28	89	2456262.037	2456936.094	0.95		
9	0.0865299	36.5521774	2	748.9668579	16.1	4.0	19	96	2456261.981	2457010.948	0.76		
10	0.0902442	55.8368454	2	28.9211788	31.0	7.3	53	91	2456968.049	2456996.970	0.82		
...	...	...	...	...	...	...	...	...	...	...	...	...	

Notes. The R.A. and decl. of the stars are listed in columns 2–3. The number of epochs and time duration of observations for each star are shown in columns 4–5. The maximum variations of RV and the weighted errors are listed in columns 6–7. S/Ns and time of exposures responding to the maximum and minimum RVs are listed in columns 8–11. The probability of being a RV variable star is provided in the last column. For the LAMOST unique spectral ID, S/N, time for each exposure, stellar parameters, and RVs, together with their errors of each epoch, see a detailed and inclusive version of the catalog online.

[a] The notes column marks the common sources between LAMOST and other surveys.

(This table is available in its entirety in FITS format.)

pollution by single stars in the simulation from Section 3.2, the purity of our catalog is estimated to be higher than ∼80%.

4.2. Crossmatch with Kepler Eclipsing Binaries

A database of thousands of Kepler eclipsing binaries (KEBs) is released by Matijevič et al. (2012, and references therein). In total 520 KEBs have been observed repeatedly by the LAMOST–Kepler project that uses LAMOST to make spectroscopic follow-up observations for the Kepler targets (De Cat et al. 2015; Zong et al. 2018). Of those, 255 stars are detected as binary stars in our catalog based on the LAMOST observations. To test the rationality of such application on the Kepler data, we simulate a sample of 1 million eclipsing stars and count the rate of the detectable binaries. The assumptions of the mock sample are similar to those described in Section 3.2. However, for the simulated eclipsing stars, we fix the inclination of their orbits as $\pi/2$. The distribution of orbital periods for the mock sample is adopted from those of the KEBs. The joint distribution of periods and ΔRV_{max} for the mock eclipsing binaries is shown in Figure 14. The box in the figure marks out the detectable stars with periods in the range of 0.1 day–5 yr and RV amplitudes higher than 10 km s^{-1}. About 60% of the eclipsing binaries are detected in the simulation. The DR will be reduced to 44% given the limitation of periods of 0.5 day–5 yr. The simulation provides an explanation for the detection ratio of ∼50% of KEBs by LAMOST.

KEBs such as KIC 11084782 and KIC 9953894 have been observed in 11 and 7 epochs by LAMOST, respectively. Their RV time series are plotted in the top panels of Figures 15 and 16. Given the orbital period measured with Kepler, we could fit the RVs of the binary system accurately with `rvfit`. The `rvfit` method fits RVs of stellar binaries and exoplanets using an adaptive simulated annealing (ASA) global minimization method, which quickly converges to a global solution minimum without the need to provide preliminary parameter values. The efficiency and reliability have been verified by Iglesias-Marzoa et al. (2015a, 2015b). As shown in the middle panels of Figures 15 and 16, the observed and fitted RVs against phases are presented. The residuals ($O - C$) are plotted in the bottom panels of the figures. The RVs from spectroscopic observations together with periods from photometric observations could constrain the orbital parameters well.

4.3. Crossmatch with GDR2 Variables

Since some stars exhibit RV variations due to periodic contraction and expansion, they will, absent of further characterization, contaminate the catalog of binary candidates. We crossmatch the variable star candidates with GDR2 variables including Cepheids, RR Lyrae, long-period variables (LPV), and short-period variables (SPV; Mowlavi et al. 2018; Roelens et al. 2018; Clementini et al. 2019). The distribution of P_v for the common stars is presented in Figure 17. From the 498 common sources, 198 variable stars are detected ($P_v > 0.6$) with LAMOST. The common sources include 19 Cepheids, 442 RR Lyrae, 34 LPV, and 3 SPV detected with Gaia. Among them, 10 Cepheids, 179 RR Lyrae, 4 LPV, and 0 SPV are identified as RV variables in our database. The TPR of the catalog is about 39% for these intrinsic variables. This value is

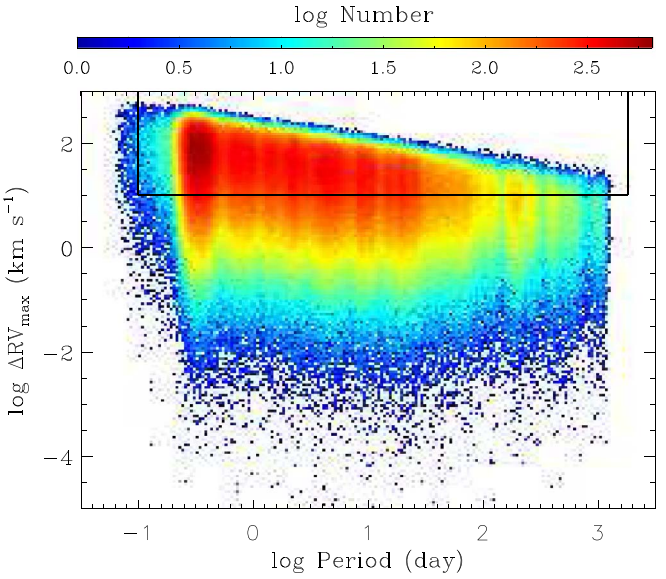

Figure 14. The joint distribution of periods and ΔRV_{max} for the mock eclipsing binaries. The box marks out the detection limit based on LAMOST's capability.

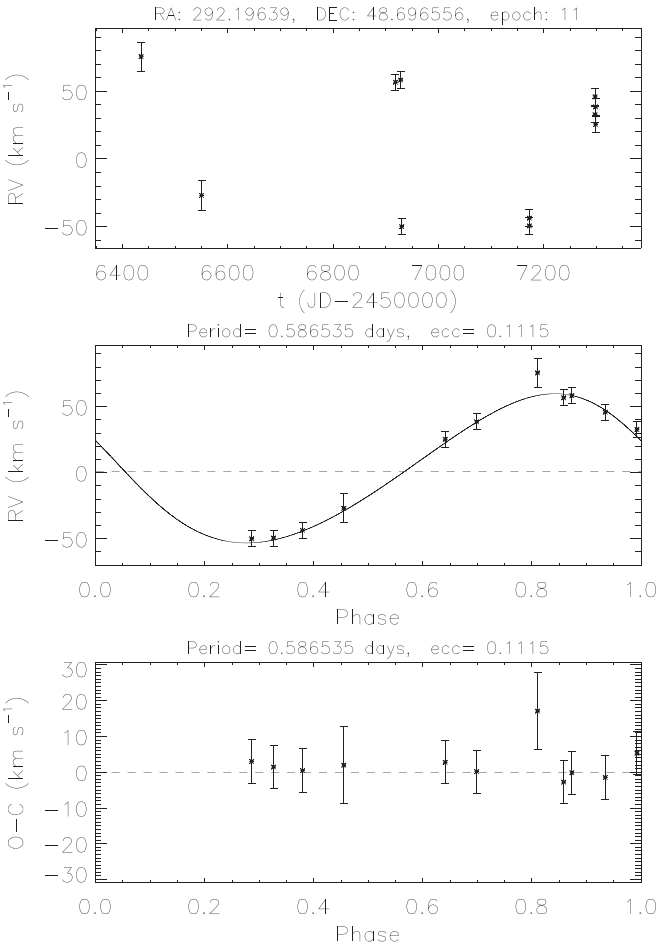

Figure 15. The comparison of RVs between observations and fittings for KIC 11084782. The top panel shows the observed RVs against time. The observed (dots with error bars) and fitted (solid line) RVs against phases are plotted in the middle panel, and the residuals ($O - C$) are plotted in the bottom panel.

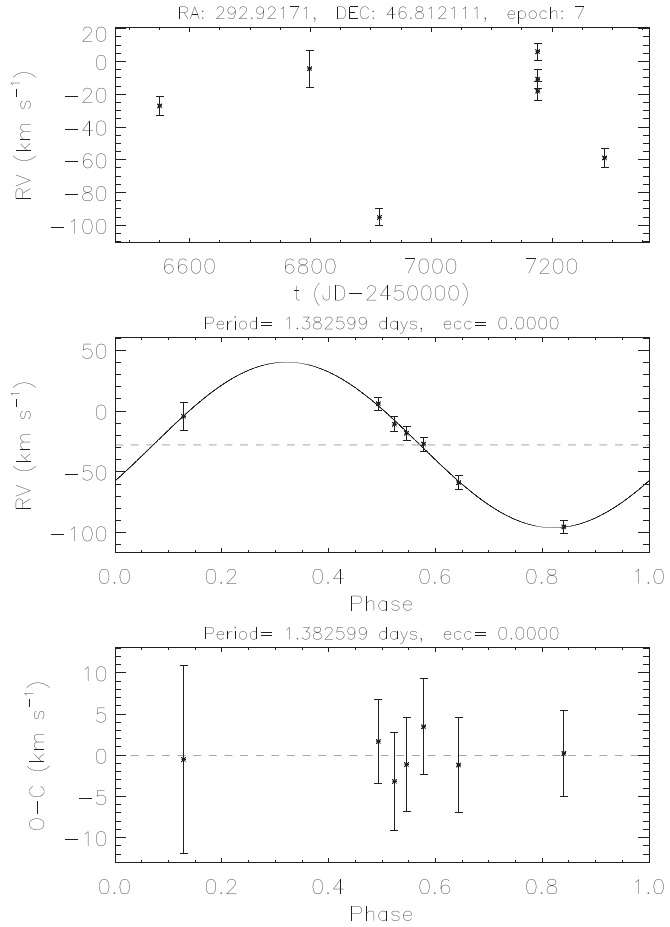

Figure 16. Same as Figure 15 but for KIC 9953894.

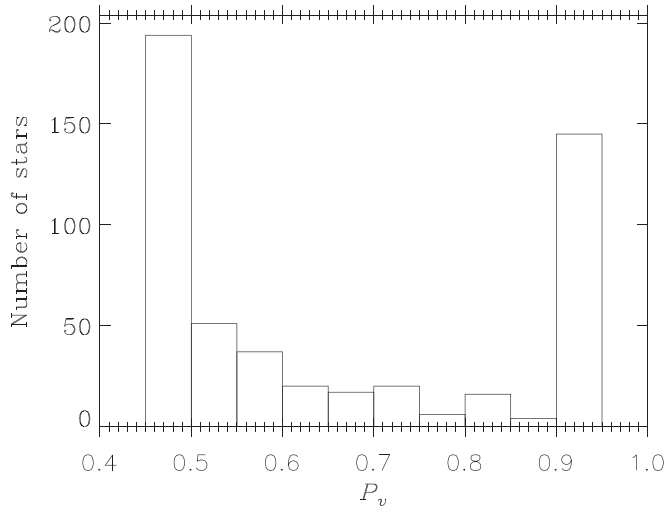

Figure 17. P_v distribution of common stars between GDR2 variables and multi-epoch observed LAMOST targets.

different than that of binary systems because of the different period distributions between intrinsic and extrinsic variables. The period–Δt diagrams for these common Cepheids and RR Lyrae are presented in Figures 18 and 19, respectively. Their

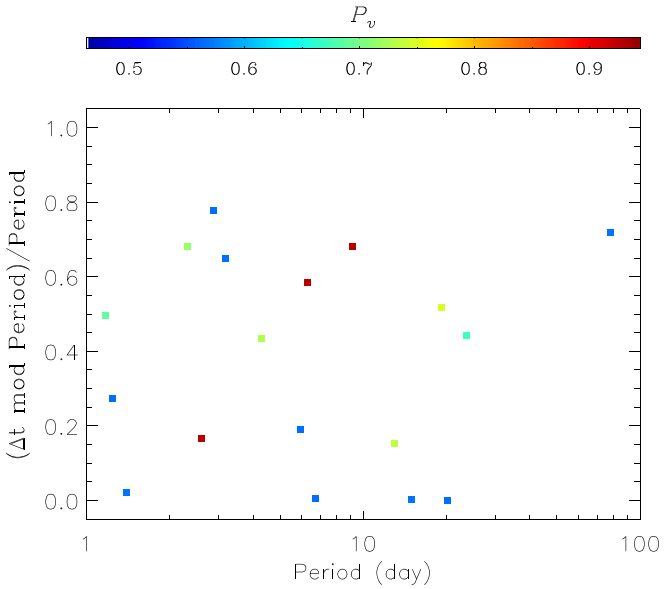

Figure 18. P_v on the period–Δt diagram for Cepheids. The colors of points denote the probabilities of RV variable stars.

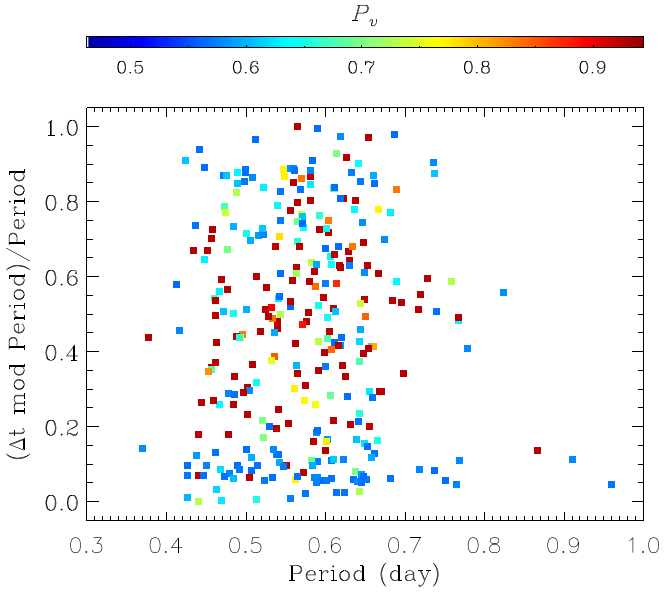

Figure 19. Same as Figure 18 but for RR Lyrae.

periods are provided by Gaia variable catalogs, while the Δt are from LAMOST observations. From the figures, we can see that the DRs are related to sampling characteristics of observations as well as stellar periods. Figure 20 quantifies the distribution of DR f_v against the (Δt mod period)/period for the common RR Lyrae between GDR2 variables and LAMOST multi-epoch targets. A Gaussian curve of the f_v with a mean of 0.48 and variance of 0.29^2 illustrates the DR depends on sampling characteristics of observations and stellar periods.

Meanwhile, we crossmatch the LAMOST multi-epoch targets with the catalog of RV standard stars from GDR2 (Soubiran et al. 2018). None of the seven common stars were identified as RV variables in our catalog.

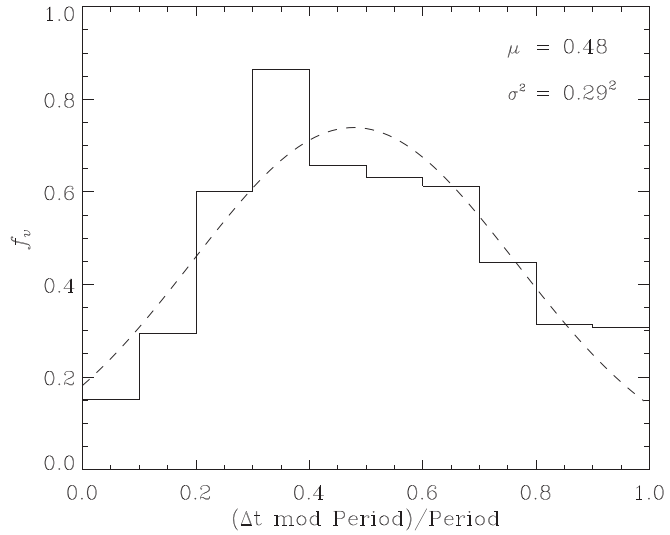

Figure 20. The distribution of detection rate f_v against the (Δt mod period)/period for the common RR Lyrae between GDR2 variables and LAMOST duplicated targets. The solid and dashed lines denote the calculated and Gaussian-fitted values of f_v, respectively. The mean and variance of the Gaussian distribution are shown in the figure.

As examples indicated in Figure 21, the phased variations of RV, $T_{\rm eff}$, and log g for RR Lyrae are presented. Their periods are measured with Gaia, and stellar parameters and RVs are derived from LAMOST spectra. Since the pulsation of an RR Lyra, its log g varies with the radius directly; meanwhile, its $T_{\rm eff}$ decreases and increases with the contraction and expansion of the star, respectively. The variations of stellar parameters and RVs could be detected through the LAMOST observations. A detailed analysis of RR Lyrae observed with LAMOST is presented in Liu et al. (2020) and interested readers are referred to that paper.

Note that the LAMOST is not adequate to detect short-period RV variables with periods shorter than two hours based on Nyquist's theorem, especially for extreme short-period ones, since the typical exposure time of LAMOST is in the order of an hour. We list the three common sources between Gaia SPV and multi-epoch observed LAMOST targets in Table 3. From the table, we can see that low-period (high-frequency) SPV could not be detected as variables with LAMOST. It demonstrates that the probability of detection is related to the period (or frequency) of the target.

4.4. Crossmatch with VSX

In order to investigate our catalog further, we crossmatch the catalog with other variable stars published in the literature. The VSX is a comprehensive relational database of known and suspected variable stars gathered from a variety of respected published sources (Watson et al. 2006). About 600,000 variable stars are collected and about 75% of them are provided with types and periods in VSX. There are 10,557 shared sources between VSX and LAMOST duplicated targets. Among them, 3044 stars are detected as RV variables in our catalog. The types of the detected stars include binary stars and pulsating stars. The comprehensive DR of VSX is about 29% by LAMOST.

4.5. Crossmatch with GCVS

The GCVS is another catalog of variable stars. The GCVS Version 5.1 contains data for 53,626 individual variable stars discovered and named as variable stars by 2017 and located mainly in the Galaxy (Version 5.1; Samus et al. 2017). An assemblage of 33,264 variables is provided with types and periods in GCVS 5.1. Among 924 common sources between GCVS 5.1 and LAMOST multi-epoch sources, 453 stars are recognized as RV variables in our catalog. The comprehensive DR of GCVS is about 49% by LAMOST.

4.6. Crossmatch with ASAS-SN

The All-Sky Automated Survey for SuperNovae (ASAS-SN) scans the extragalactic sky visible from Hawaii roughly once every five nights in the V band (Shappee et al. 2014). Catalogs of variable stars based on ASAS-SN have been released by Jayasinghe et al. (2018, 2019a, 2019b). These catalogs collect 542,526 variable stars, including 334,095 supplied with types and periods. There are 5113 common sources between the ASAS-SN variable catalogs and LAMOST multi-epoch targets. Among them, 2011 stars are recognized as RV variables in our catalog. The comprehensive DR of ASAS-SN variables is about 39% by LAMOST.

4.7. Characteristics of the Catalog

A summary of the numbers of common sources between the published catalogs and LAMOST multi-epoch targets are listed in Table 4. Note that some variable stars are identified repeatedly in different published catalogs. There are 11,035 common sources between LAMOST multi-epoch targets and the referred variable catalogs such as KEBs, GDR2 variables, VSX, GCVS, and ASAS-SN variable catalogs. There are 3163 common sources detected as RV variables in our catalog. The DR of our catalog is 29% for the variables published in the referenced catalogs.

Variable stars fall into two categories: intrinsic and extrinsic variables. Binaries belonging to extrinsic variables and pulsating stars from intrinsic ones could be detected through variations of RVs based on the LAMOST's capability. There are 80,702 stars detected as RV variables among the 818,136 stars with multiple epochs by LAMOST. As discussed in Sections 4.2 and 4.3, not only binaries are included in the catalog but also some intrinsic variables such as RR Lyrae and Cepheids. According to the CSP simulation in Section 3.2, about 8% of the sample are detected as binaries with a purity of 80%, which implies that 6.4% of the LAMOST targets with multiple epochs are binaries and 1.6% (∼13,000) are pollution by single stars given the LAMOST multi-epoch targets consist of single and binary stars. However, the 80,702 detected stars dominate about 10% of the LAMOST targets with multiple epochs, which is higher than the DR about 8% in the CSP simulation. Consequently, the others (15,251 stars) in the catalog, dominating approximately 2% of the LAMOST multi-epoch sources, probably consist of some intrinsic variables and pollution by single stars. Applying the curve of pulsating star fractions against $T_{\rm eff}$ (see Figure 11 in Murphy et al. 2019) in the LAMOST targets with multiple epochs, the number of pulsating stars covered by LAMOST is expected to be approximately 20,000. However, only pulsating stars with periods and RV amplitudes in a specified range could be detected by LAMOST. Assuming a typical DR of 30% of the

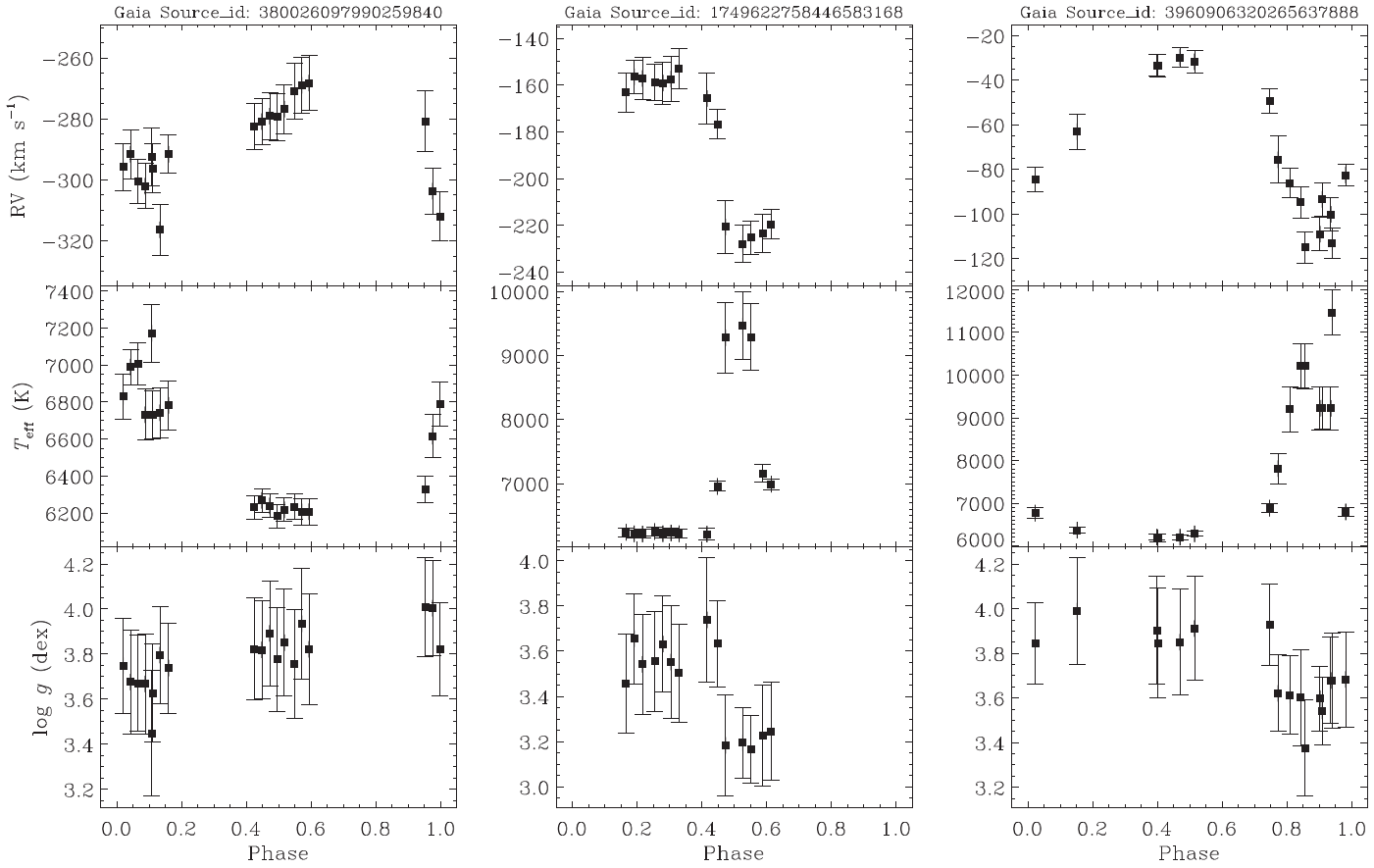

Figure 21. Phased variations of RV, $T_{\rm eff}$, and log g for examples of the RR Lyrae stars.

Table 3
Common Sources of LAMOST and Gaia SPV

No.	R.A.	Decl.	Epochs	ΔRV_{max} (km s^{-1})	$\overline{\sigma_{RV}}$ (km s^{-1})	P_v	Amplitude (mag)	Frequency (d^{-1})
1	119.46654	20.091425	2	4.88	5.79	0.46	0.22	9.30
2	205.83937	20.921980	2	22.83	14.85	0.48	0.68	43.20
3	200.50029	18.452419	2	35.28	11.54	0.58	0.94	2.03

Table 4
Numbers of Common Sources and Detected Sources by LAMOST

Catalog	Common Sources with LAMOST	Detected by LAMOST	DR
KEBs	520	255	0.49
GDR2	495	190	0.38
VSX	10,557	3044	0.29
GCVS	924	453	0.49
ASAS-SN	5113	2011	0.39

pulsators, the number of detected pulsating stars in our catalog is approximately 6000. Thus, the 15,251 stars are mainly constructed with binary stars and pulsating stars, probably. Therefore, the catalog consists of ∼62,000 binaries (77%), ∼6000 pulsating stars (7%), and pollution by ∼13,000 single stars (16%).

Based on the BSP simulation in Section 3.2, the DRs of binaries against their periods are presented in Figure 22. The DRs drop exponentially with the increasing of periods. Figure 23 displays the DR of common sources between LAMOST repeated targets and the published catalogs mentioned before. The classifications of the shared stars through cross-identifications with the previous catalogs are listed in Table 2. The distribution of DR indicates that the method adopted in this work based on ΔRV_{max} by LAMOST is sensitive to short-period RV variable stars such as short-period binaries and RR Lyrae. All the same, various types of variable stars appear in our catalog. However, most of the variables collected in our catalog, so far, are not able to be classified based on LAMOST spectra or data from other surveys.

5. Conclusions and Discussions

We analyze the probabilities of being RV variable stars based on the duplicated observations for LAMOST DR4 targets. A catalog of 80,702 RV variable star candidates is constructed. The FPR of the catalog is about 3% based on the LAMOST ability. The purity of the catalog is estimated to be better than ∼80% through simulation and cross-identifications. Both intrinsic and extrinsic variable stars are collected in the catalog. It consists of 77% binary systems and 7% pulsating stars as well as 16% pollution by single stars. The catalog is a powerful database of RV variable candidates, which could be taken as an input source for RV variable surveys.

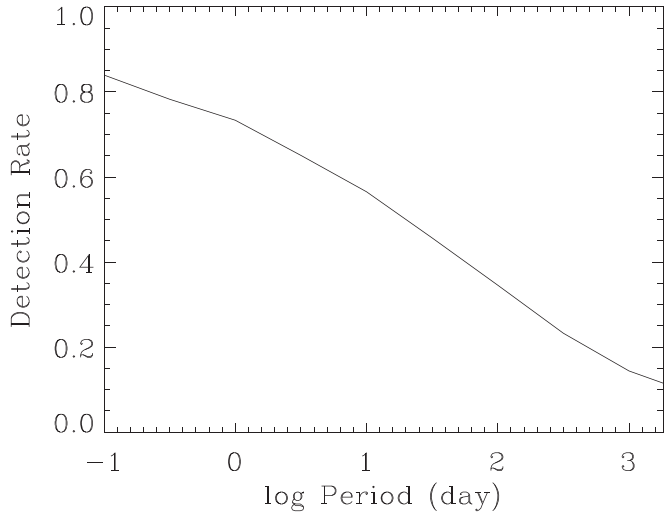

Figure 22. The DR verses orbital periods for the BSP simulation.

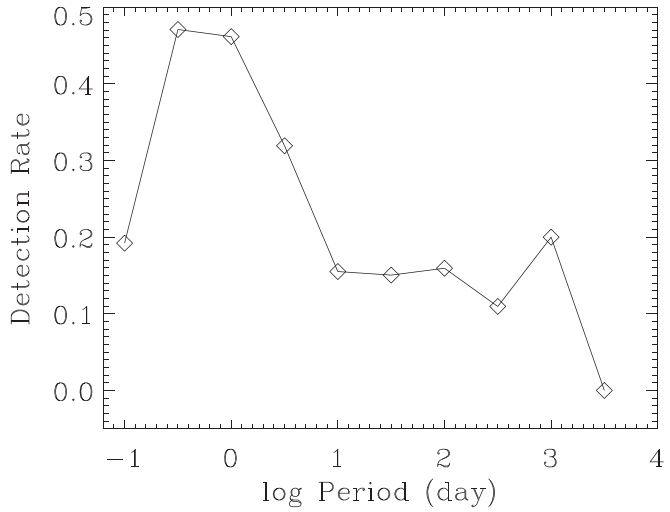

Figure 23. The DR of RV variables by LAMOST for the common sources between LAMOST multi-epoch targets and published catalogs.

Since some intrinsic variables present variability of RV, the catalog is blended with pulsating stars such as Cepheids, RR Lyrae, LPVs, and SPVs. The cross-identifications and classifications are carried out by matching with KEBs, GDR2 variables, VSX, GCVS, and ASAS-SN variables. A number of 3138 stars in our catalog are classified. Although recognized as RV variables, most of the variable stars in the catalog are not classified based on the LAMOST data or other surveys. The efficiency of the method adopted in this work relies on not only sampling frequency of observations but also on periods and amplitudes of variable stars.

The key foundation of this work is the accuracy of RVs and their uncertainties. Fortunately, overestimating uncertainties will not affect the accuracy of identifying RV variables or their candidates, although some of them would be left out. In future work, we will make use of spectral and photometric data from LAMOST and other surveys to classify the catalog of RV variable stars as a follow-up to this work. The spectra of classified stars would probably be adopted as training set to recognize spectra of unclassified RV variables based on a machine-learning method. Meanwhile, the common sources between the RV variables and X-ray sources will provide more clues about binary interactions.

This work has made use of data products from the Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST). 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 is partially supported by National Natural Science Foundation of China 11803030, 11443006, 11773005, and 11803029; National Key Basic Research Program of China 2014CB845700; China Postdoctoral Science Foundation 2016M600850; Science & Technology Department of Yunnan Province–Yunnan University Joint Funding 2019FY003005; and Joint Research Fund in Astronomy U1531244 and U1631236. The LAMOST FELLOWSHIP is supported by Special Funding for Advanced Users, budgeted and administrated by Center for Astronomical Mega-Science, Chinese Academy of Sciences (CAMS).

We would like to thank the anonymous referee for valuable comments that improved the manuscript.

ORCID iDs

Zhijia Tian https://orcid.org/0000-0003-0220-7112
Xuan Fang https://orcid.org/0000-0002-3981-7355
Maosheng Xiang https://orcid.org/0000-0002-5818-8769
Yang Huang https://orcid.org/0000-0003-3250-2876
Shaolan Bi https://orcid.org/0000-0002-7642-7583
Wuming Yang https://orcid.org/0000-0002-3956-8061
Yaqian Wu https://orcid.org/0000-0002-8337-4117
Huawei Zhang https://orcid.org/0000-0002-7727-1699
Yong Yang https://orcid.org/0000-0002-4028-1893
Jincheng Guo https://orcid.org/0000-0002-8321-1676

References

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42
Badenes, C., Mazzola, C., Thompson, T. A., et al. 2018, ApJ, 854, 147
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Clementini, G., Ripepi, V., Molinaro, R., et al. 2019, A&A, 622, A60
Coronado, J., Sepúlveda, M. P., Gould, A., & Chanamé, J. 2018, MNRAS, 480, 4302
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
De Cat, P., Fu, J. N., Ren, A. B., et al. 2015, ApJS, 220, 19
Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, ApJS, 155, 667
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 735
Drake, A. J., Beshore, E., Djorgovski, S. G., et al. 2012, AAS Meeting, 219, 428.20
Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, ApJS, 213, 9
Duchêne, G., & Kraus, A. 2013, ARA&A, 51, 269
Duquennoy, A., & Mayor, M. 1991, A&A, 248, 485
El-Badry, K., Ting, Y.-S., Rix, H.-W., et al. 2018, MNRAS, 476, 528
Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gao, S., Liu, C., Zhang, X., et al. 2014, ApJL, 788, L37
Gao, S., Zhao, H., Yang, H., & Gao, R. 2017, MNRAS, 469, L68
Hettinger, T., Badenes, C., Strader, J., Bickerton, S. J., & Beers, T. C. 2015, ApJL, 806, L2
Huang, Y., Liu, X.-W., Chen, B.-Q., et al. 2018, AJ, 156, 90
Iglesias-Marzoa, R., López-Morales, M., & Jesús Arévalo Morales, M. 2015a, rvfit: Radial Velocity Curves Fitting for Binary Stars or Exoplanets, Astrophysics Source Code Library, ascl:1505.020

Iglesias-Marzoa, R., López-Morales, M., & Jesús Arévalo Morales, M. 2015b, PASP, 127, 567
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, MNRAS, 477, 3145
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019a, MNRAS, 486, 1907
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019b, MNRAS, 485, 961
Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJL, 713, L79
Kovaleva, D., Kaygorodov, P., Malkov, O., Debray, B., & Oblak, E. 2015, A&C, 11, 119
Kroupa, P., & Petr-Gotzens, M. G. 2011, A&A, 529, A92
Liu, G. C., Huang, Y., Zhang, H. W., et al. 2020, ApJS, 247, 68
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. (Cambrridge: Cambridge Univ. Press), 310
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Maoz, D., Badenes, C., & Bickerton, S. J. 2012, ApJ, 751, 143
Matijevič, G., Prša, A., Orosz, J. A., et al. 2012, AJ, 143, 123
Maxted, P. F. L., Heber, U., Marsh, T. R., & North, R. C. 2001, MNRAS, 326, 1391
Moe, M., & di Stefano, R. 2017, ApJS, 230, 15
Mowlavi, N., Lecoeur-Taïbi, I., Lebzelter, T., et al. 2018, A&A, 618, A58
Murphy, S. J., Hey, D., van Reeth, T., & Bedding, T. R. 2019, MNRAS, 485, 2380
Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2018, AJ, 156, 18
Prugniel, P., & Soubiran, C. 2001, A&A, 369, 1048
Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, arXiv:astro-ph/0703658
Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, ApJS, 190, 1
Rimoldini, L., Holl, B., Audard, M., et al. 2019, A&A, 625, A97
Roelens, M., Eyer, L., Mowlavi, N., et al. 2018, A&A, 620, A197
Samus', N. N., Kazarovets, E. V., Durlevich, O. V., Kireeva, N. N., & Pastukhova, E. N. 2017, ARep, 61, 80
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Soubiran, C., Jasniewicz, G., Chemin, L., et al. 2018, A&A, 616, A7
Tian, Z.-J., Liu, X.-W., Yuan, H.-B., et al. 2018, RAA, 18, 052
Wang, C., Liu, X.-W., Huang, Y., et al. 2018, MNRAS, 480, 4766
Watson, C. L., Henden, A. A., & Price, A. 2006, SASS, 25, 47
Xiang, M.-S., Liu, X.-W., Shi, J.-R., et al. 2017, MNRAS, 464, 3657
Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, MNRAS, 448, 822
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
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
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723
Zong, W., Fu, J.-N., De Cat, P., et al. 2018, ApJS, 238, 30