A&A 662, A66 (2022)
https://doi.org/10.1051/0004-6361/202141570

Astronomy & Astrophysics

Stellar labels for hot stars from low-resolution spectra

I. The HotPayne method and results for 330 000 stars from LAMOST DR6⋆

Maosheng Xiang[1], Hans-Walter Rix[1], Yuan-Sen Ting[2,3,4,5,⋆⋆], Rolf-Peter Kudritzki[6,7], Charlie Conroy[8], Eleonora Zari[1], Jian-Rong Shi[9,10], Norbert Przybilla[11], Maria Ramirez-Tannus[1], Andrew Tkachenko[12], Sarah Gebruers[12], and Xiao-Wei Liu[13]

[1] Max-Planck-Institut für Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
e-mail: 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] LMU München, Universitätatssternwarte, Scheinerstr. 1, 81679 München, Germany
[7] Institute for Astronomy, University of Hawaii at Manoa, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[8] Center for Astrophysics, Harvard & Smithsonian, Cambridge, MA 02138, USA
[9] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, PR China
[10] University of Chinese Academy of Sciences, Beijing 100049, PR China
[11] Institut für Astro- und Teilchenphysik, Universität Innsbruck, Technikerstrasse 25, 6020 Innsbruck, Austria
[12] Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
[13] South-Western Institute for Astronomy Research, Yunnan University, Kunming 650500, PR China

Received 16 June 2021 / Accepted 19 February 2022

ABSTRACT

We set out to determine stellar labels from low-resolution survey spectra of hot stars, specifically OBA stars with $T_{\rm eff} \gtrsim 7500$ K. This fills a gap in the scientific analysis of large spectroscopic stellar surveys such as LAMOST, which offers spectra for millions of stars at $R \sim 1800$ and covers $3800\,\text{Å} \leq \lambda \leq 9000\,\text{Å}$. We first explore the theoretical information content of such spectra to determine stellar labels via the Cramér-Rao bound. We show that in the limit of perfect model spectra and observed spectra with signal-to-noise ratio ∼50−100, precise estimates are possible for a wide range of stellar labels: not only the effective temperature, $T_{\rm eff}$, surface gravity, $\log g$, and projected rotation velocity, $v\sin i$, but also the micro-turbulence velocity, $v_{\rm mic}$, helium abundance, $N_{\rm He}/N_{\rm tot}$, and the elemental abundances [C/H], [N/H], [O/H], [Si/H], [S/H], and [Fe/H]. Our analysis illustrates that the temperature regime of $T_{\rm eff} \sim 9500$ K is challenging as the dominant Balmer and Paschen line strengths vary little with $T_{\rm eff}$. We implement the simultaneous fitting of these 11 stellar labels to LAMOST hot-star spectra using the Payne approach, drawing on Kurucz's ATLAS12/SYNTHE local thermodynamic equilibrium spectra as the underlying models. We then obtain stellar parameter estimates for a sample of about 330 000 hot stars with LAMOST spectra, an increase by about two orders of magnitude in sample size. Among them, about 260 000 have good *Gaia* parallaxes ($\omega/\sigma_\omega > 5$), and their luminosities imply that ≳95% of them are luminous stars, mostly on the main sequence; the rest are evolved lower luminosity stars, such as hot subdwarfs and white dwarfs. We show that the fidelity of the results, particularly for the abundance estimates, is limited by the systematics of the underlying models as they do not account for nonlocal thermodynamic equilibrium effects. Finally, we show the detailed distribution of $v\sin i$ of stars with 8000−15 000 K, illustrating that it extends to a sharp cutoff at the critical rotation velocity, $v_{\rm crit}$, across a wide range of temperatures.

Key words. techniques: spectroscopic – surveys – catalogs – stars: massive – stars: fundamental parameters – stars: abundances

1. Introduction

Hot stars, here meaning stars with OBA spectral types or $T_{\rm eff} \gtrsim 7500$ K, are interesting objects for a number of reasons: they are tracers of young stellar populations in our Galaxy; they are intriguing laboratories for testing stellar structure and evolution scenarios; and they can be progenitors of and companions to black holes and supernovae. Recently, these stars have garnered more attention thanks to large sky low-resolution ($R \sim 2000$) spectroscopic surveys, such as the LAMOST Galactic surveys (Deng et al. 2012; Zhao et al. 2012) and the SDSS-V Milky Way Mapper (Kollmeier et al. 2017; Zari et al. 2021). LAMOST has collected $R \sim 1800$ optical spectra for thousands of stars identified as OB stars (Liu et al. 2019) and hundreds of thousands of A-type stars. The SDSS-V will yield spectra for many more OB stars in the next few years (Zari et al. 2021). Due to their high effective temperature, the spectra of hot stars show much weaker spectral features than those of cool stars. As a result, the latter have been the analytical focus of most Galactic archaeology surveys to date. Stellar parameter pipelines for large spectroscopic

⋆ The catalog is only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/662/A66
⋆⋆ Hubble fellow.

surveys have been mainly focused on FGK stars (e.g., Lee et al. 2008; Luo et al. 2015; Xiang et al. 2015, 2019; García Pérez et al. 2016; Casey et al. 2017; Ahumada et al. 2020; Buder et al. 2021; Steinmetz et al. 2020). Beside the easier measurements, the detailed abundances [X/H] of cool stars have been of particular interest as they are closer to the birth-material composition of the stars and can enable "chemical tagging". Furthermore, it has been shown that even low-resolution spectra can yield useful stellar labels, such as $T_{\rm eff}$, $\log g$, and abundance [X/H] for >15 elements (Ting et al. 2017a,b; Xiang et al. 2019).

For hot stars, the surface abundances can be affected by stellar internal mixing processes, making them deviate significantly from the solar abundance scale (e.g., Przybilla et al. 2010; Ekström et al. 2012; Maeder et al. 2014; Martins et al. 2015; Pedersen et al. 2018; Aerts 2021) and rendering them less useful for inferences in Galactic archaeology. On the other hand, these mixing processes can make hot stars powerful probes of stellar evolution physics. Particularly, it has long been known that "chemically peculiar" stars are common among hot stars (e.g., Renson & Manfroid 2009; Gray et al. 2016; Qin et al. 2019; Chojnowski et al. 2020; Xiang et al. 2020; Paunzen et al. 2021) and provide good constraints on stellar element transport processes (e.g., Talon et al. 2006; Vick et al. 2010; Michaud et al. 2015). Despite their interesting prospects, the parameters and detailed abundances of hot stars were only derived for relatively small samples, usually no larger than ~O(10^3) stars (e.g., Przybilla et al. 2008; Nieva & Przybilla 2012, 2014; Simón-Díaz & Herrero 2014; Kudritzki et al. 2016; Berger et al. 2018; Holgado et al. 2018).

The lack of a systematic analysis of hot-star spectra calls for flexible and sophisticated spectral modeling to deliver astrophysical labels for hot stars from the vast sets of low-resolution spectra collected by large sky surveys. This is particularly important for the LAMOST Galactic Survey as it has collected low-resolution spectra ($R \sim 1800$) for hundreds of thousands of OBA stars. Yet, a comprehensive and quantitative determination of stellar parameters and abundances for all these hot stars with LAMOST spectra does not exist, and we aim to provide it here.

In this work we demonstrate that $R \sim 1800$ spectra contain a wealth of astrophysical information for hot stars. We apply PAYNE (Ting et al. 2019) for an "industrial scale" determination of stellar labels from a massive set of LAMOST hot-star spectra. PAYNE is a spectral model fitting tool with the ability to simultaneously determine many labels from the full spectra at modest computational expense. This is owed to the flexible neural-net-based spectral interpolation algorithm that can precisely interpolate spectra in high dimensions (Ting et al. 2019). Fitting the entire spectrum allows for an optimal use of its information content, and fitting all pertinent labels simultaneously ensures that label covariances, due to line blending, are correctly accounted for. We fit a large number (11) of labels to LAMOST spectra of probable hot stars – effective temperature ($T_{\rm eff}$), surface gravity ($\log g$), microturbulence velocity ($v_{\rm mic}$), projected rotation velocity ($v\sin i$), helium abundance ($N_{\rm He}/N_{\rm tot}$), and elemental abundances [C/H], [N/H], [O/H], [Si/H], [S/H], and [Fe/H]. We dub our modeling HOTPAYNE as the original PAYNE approach has been mostly applied to FGK stars; hot-star spectra are sufficiently different, so this spectral regime requires nontrivial modifications and testing, as we will show.

We adopt the Kurucz local thermodynamic equilibrium (LTE) spectra (Kurucz 1970, 1993, 2005) as our underlying model spectra to train the HOTPAYNE neural network spectral model. As nonlocal thermodynamic equilibrium (NLTE) effects and stellar winds play important roles for hot-star spectroscopy (e.g., Auer & Mihalas 1972; Kudritzki 1976, 1979, 1998; Hubeny & Lanz 1995; Kudritzki & Puls 2000; Lanz & Hubeny 2007; Przybilla et al. 2008, 2010; Hainich et al. 2019), our resultant label estimates, based on the Kurucz LTE model spectra, are therefore inevitably affected by non-negligible systematic errors. We address such systematics by comparing our results with literature values from high-resolution NLTE spectroscopic analyses for a sample of reference stars. We emphasize from the outset that the approach presented here deserves to be followed up with label determinations based on NLTE model spectra to obtain accurate label estimates, in particular of elemental abundances for these hot stars, which we will perform in a future study. Nonetheless, the internal precision of our LTE-based abundance estimates appears to be promising for a number of interesting applications, as presented in this study.

The paper is organized as follows. In Sect. 2 we introduce the LAMOST spectra for the sample of candidate hot stars. In Sect. 3 we explore the information content of such low-resolution spectra of hot stars. In Sect. 4 we derive stellar labels from the $R \sim 1800$ LAMOST spectra for >330 000 hot stars. We present and discuss these results in Sect. 5, which is followed by a summary in Sect. 6.

2. LAMOST spectra of candidate hot stars

We compiled a sample of candidate hot stars with low-resolution ($R \sim 1800$) optical ($\lambda \sim 3800$–9000) spectra from the sixth data release (DR6)[1] of the LAMOST Galactic surveys (Deng et al. 2012; Zhao et al. 2012; Liu et al. 2014); we refer to this initial list as candidates, as our subsequent spectral analysis reveals that some of them have low temperatures of $T_{\rm eff} < 7000$ K. The LAMOST Galactic surveys target millions of stars in the full color-magnitude diagram down to about 17.8 mag in the SDSS r band (Carlin et al. 2012; Liu et al. 2014; Yuan et al. 2015). As a key component of the LAMOST Galactic surveys, the LAMOST Spectroscopic Survey at the Galactic Anti-Center (LSS-GAC) uniformly samples stars in the (g–r, r) and ($r - i$, r) space. This uniform selection implies that stars in the low density regimes of the color-magnitude diagram (e.g., OB stars) are targeted with higher completeness (Liu et al. 2014; Yuan et al. 2015; Xiang et al. 2017; Chen et al. 2018), which results in a large sample of hot-star spectra with well-characterized completeness, located mostly in the Galactic anticenter direction. The LAMOST DR6 released 9 911 337 spectra, and each has a spectral classification with the LAMOST 1D pipeline (e.g., Luo et al. 2015). Among the 9 911 337 spectra, 9 231 057 are classified as stars.

As summarized in Table 1, we define the hot-star candidates by the following three criteria. First, we adopted LAMOST DR6 stars that are classified as O, B, and A types according to the LAMOST pipeline. This is a total of 575 736 spectra.

Second, we cross-matched the LAMOST DR6 with a full-sky hot-star catalog (candidate OBA stars), devised using the same approach as in Zari et al. (2021), which is based on parallax, magnitudes, and colors from *Gaia* DR2 (Gaia Collaboration 2018) and the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006). The key difference is that Zari et al. (2021) based their OBA star catalog on *Gaia* Early Data Release 3 (eDR3; Gaia Collaboration 2021) and selected hot stars with a 2MASS K_s-band absolute magnitudes $M_{\rm K} < 0$ mag. For our study, we adopted *Gaia* DR2 (consistent with LAMOST DR6) and have

[1] dr6.lamost.org

Table 1. Sources of the LAMOST DR6 hot-star candidates.

Sources	Number of spectra
Stars classified as O,B,A type by LAMOST pipeline	575 736
Gaia DR2 selection with Zari et al. (2021)'s method	866 780
OB star catalog of Liu et al. (2019)	26 453
Total	1 163 410 (844 790 unique stars)

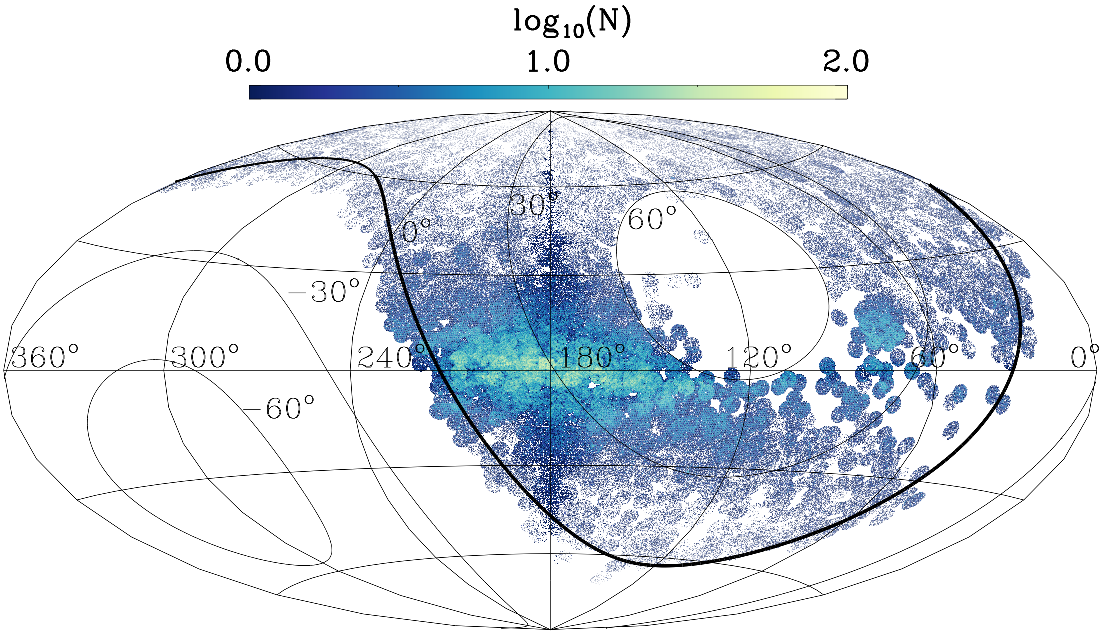

Fig. 1. On-sky distribution of our LAMOST DR6 hot-star candidates in Galactic coordinates (l, b). The map is centered on the Galactic anticenter ($l = 180°$, $b = 0°$). The thick line delineates the celestial equator (Dec = 0°), and lines of constant declination, Dec = ±30° and Dec = ±60°, are also shown. Colors indicate the logarithmic number of stars in each (l, b) cell of $0.2° \times 0.2°$. The distribution reflects both the LAMOST survey area and the candidate stars' concentration to the Galactic plane, which is expected for young stars.

selected hot stars with $M_K < 3$ mag. The cross-match led to 866 780 spectra in common.

Third, we adopted the LAMOST DR5 OB star catalog by Liu et al. (2019). It contains 26 453 spectra also found in the LAMOST DR6 database.

About 35% of the spectra for the candidate stars selected with the Zari et al. (2021) method are classified as OBA types by the LAMOST pipeline. This number is reasonable as we set a loose cut in absolute magnitude by using $M_K < 3$ mag. Utilizing a stricter cut of $M_K < 0$ mag, Zari et al. (2021) suggested that about 45% of their hot-star candidates with LAMOST spectra have effective temperature higher than 9700 K, corresponding to a stellar type earlier than A0V. Equivalently, about 52% of the OBA-type spectra classified by the LAMOST pipeline are selected as hot-star candidates with the Zari et al. (2021) method. We note that the criteria of Zari et al. (2021) are mainly set to select hot stars earlier than B7V type, for which they have achieved a high completeness rate (8282/9083 when testing on the LAMOST OB stars). As here we apply the method to the O-, B-, and A-type stars, it is not surprising that we find a smaller completeness rate.

A simple union of these three source lists led to 1 163 410 spectra for hot-star candidates. These are for 844 790 unique stars, as we keep the repeated observations in the database. Figure 1 shows the distribution of these stars in Galactic coordinates, along with the LAMOST survey footprint. Unsurprisingly, these hot-star candidates are concentrated toward the Galactic plane, which is expected since most hot stars are young stars. Many of the stars at high Galactic latitudes turn out to be hot old stars, such as hot subdwarfs (sdBs, sdOs) and blue horizontal branch (BHB) stars. The majority of the stars are toward the Galactic anticenter, as targeted by the LSS-GAC. There are also a moderate number of stars in the first and second Galactic quadrants toward the inner Galaxy, particularly in the *Kepler* field. Figure 2 shows the distribution of the *Gaia* G-band magnitude and the LAMOST spectral signal-to-noise ratio (S/N) for our sample of candidate stars. The G−band magnitudes cover a dynamic range of nearly 15 mag, and most of them are between ∼10 and 18 mag, with a peak at around 14 mag. The S/N peaks at about 63 ($\log(S/N) \sim 1.8$), but the median value is about 37 due to a significant low-S/N tail.

We corrected for the line-of-sight velocity derived from the LAMOST pipeline (Luo et al. 2015). All model fitting in this study will act on the normalized spectra. Specifically, all spectra were normalized by a smoothed version of themselves, derived by convolving the spectra with a Gaussian kernel of 50 Å width:

$$\bar{f}(\lambda_i) = \frac{\Sigma_j f_j \omega_j(\lambda_i)}{\Sigma_j \omega_j(\lambda_i)}, \quad (1)$$

$$\omega_j(\lambda_i) = \exp\left(-\frac{(\lambda_j - \lambda_i)^2}{L^2}\right), L = 50\ \text{Å}. \quad (2)$$

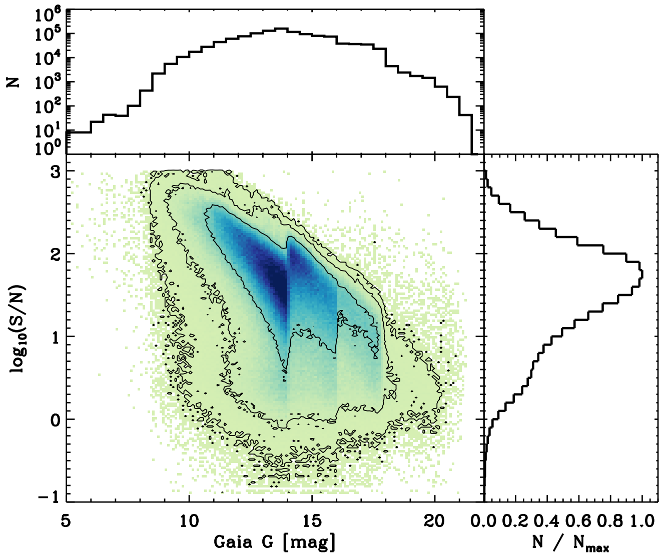

Fig. 2. Distribution of our LAMOST DR6 hot-star candidates in the *Gaia G*-band magnitude and LAMOST spectral S/N plane. The G−band magnitudes cover a dynamic range of nearly 15 magnitudes, and most of them are between $G \sim 10$ and 18 mag. The spectral S/N peaks at around 63 ($\log(S/N) \sim 1.8$) but has a median value of 37 due to a significant low-S/N tail. The isodensity contours encompass 68.3, 95, and 99.7% of the sample stars, respectively, and the colors delineate the relative stellar number density, with darker colors representing higher density.

A similar spectral normalization approach was previously adopted in the analysis of the LAMOST low-resolution spectra (Ho et al. 2017; Xiang et al. 2019). It should be noted that this is different to the "continuum" normalization applied in high-resolution spectroscopic analysis. Rather than determining the real continuum, which is hard for the case of low-resolution spectra, the purpose of our spectral normalization is to ensure that the observation spectra and the model spectra are normalized in the same manner for spectral fitting. As the spectra are normalized to the mean flux smoothed in a broad window, the method is robust for spectra of low S/N. However, a mismatch between the model spectra and observation spectra could be raised from either systematics of the former (Sect. 4) or interstellar absorption lines that presented solely in the latter. Such a mismatch will cause line profiles that differ between the normalized model and the observed spectra, causing systematic mismatch in the spectral fitting. To minimize this effect, we masked the hydrogen lines along with some other strong lines and the bands of interstellar absorption, by setting the weight of those pixels to be zero. By doing so, the resulting line profiles are consistent between models and observations, except for the cores of strong lines that suffer large model systematics.

3. The stellar label information content of low-resolution hot-star spectra

Understanding what parameters and abundances we should adopt for our spectral grid requires us to evaluate the information content of these spectra: that is, for which labels the spectra vary appreciably when labels vary and how covariate they are. One way to answer these questions is via calculating the Cramér-Rao (CR) bound (Cramér 1945; Rao 1945). The same approach was demonstrated by Ting et al. (2016, 2017a) and Sandford et al. (2020), who applied the CR bound to predict the (theoretical) precision of stellar label determination of FGK stars for different spectral resolutions and wavelength coverage.

The theoretical precision limit for a label estimate from a given spectrum depends on two aspects: (a) the gradient function of the spectra $\nabla_l := \partial f_\lambda / \partial l$, which describes the response of the spectral flux f_λ with respect to the change of stellar label l, and (b) the flux uncertainties and covariances of the spectra. Following Ting et al. (2017a), the CR bound thus can be described as

$$K_{ij}^{-1} = \nabla_i f(\lambda_1) C_{\lambda_1,\lambda_2}^{-1} \nabla_j f(\lambda_2), \quad (3)$$

where K_{ij}^{-1} is the inverse of the covariance matrix of the stellar labels, C^{-1} the inverse of the covariance of the normalized spectral flux, $\nabla_i f(\lambda_1)$ the gradient spectra with respect to label i at λ_1, and $\nabla_j f(\lambda_2)$ the gradient spectra with respect to label j at λ_2. The gradient spectra are calculated through finite model differencing.

As an example, Fig. 3 shows the gradient spectra $\partial f_\lambda / \partial l$ computed for a fiducial B star with $T_{\rm eff} = 18\,000$ K, $\log g = 4.0$, and $[{\rm Fe/H}] = 0$. The figure illustrates that hot stars are information rich. Besides the stellar parameters, there are numerous pixels in the LAMOST wavelength range (3800–9000 Å) that are informative for a few selected elements. Unsurprisingly, the most informative pixels for $T_{\rm eff}$ and $\log g$ are the hydrogen lines, both Balmer and Paschen series, with moderate contributions from helium lines and metal lines, whereas the information for $v \sin i$ resides mostly at the cores of strong absorption lines. We also note that the spectral features do not necessarily need to come from the atomic transition of a species. For some elements, such as helium, the change of abundance can modify the overall atmosphere structure appreciably. This will in turn indirectly modify the strength of spectral features.

Although the figure demonstrates that many pixels contain valuable information for various labels, nonetheless the absolute changes for individual pixels are small in most cases: $|\partial f_\lambda / \partial l|$ changes rarely more than 1% even for abundance changes of a full *dex*. These subtle and mostly covariate changes of spectral features are what inspire us to adopt a technique for full spectral fitting, such as PAYNE, in the first place (see Sect. 4.2).

A realistic calculation of CR bound requires not only gradient spectra but also the assumed uncertainties of the spectral fluxes. For simplicity, for this theoretical exercise, we assume that the spectral measurements are independent among the adjacent pixels. We adopted a S/N variation with wavelength that is typical of B stars in LAMOST and scale that value to $S/N = 100$ at 4650 Å. Some assumptions of the CR bound calculation are undoubtedly optimistic. For example, the values and uncertainties among adjacent pixels in an observed spectrum can be correlated. Furthermore, the calculation assumes a perfect knowledge of the spectral models. Experience has shown that CR bound calculations can usually be attained within a factor of two despite these imperfections (e.g., Ting et al. 2019; Sandford et al. 2020). We emphasize that this exercise here aims only to determine what labels we should adopt for the spectral grid.

The CR bound, expressed by the diagonal terms of the CR matrix $\sqrt{K_{ii}}$, is shown in Fig. 4 for our fiducial B star spectrum. The calculation predicts that, in the limit of spectral perfect knowledge, we should obtain measurements with precision of ~300 K for $T_{\rm eff}$, 0.03 dex for $\log g$, 1% for $N_{\rm He}/N_{\rm tot}$, 1 km s^{-1} for $v_{\rm mic}$, and 8 km s^{-1} for $v \sin i$ from a typical LAMOST spectrum for a hot star. The CR bounds for element abundance are within 0.1 dex for a few elements with strong features, such as C, O, Si, and Fe. For other elements, such as N, Mg, and Al, the CR

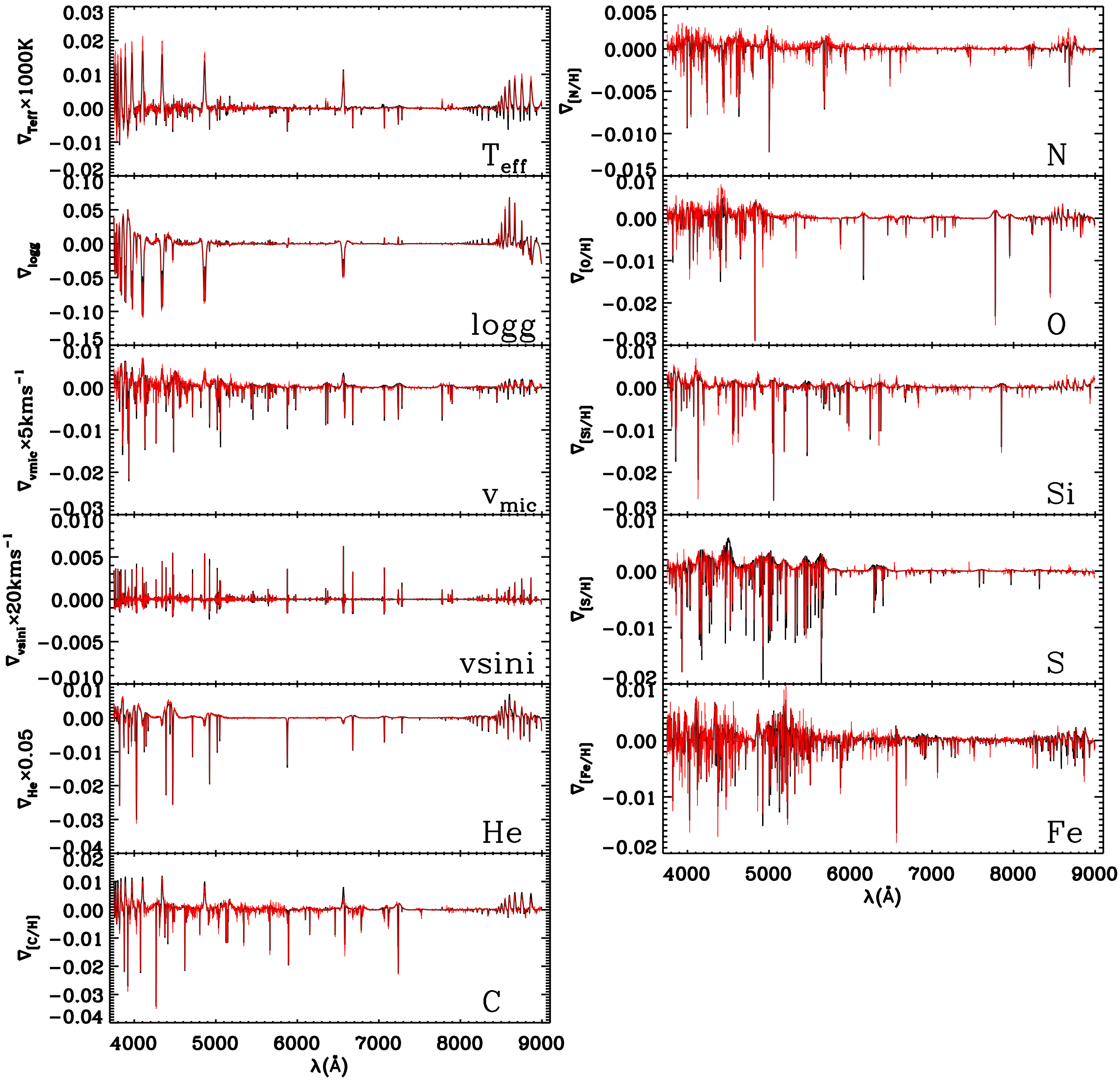

Fig. 3. Information content of low-resolution spectra for a B star. Each panel shows the gradient spectrum $\nabla_l := \partial f_\lambda / \partial l$ with respect to different labels, l, derived from finite differencing of Kurucz model spectra (black) at $T_{\rm eff} = 18\,000$ K, $\log g = 4.0$, [Fe/H] = 0, and $v\sin i = 100$ km s^{-1}. The gradient spectra generated from HotPayne are overplotted in red, illustrating the robustness of HotPayne: the good agreement to the Kurucz model suggests that HotPayne determines the labels from the relevant features instead of harnessing the astrophysical correlations. For $T_{\rm eff}$, $v_{\rm mic}$, $v\sin i$, and $N_{\rm He}/N_{\rm tot}$, the gradient spectra are multiplied by 1000 K in $T_{\rm eff}$, 5 km s^{-1} in $v_{\rm mic}$, 20 km s^{-1} in $v\sin i$, and 0.05 in $N_{\rm He}/N_{\rm tot}$ for better visibility. This figure shows that the diagnostic signatures of label changes are widely distributed across the optical spectral range. Nonetheless, most spectral features are subtle due to LAMOST's spectral resolution and the high effective temperature of the star: most abundance signatures are at the 1% level, even for abundance changes of 1 dex. Useful information about hot stars can only be extracted from a global fit to the spectrum and a careful spectral modeling.

bounds are within 0.2 dex. Although the figure only presents the CR bounds for some selected elements (with atomic number no larger than Zn), all elements with atomic number smaller than Eu, except for Li, Be, B, F, V, and Co, are included for the CR bound computation. We omit all CR bounds that are larger than 10 dex. The CR bounds for Li, Be, B, F, V, and Co are discarded because their null gradients could cause numerical artifacts in the results. Finally, we note that here we adopt a B star as reference point for this study; the CR bound will obviously vary among different stellar types.

Figure 5 presents the covariances among the selected labels, reflecting in the off-diagonal term of the CR matrix K_{ij}. As expected, the $T_{\rm eff}$ and $\log g$ are strongly correlated, as their gradient spectra share many common spectral features (H and He

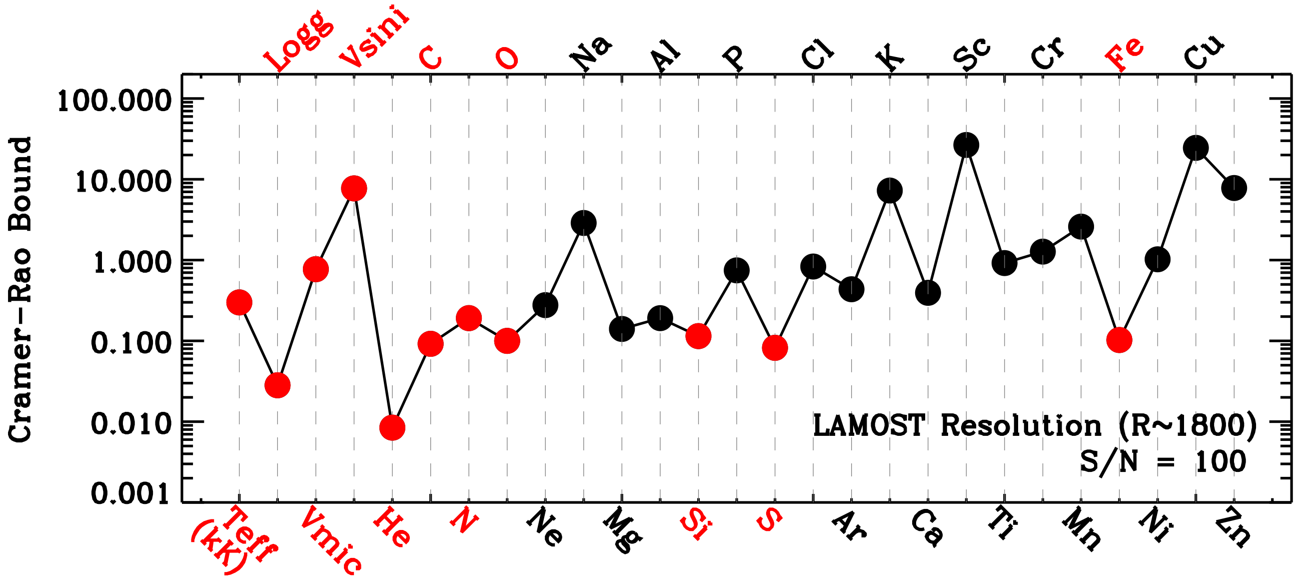

Fig. 4. Theoretical precision limits for stellar label determination from LAMOST of hot stars, expressed through the CR bound, $\sqrt{K_{ii}}$ (Eq. (3)), for a fiducial B star with $T_{\rm eff} = 18\,000$ K, $\log g = 4.0$, [Fe/H] = 0, and $v\sin i = 100\,{\rm km\,s^{-1}}$. The LSF and S/N of the Kurucz model spectra as a function of wavelength are matched to that of a LAMOST B star spectrum, with $S/N = 100$ at 4650 Å. The CR bound is expected to vary by 1/(S/N). The units of the CR bound on the y axis differ from label to label: they are 1000 K for $T_{\rm eff}$, cm s^{-2} for $\log g$, km s^{-1} for $v_{\rm mic}$ and $v\sin i$, and dex for abundance [X/H], except for He. For He, the number indicates the atomic number ratio $N_{\rm He}/N_{\rm tot}$. Labels shown in red are those we attempt to derive from the LAMOST spectra in this work.

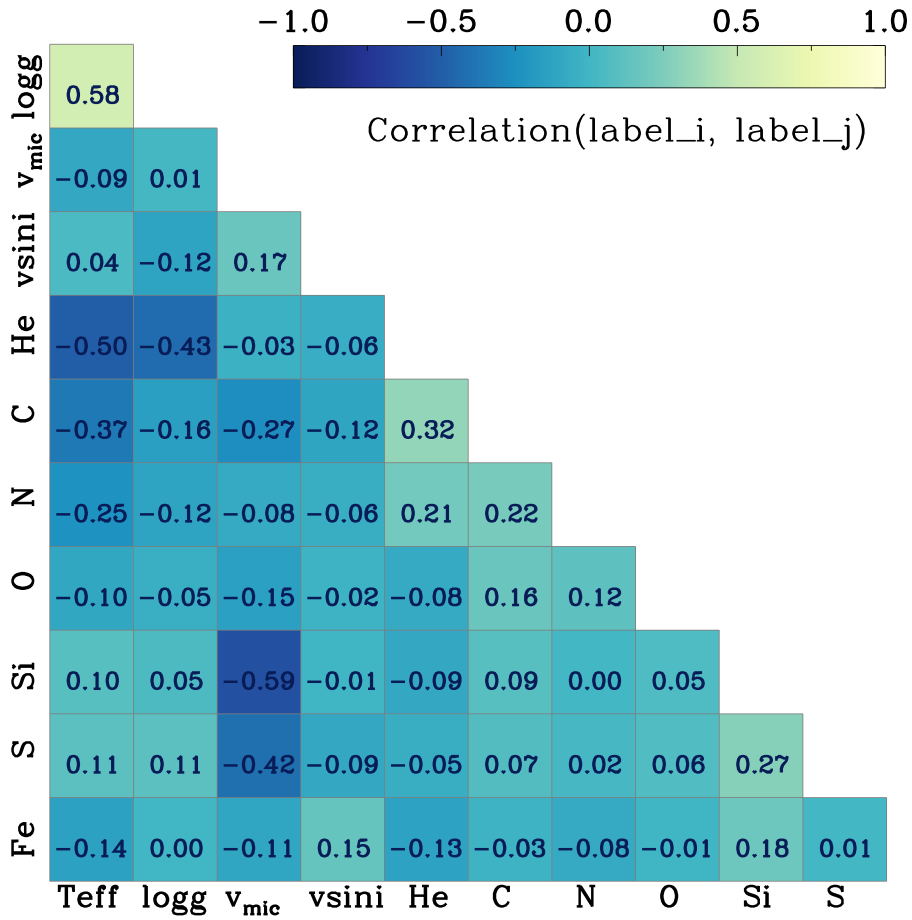

Fig. 5. Correlation coefficients among the stellar label estimates, predicted by the CR bound calculation (see Eq. (3)). These correlations were derived for the spectrum of a fiducial B-type star with $T_{\rm eff} = 18\,000$ K, $\log g = 4.0$, [Fe/H] = 0, and $v\sin i = 100\,{\rm km\,s^{-1}}$. Prominent among the correlations are those between estimates of $T_{\rm eff}$ and $\log g$, and between $N_{\rm He}/N_{\rm tot}$ and both $T_{\rm eff}$ and $\log g$. The nontrivial correlations suggest that most spectral features are somewhat blended and thus call for a simultaneous fit of all pertinent labels to the entire spectrum.

Table 2. Range and sampling method of the stellar labels of Kurucz ATLAS12 models, calculated for the HotPayne spectral modeling (see Sect. 4).

Labels	Range	Sampling
$T_{\rm eff}$	[7500, 60 000] (K)	MIST
$\log g$	[0, 5]	MIST
[Fe/H]	[−3,0.5]	MIST
$v_{\rm mic}$	[0, 15] (km s^{-1})	Random
$v\sin i$	[0, 500] (km s^{-1})	Random
$N_{\rm He}/N_{\rm total}$	[0, 0.25]	Random
[C/Fe]	[−2,3]	Random
[N/Fe]	[−2,3]	Random
[O/Fe]	[−2,3]	Random
[Si/Fe]	[−1,2]	Random
[S/Fe]	[−1,2]	Random

Notes. The MIST sampling method means we draw from the MIST grid uniformly with equivalent evolutionary phases.

lines). Similarly, the $N_{\rm He}/N_{\rm tot}$ abundance estimate is strongly anticorrelated with $T_{\rm eff}$ and $\log g$. The [Si/H] and [S/H] are strongly correlated with $v_{\rm mic}$. The [C/H] and [N/H] are moderately correlated with $T_{\rm eff}$. The $v\sin i$, [O/H] and [Fe/H] are not strongly correlated with any of the other labels. The nontrivial correlation between labels due to somewhat degenerate spectral features further demonstrates that the pertinent labels can be only be obtained by fitting all labels simultaneously through full spectral fitting.

To sum up, inspired by this exploration of the CR bounds for low-resolution spectra of hot stars, we chose to build a spectral model for these LAMOST spectra that depends on 11 labels, namely $T_{\rm eff}$, $\log g$, $v_{\rm mic}$, $v\sin i$, $N_{\rm He}/N_{\rm tot}$, [C/H], [N/H], [O/H], [Si/H], [S/H], and [Fe/H] (Sect. 4), as they are among labels we can derive from the spectra with the best precision.

4. Stellar label determination

In this section we introduce our method of determining stellar labels for hot stars from the LAMOST spectra. This includes an introduction of the Kurucz model spectral library, the application of HotPayne for the spectral fitting, as well as examinations of the resultant label estimates.

4.1. The Kurucz synthetic spectral library

We built a library of synthetic spectra generated with SYNTHE, using the Kurucz LTE atmospheric model calculated with ATLAS12 (Kurucz 1970, 1993, 2005). We considered 11 physical labels – $T_{\rm eff}$, $\log g$, [Fe/H], $v_{\rm mic}$, $v\sin i$, $N_{\rm He}/N_{\rm tot}$, and [X/H] for X = C, N, O, Si, and S for the spectral modeling. The parameter ranges are presented in Table 2. For $T_{\rm eff}$, $\log g$, and [Fe/H], we sampled the grids from the MESA Isochornes and Stellar Tracks (MIST; Dotter 2016; Choi et al. 2016) uniformly in equivalent evolutionary phase. While the other labels were uniformly drawn within the given range for each label. For all other abundances not presented in Table 2 but are required to generate the synthetic spectra models, we simply scaled them with [Fe/H], assuming the solar abundance scale of Asplund et al. (2009). The model spectra were broadened according to the $v\sin i$ value. To ensure the convergence of the model computation, we required that, at Rosseland mean opacities of $\tau_{\rm ross} = 0.1$ and of $\tau_{\rm ross} = 1$, the flux errors be smaller than 2% and the flux derivative errors be smaller than 20%. Also, we required the maximum variation in temperature for the inner atmospheric layers (40–80) between two iterations to be fractionally smaller than 10^{-4}.

In total, we generated 10 127 Kurucz model spectra. The native model spectra generated with the SYNTHE have a resolving power of $R = 500\,000$, and we degraded them to match the line spread function (LSF) of the LAMOST spectra. We adopted a Gaussian profile for the LSF, the width of which is a wavelength-dependent function, approximated by a two-order polynomial for the blue and red spectrograph arms, separately. The LAMOST spectra have $R \sim 1800$ at 5000 Å ($FWHM \sim 2.8$ Å), but the detailed spectral LSF is found to vary from fiber to fiber and exposure to exposure (e.g., Xiang et al. 2015). To generate a range of realistic LSFs for the model grid, we assign a specific LSF for each model spectrum, using individual LAMOST LSFs derived from different exposures and fibers. The spectra are further broadened to their corresponding $v\sin i$ assuming rotational profiles with a constant limb darkening coefficient of 0.6 (Gray 1992; Hubeny & Lanz 2011).

Our LTE Kurucz models come with two caveats. First, for very hot ($\gtrsim$40 000 K) stars with either relatively high metallicity ([Fe/H] $\gtrsim -0.25$ dex) or relatively low $\log g$ values ($\lesssim$4.0), those with very high CNO abundance values can lead to a low hydrogen fraction (e.g., <0.6). In this case, the model computations are hardly converged, which leads to a poor coverage in those parts of the parameter space.

Second, the model spectra are generated using the Kurucz 1D LTE atmospheric model. As mentioned in Sect. 1, it is known that the NLTE effect and stellar winds are non-negligible for hot stars, and especially strong for giants (e.g., Kurucz 1970; Kudritzki 1979, 1998; Hubeny & Lanz 1995; Lanz & Hubeny 2007; Hainich et al. 2019). In this work, we chose to apply the LTE model as a proof of concept. While the Kurucz model spectra may be less accurate than NLTE spectra, they are flexible in obtaining self-consistently computed model spectra across the broad parameter range we want, and they use relatively complete line lists. Nonetheless, as we demonstrate in the rest of the paper, further efforts based on NLTE model spectra are necessary in order to generate accurate elemental abundances.

4.2. HotPayne

Payne is a full spectral fitting tool for stellar label determination built on several key ingredients: First, it trains a flexible spectral interpolation model based on a neural network that allows one to interpolate spectra precisely in high dimensions (e.g., ~20), Furthermore, Payne emphasizes the importance to generate models self-consistently. The self-consistency here refers to the fact that we always solve for the atmospheric structure before running the radiative transport for changes in any of the stellar labels. Secondly, Payne fits all labels of consideration simultaneously from the full spectra with the goal to make full use of the information in the spectra and to properly characterize the covariances of labels arose due to line blending effects. Such an effect is particularly prominent for low-resolution spectroscopy (Ting et al. 2019).

To train Payne, we divided the Kurucz spectral library into two sets, the "training set", with 9115 spectra, and the validation test set, with 1012 spectra. We trained the neural-network model for the pixel-by-pixel spectral interpolation, predicting the flux at each wavelength as a continuous function of the labels. Our neural network model has two layers, each with 100 neurons, and the Sigmoid activation function is adopted to scale the input spectra.

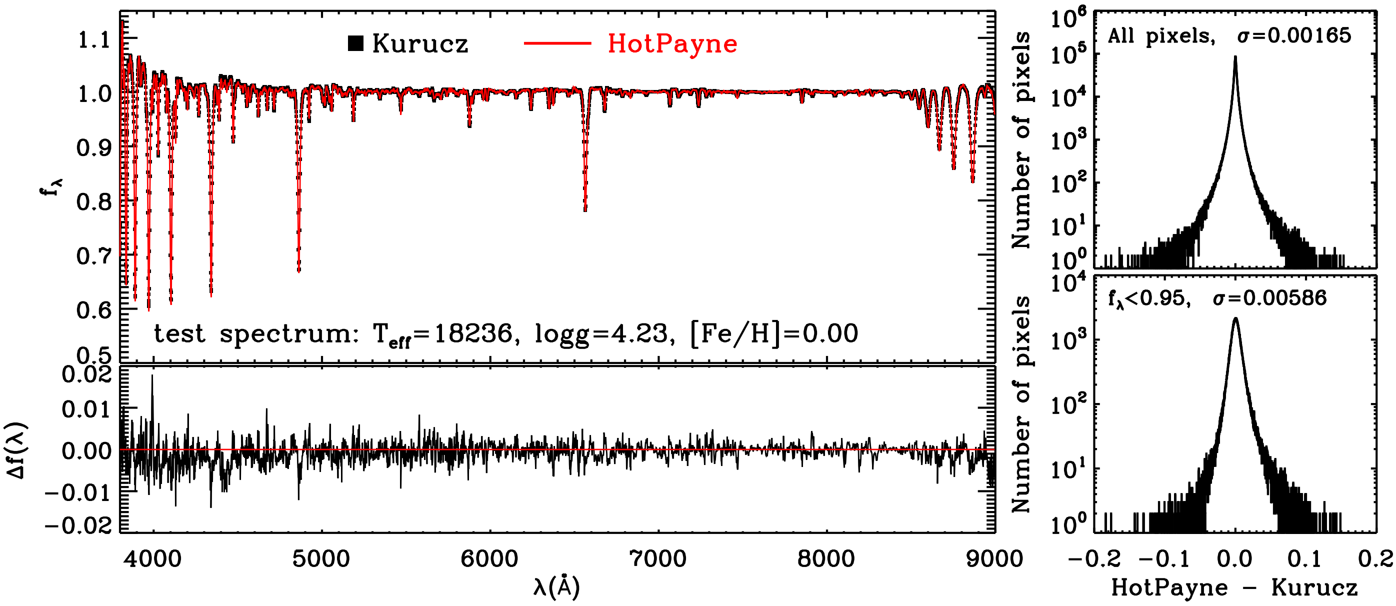

Fig. 6. Precision of the spectrum prediction by HOTPAYNE, which entails a neural-net-enabled interpolation among Kurucz model spectra. The *left panels* show the comparison between the Kurucz spectrum and its prediction by HOTPAYNE for a test star with $T_{\rm eff}$ = 18 236 K, log g = 4.23, and [Fe/H] = 0.0. *Top-right panel*: distribution of flux residuals for all of our 1012 test spectra, each with 3801 pixels in the wavelength range of $\lambda\lambda$3750–9000 Å. A Gaussian fit to the distribution yields a standard deviation of 0.165%. *Bottom-right panel*: analogous to the top-right panel, but only for pixels whose normalized spectral fluxes are smaller than 0.95, i.e., for strong absorption lines, mostly the cores of H and He lines. The standard deviation from a Gaussian fit is 0.586% for the strong absorption features.

The training is carried out in the PYTORCH environment, using the adaptive moment estimation algorithm (Kingma & Ba 2014). In the training process, we have 17 free parameters, including the 11 target stellar labels shown in Table 2, and six more coefficients describing the LSF (three for each of the blue and red arm, respectively) as introduced above. We also explored the possibility of deriving the macroturbulent velocity, $v_{\rm macro}$, which reflects the contribution of the non-rotational line profile broadening (e.g., Simón-Díaz & Herrero 2014), on top of $v\sin i$, but we found that the two parameters are too degenerate at the LAMOST resolution. We chose to only fit for $v\sin i$ in this work. Consequently, our $v\sin i$ estimates are likely contributed by both the rotational broadening and macroturbulence. Our estimates might overestimate the true $v\sin i$ in the case where the non-rotational broadening is comparable or stronger than the rotational broadening. For example, Simón-Díaz & Herrero (2014) found that for stars with $v\sin i \lesssim 120$ km s^{-1}, ignoring $v_{\rm macro}$ may lead to an overestimate of $v\sin i$ by ~25 km s^{-1} in general.

Figure 6 illustrates the precision of the spectral interpolation by HOTPAYNE, deduced from the test spectra. For all individual pixels of the 1012 test spectra, the overall dispersion of the residuals between the HOTPAYNE fits and the Kurucz model flux is 0.165%, which is equivalent to a spectral S/N of 600. For those pixels with normalized flux values smaller than 0.95, mostly the cores of strong H and He lines, the dispersion of the residuals is 0.586% (or equivalently, a S/N of 200). In other words, the interpolation errors from PAYNE are largely negligible given the typical S/N of LAMOST is lower than 200.

To assess the internal precision of stellar label determination with HOTPAYNE, we make a sanity check using the Kurucz test spectral sample. That is, we derived the stellar labels of the Kurucz test spectra using HOTPAYNE and then compared them to the true labels. We added noise to the test spectra (with $S/N \sim 100$ at 4650 Å), as for the CR bound calculations. For reason that we elaborate on in Sect. 4.3, we adopted a fixed LSF by taking the mean LAMOST LSF and did not fit for them; we analyzed the LAMOST data the same way.

Figure 7 plots the differences (residuals) between the HOTPAYNE stellar label estimates and the true values as a function of the true $T_{\rm eff}$, and color-coded by the true values of the labels. On average, the model-to-model fits show that we can recover the stellar labels, except for the very metal poor end. A Gaussian fit to the residuals suggests that, on the whole, the internal precision is about 2.1% for $T_{\rm eff}$, 0.07 dex for log g, 1.2 km s^{-1} for $v_{\rm mic}$, 12.1 km s^{-1} for $v\sin i$, 0.02 for $N_{\rm He}/N_{\rm tot}$, 0.09 dex for [O/H], about 0.15 dex for [C/H], 0.18 dex for [Fe/H], 0.20 dex for [Si/H], 0.26 dex for [N/H], and 0.39 dex for [S/H]. We note that for some labels such as [N/H] and [S/H], the error is partially dominated by a few outliers, a consequence of the strong variations in label precision across the parameter ($T_{\rm eff}$ and [X/H]) space. For stars with $T_{\rm eff} \sim 18\,000$ K and [Fe/H] ~ 0 dex, we find that the precision is consistent with the CR bound (Fig. 4) within a factor of two.

The precision of HOTPAYNE's label estimates vary significantly with both $T_{\rm eff}$ and [X/H]. For stars with $T_{\rm eff} < 11\,000$ K, the log g residuals exhibit a strong negative trend with $T_{\rm eff}$, causing a $T_{\rm eff}$-dependent systematic bias in log g of up to about 0.2 dex at the low-$T_{\rm eff}$ limit. A similar trend is also presented in the [Fe/H], [Si/H], and [S/H] panels, causing a $T_{\rm eff}$-dependent systematics of about 0.1 dex in [Fe/H], and larger values in [Si/H] and [S/H]. This trend reflects an imperfection of our spectral emulation for this cool part of the parameter space, probably due to either insufficiently flexible neural networks or nonoptimal training process (e.g., gets stuck in a local minimum). There are also many more metal lines in the low-temperature spectra, which increase significantly the complexity of the emulation. At this cool end, our test shows that the $N_{\rm He}/N_{\rm tot}$ estimates can be challenging but the $N_{\rm He}/N_{\rm tot}$ estimates are robust for stars

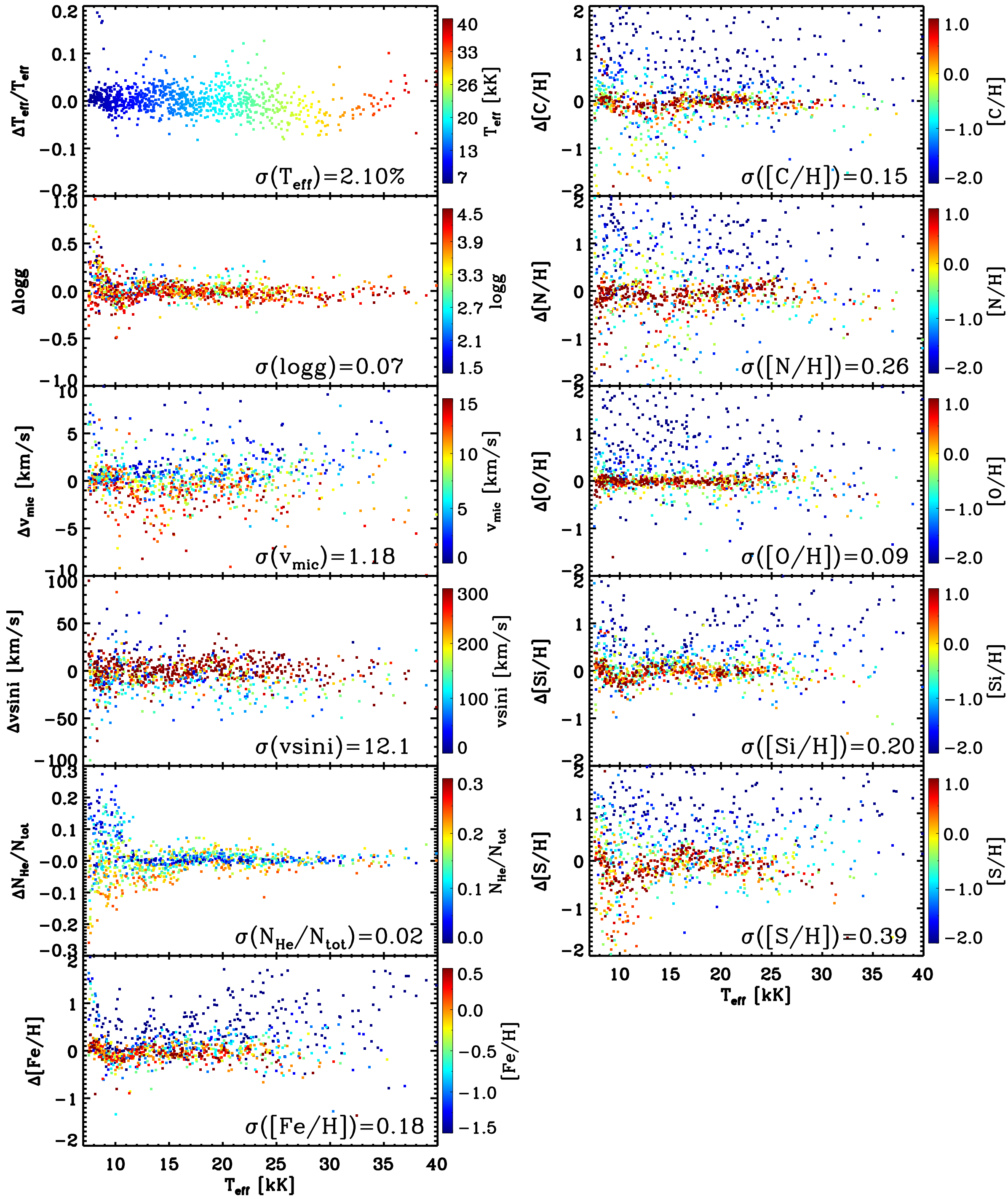

Fig. 7. Precision of the HotPayne label estimates as a function of $T_{\rm eff}$. Plotted are the differences (residuals) between the labels from HotPayne and the true values for a test set of Kurucz model spectra with $S/N \sim 100$, color-coded by the true values. The numbers marked in the figure are the dispersion derived from a Gaussian fit to the residuals. The precision of the HotPayne label determination varies with $T_{\rm eff}$ and [X/H]. For metal-poor stars of [X/H] < -1 dex, the [X/H] estimates may suffer large uncertainties. However, we note that in the real case of the LAMOST survey, we do not expect there to be many hot stars with such low metallicities.

with $T_{\rm eff} > 11\,000$ K, because the He lines are more prominent at this temperature range. On the high-temperature side of $T_{\rm eff} > 25\,000$ K, the $T_{\rm eff}$ estimates exhibit a somewhat artificial trend, with an amplitude of $\lesssim$3%. However, we note that for such a high-temperature case, the NLTE effects may lead to much larger systematics in the $T_{\rm eff}$ estimates (see Sect. 4.7). For the metal-poor stars of [X/H] $\lesssim -1$ dex, the abundance estimates may suffer large deviations from the true values, especially for [C/H], [N/H], [O/H], and [S/H]. This is just the manifestation that, for metal-poor stars, the spectral features become too weak to allow for any abundance estimation. However, we note that in the real case of LAMOST survey, we do not expect that there are many hot stars with such a low metallicity, except for some chemically peculiar stars with very low CNO abundances due to atomic diffusion (e.g., Michaud et al. 2015).

While the model-to-model fit remains optimistic, we caution that when fit to the LAMOST spectra, the fit is likely to incur larger uncertainties due to the mismatch between model and observed spectra due to model imperfections.

4.3. The line spread function

The LSF of the individual LAMOST spectra vary between fibers and between exposures, which is expected to have an impact on the stellar label determination (e.g., Xiang et al. 2015). By design, our HotPayne module can fit the LSF from the spectra simultaneously with the stellar labels. However, we found it challenging to determine the LSF from fitting Kurucz models to the LAMOST spectra, because such fits are impacted by even small, systematic line core mismatches owed to systematics of the Kurucz models. A more sensible way is to take the LSF of each LAMOST spectrum derived from arc lamps and sky emission lines as input, and fix it in the HotPayne fitting. However, currently such LSF data for LAMOST spectra are still missing.

As a compromise, we instead adopted the mean LSF of the LAMOST spectra derived by the author using arc lamps and sky emission lines from the LSS-GAC DR2 (Xiang et al. 2017). We tested the impact of the fiber-to-fiber and exposure-to-exposure variations in the LSF on the stellar label determination by assigning each Kurucz test spectrum (Sect. 4.2) a unique LSF from the LSS-GAC DR2, and we fit them with the HotPayne model spectra with the mean LSF. We find that ignoring the variations in LSF will cause larger uncertainties on the $v\sin i$ estimates, which typically amount to 17 km s^{-1} (compared to the 12 km s^{-1} from the mean LSF as shown in Fig. 7), as a consequence of the LSF and $v\sin i$ covariance. For a minor number of spectra with very different LSF, the $v\sin i$ estimates may differ by larger than 30 km s^{-1} from the results of the mean LSF. Whereas for the other labels, the LSF has only marginally effects on the HotPayne determination in most cases.

4.4. Initial estimates of $T_{\rm eff}$ from line indices

To facilitate the spectral fitting with HotPayne to the LAMOST data set, we first made an initial estimate of $T_{\rm eff}$ based on the indices (equivalent widths) of the Ca II 3933 Å, H I and Ca II 3967 Å, and H I 4101 Å lines. To compute the line indices, we adopted the λ3912–3922 Å and λ3946–3956 Å windows for continuum estimates of the Ca II 3933 line. Similarly, the continuum for H I+Ca II 3967 Å is derived from the λ3946–3956 Å and λ3980–3990 Å windows, and the continuum for H I 4101 Å line is derived from the λ4080-4090 Å and λ4114–4124 Å windows.

We used a subset of 57 039 LAMOST spectra as the reference set to build an empirical relation between line indices and $T_{\rm eff}$. This reference set includes 16 002 stars from the LAMOST OB sample of Liu et al. (2019), and the remaining ones are selected from the LAMOST and Zari et al. (2021) common star sample, the majority of which are A- and F-type stars according to the LAMOST classification. The $T_{\rm eff}$ of these reference spectra are estimated with HotPayne assuming an initial $T_{\rm eff}$ set to be the mean $T_{\rm eff}$ of our training library spectra. For late F-type stars, the $T_{\rm eff}$ estimates with HotPayne might be inaccurate as they are outside the $T_{\rm eff}$ coverage of the Kurucz training set (i.e., cooler than <7500 K). We therefore adopted the $T_{\rm eff}$ provided in Xiang et al. (2019) if it has smaller χ^2 than the HotPayne fits. We fitted a three-order polynomial to this reference set, thereby predicting $T_{\rm eff}$ from line indices. This yields the following relation:

$$\begin{aligned}\log T_{\rm eff} = {} & 4.71047 - 4.43455\times10^{-2}C_{3933} - 5.37380\times10^{-2}C_{3967} \\ & + 3.46889\times10^{-2}C_{4101} - 5.68345\times10^{-3}C_{3933}^2 \\ & - 1.26749\times10^{-2}C_{3967}^2 - 2.40555\times10^{-2}C_{4101}^2 \\ & + 1.26779\times10^{-2}C_{3933}C_{3967} - 4.19722\times10^{-2}C_{3933}C_{4101} \\ & + 2.34087\times10^{-2}C_{3967}C_{4101} \\ & + 8.44885\times10^{-4}C_{3933}C_{3967}C_{4101} \\ & - 9.98430\times10^{-4}C_{3933}^2C_{3967} + 2.23513\times10^{-3}C_{3933}^2C_{4101} \\ & + 6.24433\times10^{-4}C_{3933}C_{3967}^2 + 1.40913\times10^{-3}C_{3933}C_{4101}^2 \\ & - 1.78560\times10^{-3}C_{3967}^2C_{4101} + 1.85542\times10^{-3}C_{3967}C_{4101}^2 \\ & + 5.01277\times10^{-4}C_{3933}^3 + 5.10836\times10^{-4}C_{3967}^3 \\ & + 1.52060\times10^{-4}C_{4101}^3,\end{aligned}$$

where C_{3933}, C_{3967}, and C_{4101} are the equivalent widths of the Ca II 3933 Å, H I+Ca II 3967 Å, and H I 4101 Å lines, respectively. For OB stars, the Ca line index may suffer from some uncertainties due to interstellar absorption, but since our estimates from the line indices only serve as an initial guess to optimize the running time, such uncertainties have minimal impact on our final estimates.

4.5. Spectral masking and the spurious absence of early A-type stars

When fitting the LAMOST spectra, we also masked the wavelength windows with known telluric or interstellar medium absorption. In particular, we masked the known strong diffuse interstellar bands from the Diffuse Interstellar Band (DIB) directory (Hobbs et al. 2008), the Na I λ5890, 5896 doublet, and the Ca II K λ3933 line. We also discarded the blue (λ < 3820 Å) and red (λ > 8880 Å) edges of the LAMOST spectra, as well as the dichroic region (5700–6050 Å) of the blue and red arms of the LAMOST spectrographs. The masked wavelength regions are shaded in Fig. 8.

Our preliminary fits lead to resultant stellar parameters that exhibit a desert of early A-type and late B-type dwarf stars ($9000 \lesssim T_{\rm eff} \lesssim 11\,000$ K) in the $T_{\rm eff}$–$\log g$ diagram (left panel of Fig. 9). This desert is unexpected in the context of Galactic evolution and star formation/evolution scenarios. A careful inspection of the spectral models reveals that this artifact is possibly related to the change of signs in the gradient spectra of hydrogen lines for the A0-type stars. For A0 stars ($T_{\rm eff} \sim 9500$ K), the strength of hydrogen lines reaches a maximum, so that the gradient spectra of hydrogen lines for A0 stars reach a minimum, with absolute values close to 0 (see Fig. A.1). This causes an inflection point in the spectral gradient features of the hydrogen lines as a function of $T_{\rm eff}$. As a consequence, the parameter estimates are sensitive to the initial $T_{\rm eff}$ values. Depending on the initial guesses from the line indices, the fits tend to pile up due to the "gradient gap", causing the desert in the $T_{\rm eff}$-$\log g$ diagram as seen in the left panel of Fig. 9. These effects are illustrated in the right panel of Fig. 9.

As a remedy, we masked all hydrogen lines for label determination for these early A-type and late B-type stars. The masked regions are shown in the bottom panels of Fig. 8. Masking the hydrogen lines can reduce the spectral information content and

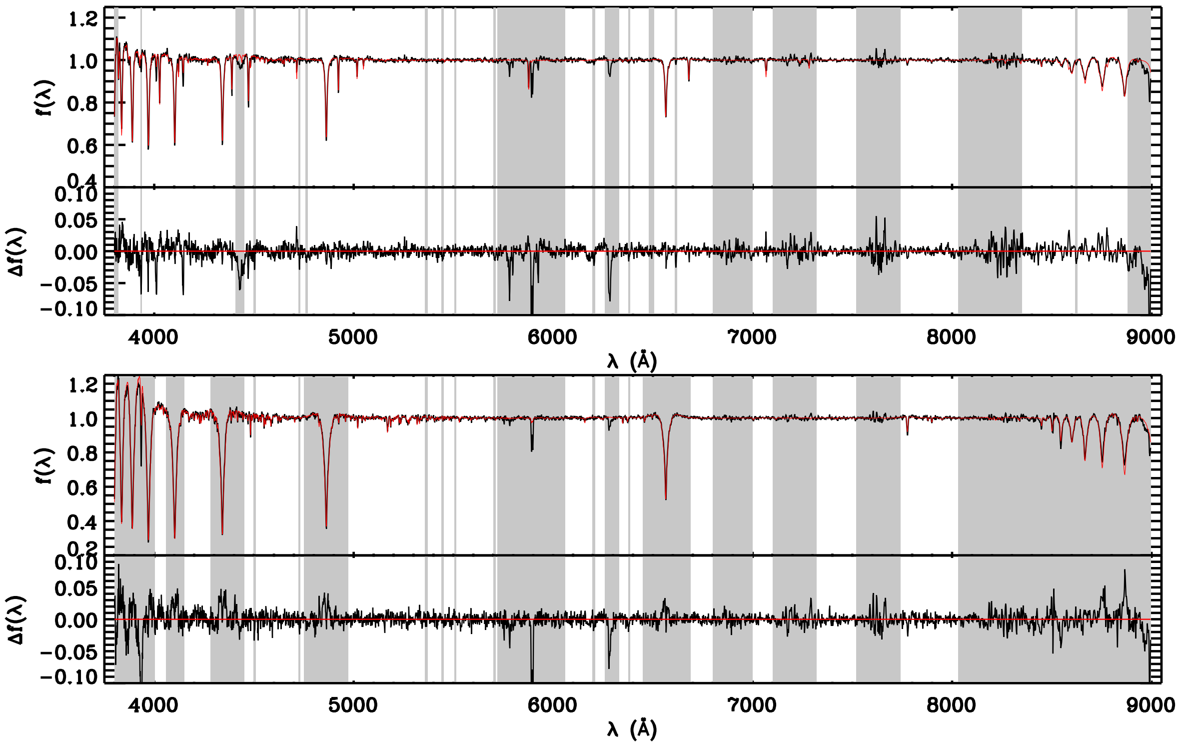

Fig. 8. Example of the HotPayne fit for a B-type star (*top*) and an early A-type star (*bottom*). The LAMOST spectra are shown in black, and the best-fit HotPayne models are shown in red. The masked telluric bands, the DIB, the interstellar absorption of Na and Ca, and the dichroic region are marked by shaded regions. For the case of early A-type stars, the wavelength windows of hydrogen lines (Balmer and Paschen series) are also masked (see text).

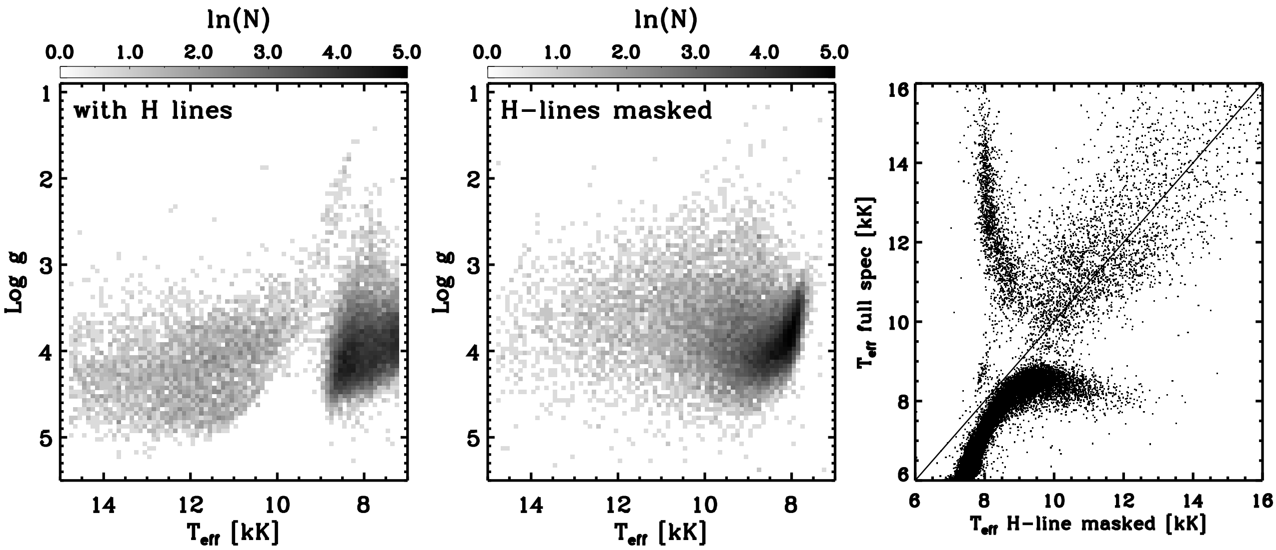

Fig. 9. $T_{\rm eff}$-$\log g$ diagrams for $T_{\rm eff}$ derived from the LAMOST spectra without (*left*) and with (*middle*) the hydrogen lines masked. *Right panel*: direct comparison of the $T_{\rm eff}$ estimates in the left and middle panels. The inclusion of hydrogen lines for the label determination causes an artificial desert of early A-type (and late B-type) dwarf stars in the $T_{\rm eff}$-$\log g$ diagram due to the reverse of spectral gradients $\partial f_\lambda/\partial\lambda$ for the A0 stars. Masking the hydrogen lines resolves this issue (middle panel). In this work, we consider label estimates by combining these two estimates, putting more weight on the masked estimates in the $T_{\rm eff}$ range where the hydrogen lines gradient reversal happens.

lead to lower precision for the label determinations, particularly for $T_{\rm eff}$ and $\log g$. Nonetheless, an examination of the CR bound for early A-type stars suggests that the hydrogen-masked spectra are still able to yield robust estimates for many of the labels (see Fig. B.1). The $T_{\rm eff}$–$\log g$ diagram deduced for a subset of LAMOST stars after masking the hydrogen lines is shown in the middle panel of Fig. 9. Masking the hydrogen lines resolved the non-sensible desert. In order to benefit from the full spectral fitting as we can while mitigating the impact of hydrogen lines on the A0 stars, we combined the two sets of label estimates to give the ultimate label estimates[2]. The combination is carried out using the weighted mean value,

$$\boldsymbol{l} = \frac{\omega_1 \boldsymbol{l}_1 + \omega_2 \boldsymbol{l}_2}{\omega_1 + \omega_2}, \tag{4}$$

where $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ represent the labels derived from spectra with and without the hydrogen lines masked, respectively. We adopted the Hann function as the weights for labels derived from the H-masked spectra,

$$\omega_1(x) = \begin{cases} \frac{1}{2}\left(1 + \cos\left(\frac{2\pi x}{L}\right)\right) = \cos^2\left(\frac{\pi x}{L}\right), & \text{if } |x| \leq L/2 \\ 0, & \text{if } |x| > L/2, \end{cases} \tag{5}$$

where $x = T_{\rm eff} - 9500$ K, and the $T_{\rm eff}$ is that from the H-masked spectra. The weights for the labels derived from the full spectra are

$$\omega_2 = 1 - \omega_1. \tag{6}$$

As for the smoothing length, for stars with $T_{\rm eff}^{\rm fullspec} > 9500$ K, we adopted a smoothing length L of 5000 K, such that $\omega_1 = 0$ for stars with $T_{\rm eff}^{\rm H-masked} > 12\,000$ K. And for stars with $T_{\rm eff}^{\rm fullspec} < 9500$ K, we adopted a length L of 3000 K, such that $\omega_1 = 0$ for stars with $T_{\rm eff}^{\rm H-masked} < 8000$ K.

4.6. Examination with repeat observations

About 30% of the LAMOST spectra are repeat observations of common targets (e.g., Yuan et al. 2015; Xiang et al. 2017). The label estimates from these repeated observations can differ due to a number of reasons, such as the photon noise, the fiber-to-fiber and temporal variation in the LSF, and imperfect calibration processes (e.g. flat-fielding, sky and scatter light subtraction, etc.). Comparing the label estimates among repeated observations thus gives an estimate of the precision of the label estimates[3]. To estimate the quality of our fit, we adopted only the repeated observations that have comparable (<20% difference) spectral S/N. The precision of the label is estimated through the square root of the dispersion among repeated observations.

The precision of our estimates as a function of S/N is shown in Fig. 10. The figure shows that, at S/N below 100, the photon noise dominates the spectral uncertainties, with the precision improve as 1/(*S/N*), as expected from the CR bound calculation. The precision only gradually improves beyond $S/N > 100$ because the instrument and calibration errors dominate the spectral uncertainties. In the case of $S/N < 50$, the label estimates have a precision of 700 K in $T_{\rm eff}$, 0.19 dex in $\log g$, 4.1 km s^{-1} in $v_{\rm mic}$, 51.5 km s^{-1} in $v\sin i$, 0.03 in $N_{\rm He}/N_{\rm tot}$, 0.27 dex in [O/H], ∼0.5 dex in [Si/H], and [Fe/H], ≳1.0 dex in [C/H], [N/H], and [S/H]. Whereas in the case of $S/N > 100$, the uncertainties reduce to 260 K in $T_{\rm eff}$, 0.08 dex in $\log g$, 0.85 km s^{-1} in $v_{\rm mic}$, 13.0 km s^{-1} in $v\sin i$, 0.01 in $N_{\rm He}/N_{\rm tot}$, ∼0.20 dex in [C/H], [N/H], and [S/H], ≲0.10 dex in [O/H], [Si/H], and [Fe/H]. In general, these values for the case of $S/N \simeq 100$ are consistent with the CR bounds.

[2] Since those with $T_{\rm eff}^{\rm H-masked} < 9500$ K and $T_{\rm eff}^{\rm fullspec} > 9500$ K and $T_{\rm eff}^{\rm H-masked} > 9500$ K and $T_{\rm eff}^{\rm fullspec} < 9500$ K are likely affected by imperfect initial $T_{\rm eff}$ estimates from the line indices. For those objects, we opted to redetermine their full-spectra-based labels using the H-masked $T_{\rm eff}$ as the initial estimates.

[3] Stellar variability could be an issue as well, but we do not expect it to dominate over the measurement uncertainties for the majority of stars.

Table 3. Sources of high-resolution optical spectra from ESO archives.

Star ID	Wavelength range	Instrument
HD 149438	3800–6800 Å	HARPs
HD 36512	3700–9000 Å	UVES
HD 36822	3700–9000 Å	UVES
HD 36960	3700–9000 Å	UVES
HD 41117	3700–9000 Å	UVES
HD 52089	3700–9000 Å	UVES
HD 79186	3700–9000 Å	UVES
HD 172167	6800–9000 Å	UVES
	3700–4000 Å	UVES
	4000–6800 Å	Takeda et al. (2007)
HD 34816	3800–9000 Å	XSHOOTER
HD 35299	3700–9000 Å	XSHOOTER
HD 52382	3700–9000 Å	XSHOOTER
HD 87737	3700–9000 Å	XSHOOTER
HD 92207	3700–9000 Å	XSHOOTER
HD 111613	3700–9000 Å	XSHOOTER
HD 209008	3700–9000 Å	XSHOOTER

4.7. Comparison with literature

The LAMOST surveys mostly target stars fainter than 10 mag in the SDSS *r*-band; however, most of the literature stars are brighter. As a consequence, there is a lack of a meaningful sample of reference hot stars that have well-established parameters and abundances from high-resolution, NLTE-based spectroscopy. For the current purpose, we compiled our reference stars from two sources, the first is a collection of high-resolution, NLTE-based stellar labels of 11 A-type and B-type stars from Przybilla et al. (2006, 2010), and Nieva & Przybilla (2012). These stars were not observed by the LAMOST survey but have available high-resolution full-optical spectra from the European Southern Observatory (ESO) science archive. We degraded them to the LAMOST resolution with the mean LSF to perform the label determination with HotPayne. For verifying stellar parameters ($T_{\rm eff}$, $\log g$, $v\sin i$) of supergiants, our reference set contains four B-type supergiants, namely HD 41117, HD 52089, HD 52382, and HD 79186, whose reference parameters are from Haucke et al. (2018). Table 3 presents a summary of the sources of high-resolution spectra for these reference stars. Their stellar parameters and abundances from high-resolution NLTE analysis in literature are presented in Tables 4 and 5, respectively.

The other stars in our reference sample are slowly pulsating B-type (SPB) stars from Gebruers et al. (2021). The Gebruers et al. (2021) sample consists 20 SPB stars in the *Kepler* field, and with light curves from the *Kepler* mission (Borucki et al. 2010). The stellar labels of these SPB stars are estimated from high-resolution spectra, utilizing LTE-based stellar model spectra (Gebruers et al. 2021). Of the 20 SPB stars of Gebruers et al. (2021), 18 have LAMOST spectra from the LAMOST-*Kepler* project (De Cat et al. 2015; Fu et al. 2020). Due to their intrinsic brightness, these LAMOST spectra all have $S/N > 200$.

Figure 11 presents an one-to-one comparison of the labels estimates from HotPayne with their reference values for $T_{\rm eff}$, $\log g$, and $v\sin i$. The comparison suggests that our estimates for these labels are in decent agreement with the literature values,

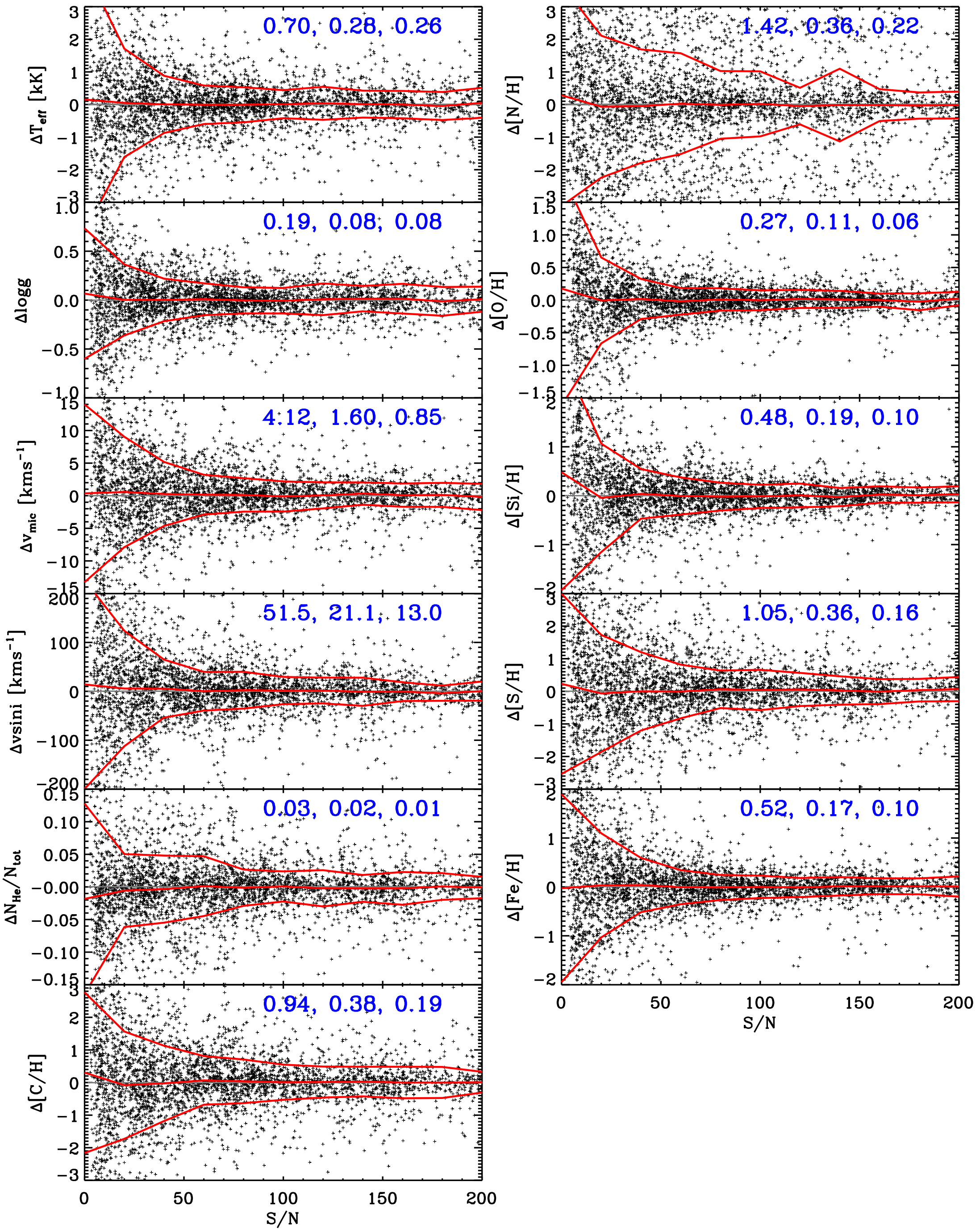

Fig. 10. Label differences between repeat observations as a function of S/N for stars with $T_{\rm eff} > 10\,000$ K. Only repeat observations carried out on different observation nights and with S/N differences of less than 20 percent are adopted. The x axis shows the mean S/N of the repeat observations. The solid lines in red show the median differences as well as the standard deviations at different S/N. The numbers marked in the figure are precision (the scatter divided by $\sqrt{2}$, i.e., assuming both stars contribute equally) for stars with $S/N < 50$, $50 < S/N < 100$, and $S/N > 100$, respectively.

except for the $T_{\rm eff}$ of stars hotter than 25 000 K. For these hot stars, our $T_{\rm eff}$ estimates are systematically lower than the literature values by about 3000 K, due to the NLTE effects. The NLTE models generally predict stronger lines than LTE models (see Figs. C.1 and C.2). As a result, the LTE analysis favors a lower $T_{\rm eff}$ for these hot stars. The $\log g$ shows good agreement with our results, demonstrating the ability of HotPayne to distinguish giants from dwarfs. The A-type supergiant giant star HD 92207 exhibits large deviation in $\log g$, as Przybilla et al. (2006) gives an estimate of 1.2 dex for $\log g$ while our estimate is 2.0 dex. This discrepancy is likely caused by NLTE effects, which are particularly strong for supergiants (e.g., Kudritzki 1976, 1979). The $v\sin i$ shows decent agreement with literature, particularly for stars rotating faster than 100 km s^{-1}, demonstrating the ability of HotPayne to distinguish fast rotating stars from slow rotating stars.

Figure 12 illustrates a similar comparison but $v_{\rm mic}$, $N_{\rm He}/N_{\rm tot}$, and [X/H]. The differences between the HotPayne estimates and the literature values are plotted as a function of $T_{\rm eff}$ of the stars. For stars hotter than 15 000 K, our estimates of $v_{\rm mic}$ are

Table 4. Stellar parameters of our reference stars from literature.

Star ID	$T_{\rm eff}$	$\log g$	[Fe/H]	$v_{\rm mic}$	$v\sin i$
HD 111613	9150 ± 150	1.45 ± 0.10	-0.11 ± 0.10	7.0 ± 1.0	19.0 ± 3.0
HD 92207	9500 ± 200	1.20 ± 0.10	-0.16 ± 0.07	8.0 ± 1.0	30.0 ± 5.0
HD 172167	9550 ± 150	3.95 ± 0.10	-0.54 ± 0.10	2.0 ± 0.5	22.0 ± 2.0
HD 87737	9600 ± 150	2.00 ± 0.15	-0.07 ± 0.10	4.0 ± 1.0	0.0 ± 3.0
HD 209008	$15\,800 \pm 200$	3.75 ± 0.05	0.03 ± 0.08	4.0 ± 1.0	15.0 ± 3.0
HD 35299	$23\,500 \pm 300$	4.20 ± 0.05	0.03 ± 0.10	0.0 ± 1.0	8.0 ± 1.0
HD 36960	$29\,000 \pm 300$	4.10 ± 0.07	-0.02 ± 0.09	4.0 ± 1.0	28.0 ± 3.0
HD 36822	$30\,000 \pm 300$	4.05 ± 0.10	0.02 ± 0.04	8.0 ± 1.0	28.0 ± 2.0
HD 34816	$30\,400 \pm 300$	4.30 ± 0.05	0.04 ± 0.07	4.0 ± 1.0	30.0 ± 2.0
HD 149438	$32\,000 \pm 300$	4.30 ± 0.05	0.04 ± 0.09	5.0 ± 1.0	4.0 ± 1.0
HD 36512	$33\,400 \pm 200$	4.30 ± 0.05	0.03 ± 0.03	4.0 ± 1.0	20.0 ± 2.0
HD 41117	$19\,000 \pm 1000$	2.3 ± 0.1	–	10 ± 5	40
HD 52089	$23\,000 \pm 1000$	3.0 ± 0.1	–	8 ± 2	10
HD 52382	$21\,500 \pm 1000$	2.45 ± 0.1	–	10 ± 2	55
HD 79186	$15\,800 \pm 500$	2.0 ± 0.1	–	11 ± 1	40

Notes. For the B-type supergiants HD 41117, HD 52089, HD 52382, and HD 79186, the parameters are from Haucke et al. (2018). For the other stars, the stellar parameters are from Przybilla et al. (2006, 2010), and Nieva & Przybilla (2012).

Table 5. Elemental abundances of our reference stars from NLTE analyses of high-resolution spectra.

Star ID	$N_{\rm He}/N_{\rm tot}$	[C/H]	[N/H]	[O/H]	[Si/H]	[S/H]
HD 111613	0.105 ± 0.020	-0.26 ± 0.07	0.57 ± 0.10	0.01 ± 0.04	-0.06 ± 0.29	-0.05 ± 0.08
HD 92207	0.12 ± 0.02	-0.10 ± 0.20	0.42 ± 0.04	0.10 ± 0.07	-0.18 ± 0.05	0.00 ± 0.08
HD 172167	0.090 ± 0.010	-0.20 ± 0.11	-0.14 ± 0.06	-0.12 ± 0.05	-0.57 ± 0.05	-0.06 ± 0.03
HD 87737	0.13 ± 0.02	-0.41 ± 0.10	0.58 ± 0.09	0.09 ± 0.05	0.07 ± 0.19	0.03 ± 0.07
HD 209008	0.086[(1)]	-0.10 ± 0.09	-0.03 ± 0.11	0.11 ± 0.11	-0.09 ± 0.04	–
HD 35299	0.086[(1)]	-0.08 ± 0.09	-0.01 ± 0.08	0.15 ± 0.09	0.05 ± 0.05	–
HD 36960	0.086[(1)]	-0.08 ± 0.09	-0.11 ± 0.11	-0.02 ± 0.08	0.05 ± 0.07	–
HD 36822	0.086[(1)]	-0.15 ± 0.14	0.09 ± 0.10	-0.01 ± 0.10	0.05 ± 0.07	–
HD 34816	0.086[(1)]	-0.05 ± 0.05	-0.02 ± 0.15	0.02 ± 0.09	0.03 ± 0.06	–
HD 149438	0.089 ± 0.009	-0.13 ± 0.12	0.33 ± 0.12	0.08 ± 0.08	0.01 ± 0.06	–
HD 36512	0.086[(1)]	-0.08 ± 0.14	-0.04 ± 0.11	0.06 ± 0.09	0.03 ± 0.07	–

Notes. The stellar abundances are from a collection of literature values in Przybilla et al. (2006, 2010), and Nieva & Przybilla (2012). [(1)]A fixed He abundance of 0.086 is adopted for the elemental abundance determination.

systematically larger than the literature values by up to 10 km s^{-1}, due to NLTE effect, which is more prominent for the hotter stars (see the appendix). At cooler end ($T_{\rm eff}$ < 15 000 K), our $v_{\rm mic}$ estimates for dwarfs are in decent agreement with literature values. For example, for Vega (HD 172167), the literature $v_{\rm mic}$ is 2 km s^{-1}, and we obtain 4.4 km s^{-1}. Whereas for the A-type supergiants, our estimates are higher than literature values, again because of NLTE effect, which is more prominent for supergiants than for dwarfs. Although not shown here, for the LAMOST stars with $T_{\rm eff}$ < 9000 K, our results exhibit a typical $v_{\rm mic}$ of ~2 km s^{-1}, which is in line with expectations (e.g., Gebran et al. 2014). The $N_{\rm He}/N_{\rm tot}$ is systematically overestimated by ~0.10, except for the A-type supergiants. This is due to NLTE effects, as explained above, NLTE models have deeper He lines than LTE spectra, which in turn favor lower $N_{\rm He}/N_{\rm tot}$ estimates.

As for the abundances, the comparison reveals a $T_{\rm eff}$-dependent systematic trend in our abundance estimates, largely a consequence of NLTE effects. In particular, for stars with $T_{\rm eff}$ > 25 000 K, our estimates are generally lower than the reference values by 0.3 dex for [Fe/H], 0.5–1.0 dex for [C/H], [Si/H], and [S/H]. It should be noted that for HD 36512 (33 400 K, in black, and barely visible in the plot), our estimates of [Si/H] and [Fe/H] are are erroneously low (<−2 dex). The exact cause is hard to be determined, but it is likely due to the imperfect fringing subtraction of spectra taken by the Ultraviolet and Visual Echelle Spectrograph (UVES). The effect is the most prominent for HD 36152. For stars of 11 000 < $T_{\rm eff}$ < 25 000 K, our [Fe/H] estimates are in decent agreement with Gebruers et al. (2021), but systematically higher than Przybilla et al. (2010); Nieva & Przybilla (2012). Both ours and Gebruers et al. (2021) are based on LTE models, and the good agreement validates our method, further elucidating that the differences with the other studies are largely due to the NLTE effect. We have only one star with NLTE abundances at $T_{\rm eff}$ ~ 15 000 K. Beside NLTE effects, imperfections of line lists may have also played a role for the difference. Particularly, for C we adopted the older line list of Jorgensen et al. (1996), which may need to be updated with more recent line lists (e.g., Masseron et al. 2014). The [O/H] seems to be in decent agreement with literature, except for the A-type supergiants and the stars with $T_{\rm eff}$ > 30 000 K, both of which may

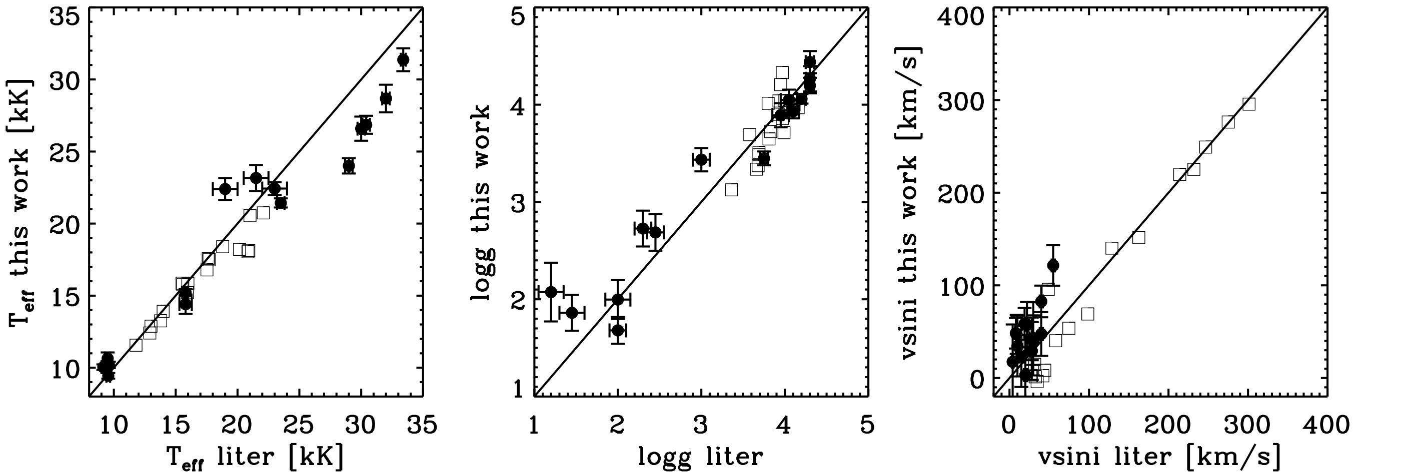

Fig. 11. Comparison of labels between the HotPayne estimates and high-resolution spectroscopic results for our reference stars. For all the dots in black, the spectra are collected from the ESO archive (Table 3) and degraded to the LAMOST resolution for stellar label determination with HotPayne. For these stars, the literature labels are adopted from Przybilla et al. (2006, 2010), Nieva & Przybilla (2012), and Haucke et al. (2018). Open squares represent the SPB stars of Gebruers et al. (2021), which have LAMOST observational counterparts. The Gebruers et al. (2021) high-resolution spectroscopic labels of these SPB stars are derived using LTE-based model spectra. The HotPayne estimates show good agreements with the literature values. For stars with $T_{\rm eff}$ > 25 000 K, the LTE-based estimation with HotPayne underestimates the $T_{\rm eff}$ due to NLTE effects. Similarly, the $\log g$ for supergiants ($\log g$ < 2) are overestimated due to the differences between the LTE and NLTE modeling.

have suffered strong NLTE effects. The reasonable agreement of [O/H] with literature values is not surprising, because oxygen has the most prominent features in the spectra (see Figs. C.1 and C.2). Nonetheless, it seems that even for [O/H], there can be a systematic uncertainty at a level of 0.2 dex from NLTE effects.

5. Results and discussion

5.1. The final sample

We applied HotPayne to our 1 160 000 LAMOST hot-star-candidate spectra. A considerable fraction of these candidates are found to be cool stars with $T_{\rm eff}$ < 7000 K, due to contamination from our target selection. Since our Kurucz spectral model grid covers only $T_{\rm eff}$ > 7500 K (Table 2), the labels for stars with $T_{\rm eff}$ < 7500 K are estimated with HotPayne via extrapolation. We therefore discarded all stars with $T_{\rm eff}$ < 7000 K in our sample. We note that we chose to keep the stars with 7000 < $T_{\rm eff} \lesssim$ 7500 K because this regime was poorly investigated in previous efforts to estimate labels for FGK stellar sample (e.g., Xiang et al. 2019), and our labels might be useful for the general purpose.

We discarded results from spectra with S/N < 5 from our final catalog to ensure the precision of the label estimates. Furthermore, we found that the $T_{\rm eff}$ estimates for a considerable number of (∼5%) intrinsically cool ($T_{\rm eff}$ < 7000 K) stars are erroneously estimated to be hotter than 7000 K due to their low spectral S/N. We also discarded these stars from our sample. These outliers are identified based on their de-reddened colors (see Sect. 5.3 for the derivation of de-reddened colors) and effective temperature derived with the data-driven Payne tool (DD-Payne; Xiang et al. 2019), which provides independent temperature estimates for FGK stars of $T_{\rm eff}$ < 7000 K with a data-driven approach, taking stellar labels from the high-resolution surveys (APOGEE and GALAH) as training sets. Specifically, if a star has a de-reddened *Gaia BP−RP* color redder than 0.4 ($(BP-RP)_0$ > 0.4) and, simultaneously, DD-Payne gives $T_{\rm eff}$ < 7000 K, we deem the star to be an intrinsically cool one, but the current work erroneously gives a high temperature. This criterion is applied to all our sample stars except for those from the OB star sample of Liu et al. (2019), as this sample contains mostly intrinsic hot stars. Ultimately, these criteria leave 332 172 unique stars with 454 693 measurements (spectra) in our final hot-star sample.

While we have shown that our abundances are internally consistent through repeat spectra (Fig. 10), the comparison with the literature values reveals that NLTE effects can remain critical, especially for hot stars, we opted to divide our results into two separate catalogs: a parameter catalog that includes $T_{\rm eff}$, $\log g$, $v\sin i$, [Fe/H], and [Si/H][4], and an abundance catalog that includes $v_{\rm mic}$ and abundances for other elements. Currently, only the parameter catalog is made public available, while the abundance catalog is available only by request to avoid any misuse of the results. In the following we present the results only for stellar labels in the stellar parameter catalog. We also note that we are in the process of applying the same method to NLTE models and will release the full NLTE abundances in the coming study.

5.2. Flagging bad fits

As we are working with a vast set of spectra, collected by a complex instrument system (16 spectrographs, 32 charge-coupled devices, 4000 fibers) of LAMOST (Cui et al. 2012), inevitably there will be some "bad" spectra with erroneous label estimates, either due to intrinsic peculiar properties, for example strong emission-line objects, or due to data reduction artifacts (Xiang et al. 2021). To ensure the robustness of our results, we first identifed such cases by flagging results with unusually large χ^2 in the spectral fitting. We quantified the median and dispersion[5] of

[4] [Si/H] is presented as a representative α-element. As we show in Sect. 5.5, despite the systematics due to the NLTE effects, the [Si/H] estimates are still informative at least for chemically peculiar stars.

[5] The dispersion is defined as a resistant estimate of the standard deviation, adopting the robust_sigma.pro in the IDL astronomical library. See, for instance, Hoaglin et al. (1983) for the method of resistant estimate of the dispersion of a distribution.

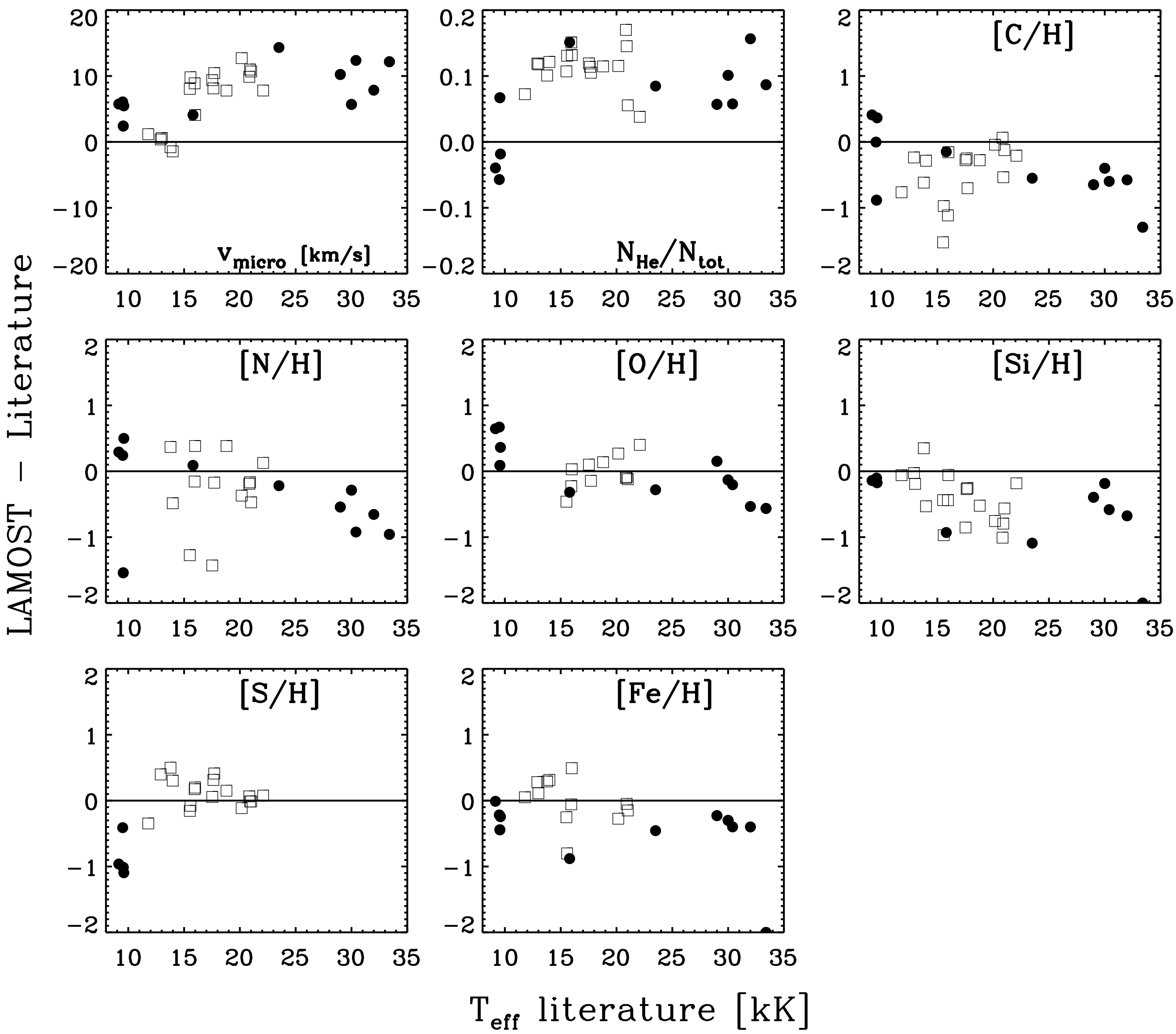

Fig. 12. Difference (LAMOST minus literature) between stellar labels estimated from the LAMOST $R \sim 1800$ spectra with HotPayne and the literature values deduced from high-resolution spectroscopy for the reference stars, as a function of $T_{\rm eff}$. The symbols have the same meanings as in Fig. 11. The differences are likely dominated by the differences between the LTE (this study) and NLTE modeling. See the text for details.

the reduced χ^2 as a function of $T_{\rm eff}$ and S/N. We then defined a flag as "chi2ratio" in the catalog to describe the deviation of the reduced χ^2 for individual stars from the median value of stars with similar S/N and $T_{\rm eff}$, divided by the dispersion. We modeled the reduced χ^2 for both the median and dispersion as a function of $T_{\rm eff}$ and S/N with a three-order polynomial. In the analysis below, we discarded the stars with reduced χ^2 larger than 10σ from the median value ("chi2ratio > 10"). We did not adopt a constant χ^2 cut because the typical χ^2 of the spectral fitting is found to vary with the $T_{\rm eff}$ and S/N of the spectra, due to both imperfect flux uncertainty estimates of the LAMOST spectra (i.e., overestimating the flux uncertainty for low-S/N stars or underestimating the flux uncertainty for high-S/N stars). In either case, they can lead to a S/N-dependent χ^2 distribution. Furthermore, since the model accuracy varies as a function of $T_{\rm eff}$, for instance, the Kurucz spectra are expected to be less accurate for very hot ($T_{\rm eff} > 25\,000$ K) stars, modeling chi2ratio as a function of $T_{\rm eff}$ is to ensure that the χ^2 criterion is adjusted according to the model accuracy.

5.3. $T_{\rm eff}$ and $\log g$

Figure 13 presents the distribution of our sample stars with $S/N > 30$ in the $T_{\rm eff}$–$\log g$ diagram overplotted with the MIST stellar evolutionary tracks (Paxton et al. 2011; Choi et al. 2016), the numbers marked in red are the initial mass of the stellar evolutionary tracks. The majority of the sample stars have relatively cool temperatures, corresponding to stellar masses of 1.5–7.5 $M_\odot$. While there are also a number of stars whose locations in the diagram are consistent with stellar mass higher than 20 $M_\odot$. At the high-temperature end, there are a group of stars with $\log g \gtrsim 5$, which are outside the coverage of the stellar evolution tracks. These stars are mostly hot subdwarfs (e.g., sdBs and sdOs), whose origin is still a research topic (e.g., Webbink 1984; Han et al. 2002, 2003; Lanz et al. 2004; Miller Bertolami et al. 2008; Heber 2009; Justham et al. 2011; Zhang & Jeffery 2012). Hot subdwarfs have been extensively identified from the LAMOST database by Luo et al. (2016, 2019, 2020). The positions of the hot subdwarfs in the $T_{\rm eff}$–$\log g$ plane are consistent

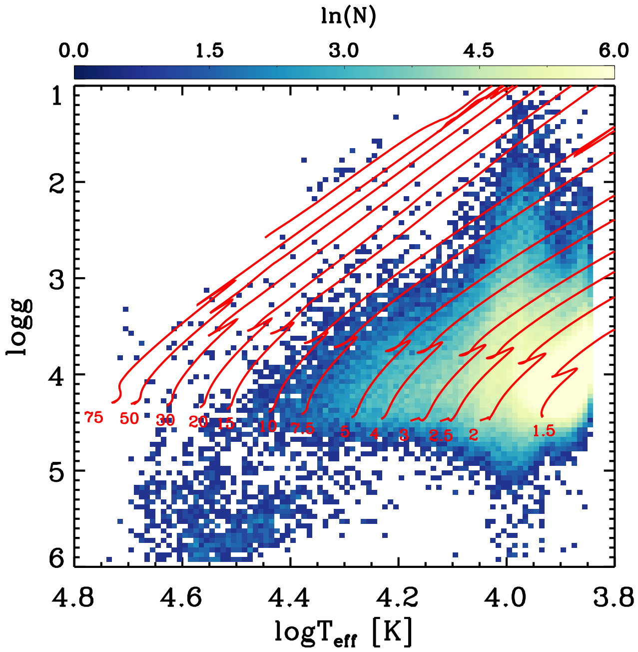

Fig. 13. Stellar number density distribution in the Kiel ($T_{\rm eff}$–$\log g$) diagram. Only 211 245 stars with $S/N > 30$ are shown. Also shown in the figure are the MIST stellar evolution tracks (Paxton et al. 2011; Choi et al. 2016) with [Fe/H] = −0.5 and $V/V_{\rm c} = 0.4$. The initial stellar masses of the tracks are marked (in unit of $M_{\odot}$). The stars with $\log g > 5$ and $\log T_{\rm eff} > 4.2$ are hot subdwarfs. The vertical stripe at $\log T_{\rm eff} \sim 4.0$ is an artifact for chemically peculiar stars (e.g., Ap/Bp stars and Am stars) whose $\log g$ have been underestimated as a consequence of masking the hydrogen lines for label determination.

with Luo et al. (2016, e.g., see their Fig. 4), which is remarkable since our $\log g$ estimates for these stars are obtained via extrapolation of HotPayne. Interestingly, Fig. 13 exhibits an upper border of the $\log g$ values that increase with $T_{\rm eff}$ for the subdwarfs. A detailed study for the subdwarfs' distribution morphology in the $T_{\rm eff}$-$\log g$ diagram is beyond the scope of this paper, but it deserves further dedicated work. It is also possible that a small number of these stars are "contamination" by unrecognized white dwarfs. Finally, we caution about the usage of our stellar labels for hot subdwarfs and white dwarfs, as they are estimated with model spectra that are extrapolations beyond the trained regime of HotPayne.

At $T_{\rm eff} \sim 10\,000$ K ($\log T_{\rm eff} \sim 4$), there is a vertical stripe of stars that extend to low $\log g$ values (<2). As we discuss further in Sect. 5.5, this is an artifact due to the existence of a large number of chemically peculiar stars, such as ApBp stars and Am stars. The $\log g$ of these stars may have been significantly underestimated as a consequence of masking the hydrogen lines for label determination for these early A-type and late B-type stars (Sect. 4.4). We note that such stars are prevalent in the hot-star regime; it is suggested that a high fraction of A-stars could be chemically peculiar (e.g., Gray et al. 2016; Qin et al. 2019; Xiang et al. 2020), and this is why their high number density constitutes a prominent stripe in the $T_{\rm eff}$–$\log g$ plane. Such an artifact reveals one of the limitations of the current model, which tends to yield incorrect $\log g$ for these chemically peculiar stars. However, as we show in Sect. 5.5, the abundance estimates of these stars are nonetheless robust enough, and thus these stars can be easily flagged through their abundance estimates.

About 98% of our sample stars have parallax measurements from the *Gaia* eDR3 (Gaia Collaboration 2021), which allow us to look into their luminosity (absolute magnitude). We derive their absolute magnitude in the 2MASS K-band (Skrutskie et al. 2006) using the distance modulus, adopting the *Gaia* eDR3 distance of Bailer-Jones et al. (2021). The interstellar extinction is corrected using the reddening $E(B-V)$ interpolated from the 3D reddening map of Green et al. (2019) and a K-band extinction coefficient of 0.34 (e.g., Yuan et al. 2013). Figure 14 presents the stellar distribution in the $T_{\rm eff}$–$M_{K{\rm s}}$ diagram for a subset of 195 879 stars with good spectral S/N ($S/N > 30$) and good parallax ($\omega/\sigma_{\omega} > 5$). The figure illustrates a good correlation between $M_{K{\rm s}}$ and $\log g$, suggesting our $\log g$ estimates are robust for the overall sample stars. The figure also illustrates that at a given $T_{\rm eff}$, the $v\sin i$ exhibits little correlation with $M_{K{\rm s}}$. Finally, while relatively cool stars exhibit a wide range of $v\sin i$ values, the majority of hot stars with $T_{\rm eff} \gtrsim 25\,000$ K tend to exhibit large $v\sin i$ values (except for hot subdwarfs).

In Fig. 15 we show the (BP–RP) color of *Gaia* eDR3 as a function of our $T_{\rm eff}$ estimates for stars with $S/N > 30$. There is a clear trend of (BP–RP) with Galactic latitude, which is due to the reddening effect. For stars at high Galactic latitude ($|b| > 30°$), there is a clear relation between $T_{\rm eff}$ and (BP–RP), which is consistent well with the $T_{\rm eff}$-dependent trajectory of synthetic color from the MIST isochrones. Stars at low Galactic latitude exhibit a much broader (BP–RP) distribution, due to their large reddening effect. The right panel of Fig. 15 presents similar results, but correcting the color for interstellar reddening $(BP–RP)_0$ estimated using the 3D reddening map of Green et al. (2019) and a total-to-selective extinction coefficient $R_{BP} = 3.24$ and $R_{RP} = 1.91$ (Chen et al. 2019). After de-reddening, the colors of the stars agree approximately with the expectation from the MIST isochrones. Interestingly, even after correcting for the reddening, a large fraction of hot stars with $T_{\rm eff} \gtrsim 25\,000$ K appear to be too red, with $(BP–RP)_0 \sim 0.0$. These stars are likely either associated with star formation regions (for young stars) or have dust envelopes (for hot subdwarfs), as they suffer (locally) larger reddening than the Green et al. (2019) 3D "average" dust map (see also Xiang et al. 2021). The NLTE effects should push the temperature even higher than the current estimates and would only further exacerbate this discrepancy.

5.4. $v\sin i$

The $v\sin i$ of our sample stars with $S/N > 50$ and $\log g < 5$ are presented in Fig. 16. Here we adopt a stricter S/N cut than above for the $T_{\rm eff}$-$\log g$ plot ($S/N > 30$) to ensure the robustness of the $v\sin i$ for the stars presented in the figure. The $v\sin i$ estimates spread a broad range of values, from no rotation to faster than 600 km s^{-1}. We note that our $v\sin i$ estimates could have negative values from the extrapolation of the model spectra beyond the well-trained regime of HotPayne. A negative $v\sin i$ simply means that the star has little or no rotation. In the final catalog, we set all negative $v\sin i$ estimates to be 0 km s^{-1}, but in Fig. 16 we keep the negative estimated values as they can serve as a good proxy for the uncertainties of the $v\sin i$ estimates. The extent of the negative tail suggests that the measurement uncertainties are on the order of 30 km s^{-1}, which is in agreement with validation using the repeat observations (Fig. 10). We note that this number does not reflect any possible systematic errors. For stars of $T_{\rm eff} < 10\,000$ K, the $v\sin i$ distribution peaks at ∼30 km s^{-1}, with a significant high-velocity tail extending to ∼300 km s^{-1}, above which the number density free falls. The peak value shifts to slightly larger values for hotter stars, reaching ∼50 km s^{-1} for $T_{\rm eff} > 16\,000$ K. This is qualitatively consistent with previous studies based on high-resolution spectroscopy, which have

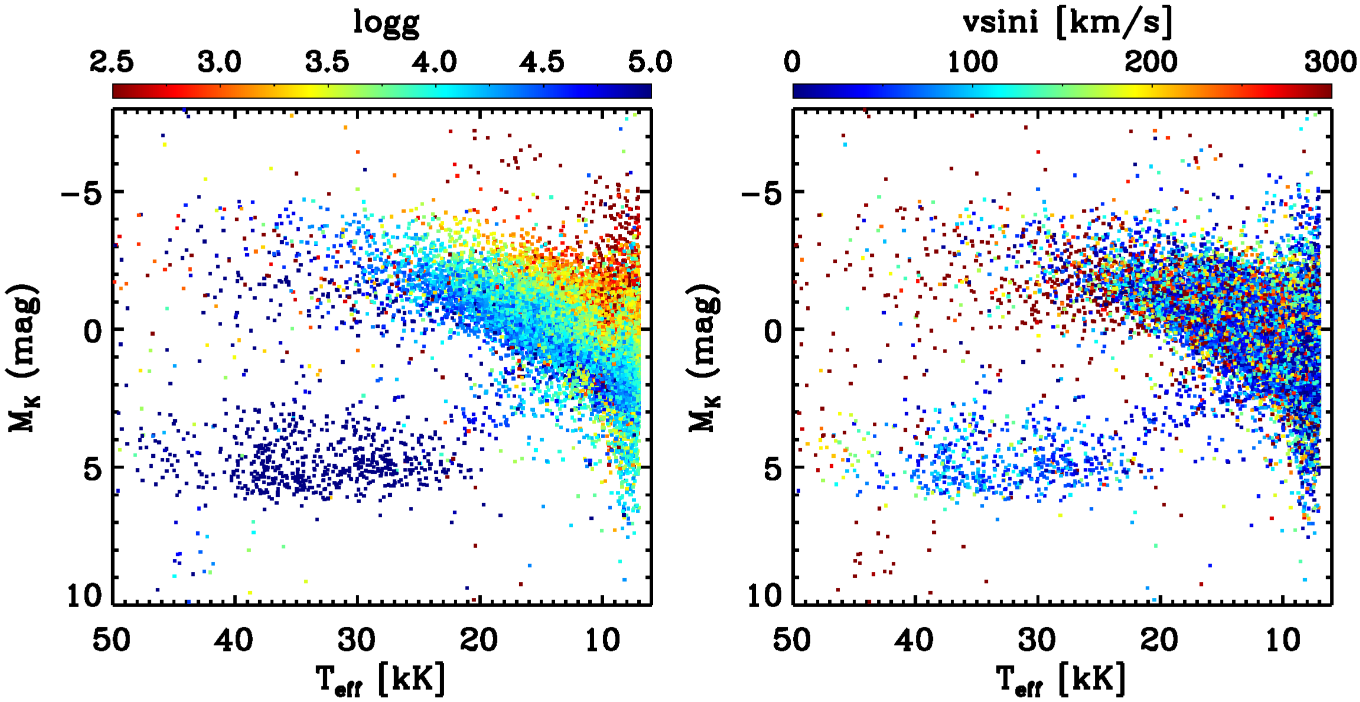

Fig. 14. Stellar distribution in the $T_{\rm eff}$–$M_{K{\rm s}}$ diagram color-coded by $\log g$ (*left*) and $v\sin i$ (*right*). Only 194 724 stars with good spectral S/N ($S/N > 30$) and good parallax ($\omega/\sigma_\omega > 5$) are shown. The expected correlation between $\log g$ with $M_{K{\rm s}}$ at a given $T_{\rm eff}$ demonstrates the robustness of the spectroscopic $\log g$ estimates. The stars with $T_{\rm eff} \gtrsim 20\,000$ K and $M_{K{\rm s}} \sim 5$ mag are hot subdwarfs. In the *left panel*, a maximal $\log g$ value of 5 is displayed in the color bar, and all stars with $\log g > 5$ are shown with the same color as $\log g = 5$ (dark blue).

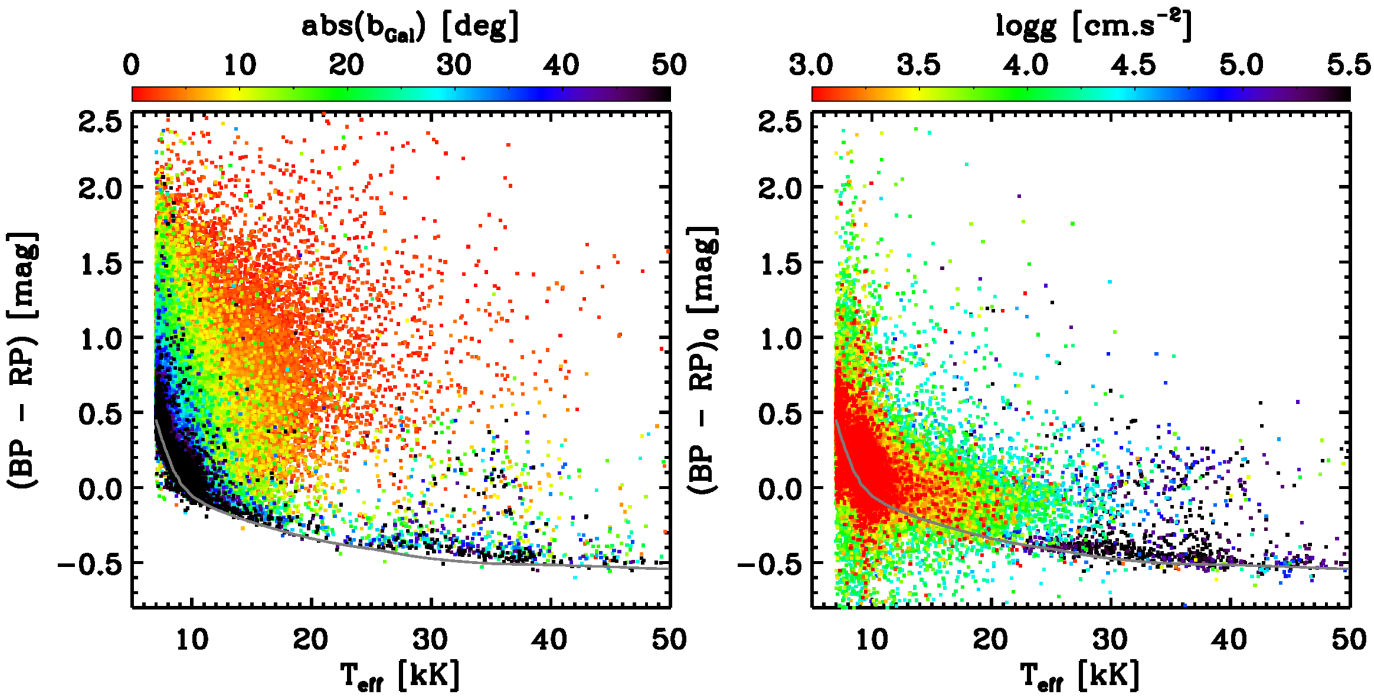

Fig. 15. Observed and de-reddened (BP–RP) color as a function of effective temperature. Only stars with spectral $S/N > 30$ are shown. The *left panel* shows the observed (BP–RP) color vs. $T_{\rm eff}$, with the stars color-coded by Galactic latitude as a rough proxy for the expected dust reddening. The gray line delineates the MIST isochrones' prediction for (BP–RP) color as a function of $T_{\rm eff}$ for ZAMS stars with initial [Fe/H] = 0. This panel illustrates that our sample stars in low Galactic latitudes suffer from a large reddening effect, leading to much redder colors than stars at high Galactic latitudes. The *right panel* shows $(BP{-}RP)_0$, the de-reddened color, based on the 3D reddening map of Green et al. (2019), with the stars sorted and color-coded by $\log g$. The panel illustrates that such de-reddened colors qualitatively agree with the isochrones, but the significant spread in color at any given temperature implies that the photometric colors alone would be too imprecise to estimate $T_{\rm eff}$ robustly for stars above 10 000 K. The very hot stars with little reddening and very high $\log g$ are presumably near subdwarf B stars of modest luminosity (see Fig. 14).

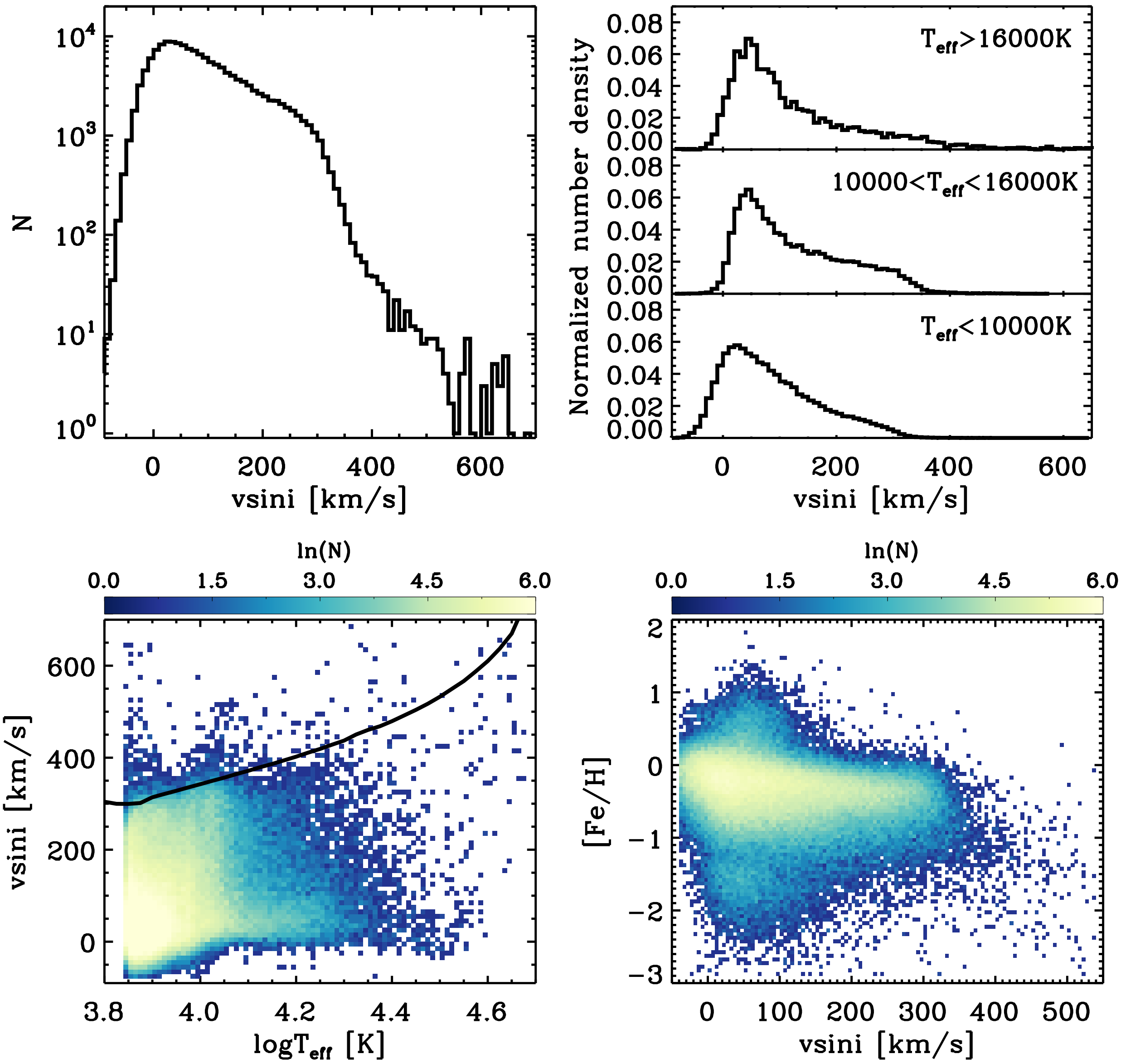

Fig. 16. Distribution of $v\sin i$. *Top left*: histogram of $v\sin i$ for the overall sample in this work. Only results from stars with $S/N > 50$ and $\log g < 5$ are shown. *Top right*: histogram of $v\sin i$ for stars in different $T_{\rm eff}$ ranges, as marked in the figure. *Bottom left*: stellar number density distribution in the $T_{\rm eff}$–$v\sin i$ plane. The solid line delineates the critical rotation velocity, $V_{\rm c}$, as a function of $T_{\rm eff}$, derived based on the ZAMS MIST isochrones (Choi et al. 2016). *Bottom right*: stellar number density distribution in the [Fe/H]–$v\sin i$ plane. Since we do not impose a prior limit on $v\sin i$, our fit allows for negative $v\sin i$ values. A negative $v\sin i$ simply means that the star has a small rotation velocity, and the distribution at the negative tail serves as a proxy for the measurement uncertainties.

suggested that most of the early-type stars in the nearby Galactic disk rotate slower than 300 km s^{-1}, although a minor number of them can rotate faster than 400 km s^{-1} (e.g., Abt et al. 2002; Huang et al. 2010; Bragança et al. 2012; Zorec & Royer 2012; Simón-Díaz & Herrero 2014; Li 2020).

Our results show that below 300 km s^{-1}, the stellar number density exhibits a descending power-law. Abt et al. (2002) suggested that the $v\sin i$ distribution of late B-type (B8-B9.5 III-V) stars exhibits double peaks, contributed by a set of slow rotating chemically peculiar stars and a set of fast rotating normal stars (see also, e.g., Dufton et al. 2013). We examine this idea via the $v\sin i$ distribution for our sample stars in different $T_{\rm eff}$ bins. The middle of the top-right panels of Fig. 16 presents the results for stars with $10\,000 < T_{\rm eff} < 16\,000$ K, which are composed of late B-type stars. It shows a prominent peak of slow rotating stars at $v\sin i \sim 30$ km s^{-1}. While our results do not exhibit a clear double-peak feature, it indeed shows a significant tail of fast rotating stars (up to $v\sin i \sim 300$ km s^{-1}). The reason for the difference is unclear yet, and deserves to be further understood. This could be either because the relatively large measurement uncertainty of our estimates has smeared an intrinsic double-peak feature or because the intrinsic strength of the second peak is weaker than presented in previous work, particularly considering that the sample size in the literature is small (1/50 of the current sample). We note that $v\sin i$ estimates from medium-resolution ($R = 7500$) spectra of LAMOST for 40 034 early-type stars have recently become available (Sun et al. 2021). From this data set with better precision in $v\sin i$ estimates, we confirmed the missing of a clear double-peak feature as found here. Independent of the temperature, our results show that slow rotators ($\lesssim$100 km s^{-1}) dominate over fast rotators for all stellar types from 7000 K ($\log T_{\rm eff} \sim 3.85$) to 25 000 K ($\log T_{\rm eff} \sim 4.4$) (see the bottom-left panel of Fig. 16). The bottom-left panel of Fig. 16 also shows a significant fraction of slow rotators even above 25 000 K ($\log g > 4.4$). This can also be seen in the right panel of Fig. 14, which displays numerous hot ($T_{\rm eff} > 25\,000$ K), bright ($M_{K{\rm s}} < 0$) stars with $v\sin i < 200$ km s^{-1}.

The bottom-left panel of Fig. 16 illustrates that at the high-velocity end, the stellar density exhibits a sharp cutoff, and the cutoff velocity increases with $T_{\rm eff}$. The cutoff is particularly prominent for stars with $T_{\rm eff} < 20\,000$ K ($\log T_{\rm eff} \sim 4.3$), above which the cutoff becomes less well defined due to the sparsity of stars. Theoretically, a star cannot be formed with infinity rotation velocity because the rotation will induce centrifugal force, which increases with the rotation velocity. Once the centrifugal force is larger than the gravity, the rotating disk that accretes materials and angular momentum to sustain the star formation will break up. This causes an upper limit of the rotation velocity, which is referred as the critical velocity, or the break-up velocity (e.g., Maeder 2009). The critical velocity can be expressed as

$$V_c = \sqrt{\frac{2}{3}\frac{GM}{R_{\rm p}}}, \tag{7}$$

where G is gravity constant, M the stellar mass of a zero-age main-sequence (ZAMS) star, $R_{\rm p}$ the radius at the pole of the star. We assume that the $R_{\rm p}$ is approximately the same as the stellar radius in spherical models. In the bottom-left panel of Fig. 16, we show the critical velocity as a function of effective temperature. Here the critical velocity is derived using ZAMS stars with solar abundances. We retrieve the parameters (mass, $T_{\rm eff}$, and radius) of the ZAMS stars from the public MIST stellar isochrones (Paxton et al. 2011; Choi et al. 2016). Technically, for each $T_{\rm eff}$, we define the main-sequence star with the maximal $\log g$ as an approximation of the ZAMS star of that $T_{\rm eff}$.

It shows that, except for a few outliers, both the values and their $T_{\rm eff}$ dependence of the observed $v\sin i$ border are consistent with the theoretical critical velocity, which increases from about 300 km s^{-1} at $T_{\rm eff} = 7000$ K to >600 km s^{-1} at $T_{\rm eff} = 40\,000$ K ($\log T_{\rm eff} \sim 4.6$). This good agreement suggests that there are indeed a considerable fraction of stars rotating with a rate close to the critical velocity.

We note that some stars exhibit $v\sin i$ larger than the critical velocity. In principle, it is possible that a star might temporarily rotate faster than the critical velocity as a consequence of binary interaction (de Mink et al. 2013). However, upon inspecting the spectra and intrinsic brightness of these outliers, we find that they are mainly composed of nearly equal-brightness binary stars that exhibit large velocity difference between the individual components thus their spectral lines are broader than single stars. In addition, these outliers also include stars whose $v\sin i$ are erroneously estimated due to a number of reasons, such as the occurrence of strong emission lines, which is particularly common for O-type stars, the erroneously spectral radial velocity correction, the contamination of subdwarfs and white dwarfs, and the failure of the model spectra at the high-$T_{\rm eff}$ end. For stars with $T_{\rm eff} \gtrsim 25\,000$ K, the inaccurate line strength of the Kurucz LTE model spectra may cause problems for the $v\sin i$ estimates. For example, by matching the He line profile of LAMOST with NLTE model spectra, Li (2020) suggests that LAMOST J040643.69+542347.8 ($T_{\rm eff} \sim 35\,000$ K) has a $v\sin i$ of $\sim$540 km s^{-1}, while our estimate is 261 ± 11 km s^{-1}, significantly underestimated due to the poor reproduction of the spectral line profile by our LTE model spectra at such high $T_{\rm eff}$.

The bottom-right panel of Fig. 16 presents the relation between $v\sin i$ and [Fe/H]. While there is no strong trend between $v\sin i$ and [Fe/H] for the majority of stars, a small group of stars with $v\sin i \lesssim 150$ km s^{-1} exhibit high metallicity ([Fe/H] $\gtrsim 0.2$). These stars are likely magnetic ApBp stars. We elaborate further on this point in the next section. On the metal-poor side, there is also an excess of stars with slow rotation velocities. These are found to be mostly BHB stars, presumably in the halo; due to their old age, they are rotating much more slowly than the typical young hot stars in the disk.

5.5. Abundance [Fe/H] and [Si/H]

In light of the large systematics and uncertainties of our LTE-based abundance estimates, we opt to present only the [Fe/H] and [Si/H] in our catalog, while leaving the other abundances to future work. The [Fe/H] is an essential metallicity indicator for relatively cool stars ($T_{\rm eff} \lesssim 10\,000$ K) both because the [Fe/H] of cool stars can have a broad range of values, as they come from different Galactic components (e.g., disk and halo), and because [Fe/H] can be robustly estimated for cool stars owing to their prominent features in the spectra. Silicon is a representative α-element. While the [Si/H] estimates may have large systematics (Fig. 12), we expect their relative values are robust and informative for identifying chemically peculiar stars (Fig. 17).

The left panel of Fig. 17 shows the stellar distribution in the $T_{\rm eff}$–[Fe/H] plane. A prominent feature in the figure is the v-shape pattern of [Fe/H] – typical [Fe/H] value for the bulk of stars decreases from -0.2 dex at the cool end ($T_{\rm eff} \sim 7000$ K) to a minimal value of about -0.7 dex at $T_{\rm eff} \sim 17\,000$ K, beyond which the [Fe/H] slightly increases with $T_{\rm eff}$. This v-shape trend seems to be consistent with the high-resolution results of Gebruers et al. (2021). The origin of this trend is unclear. It is possibly an artifact due to NLTE effects. Beyond this trend, there is a group of stars with super-solar metallicity at $T_{\rm eff} < 20\,000$ K, which are clearly separated from the bulk of stars in the [Fe/H]–$T_{\rm eff}$ plane. For many of these Fe-rich stars, the [Si/H] are also significantly enhanced (see the right panel of Fig. 17). The high [Si/H] nature suggests these stars are likely chemically peculiar stars whose origin may be related to a strong magnetic field: magnetic ApBp stars (e.g., Abt et al. 1972, 2002; Preston 1974; Smith 1996; Mathys et al. 1997; Michaud et al. 2015; Hümmerich et al. 2020). These stars have been shown to be slow rotators with $v\sin i < 150$ km s^{-1} (Fig. 16), which is consistent with previous results (e.g., Abt et al. 2002; Preston 1974; Michaud et al. 2015). Besides the ApBp stars, our sample may also contain a significant number of other types of chemically peculiar stars (such as Hg-Mn stars, Am/Fm stars, He-deficient stars, etc.) as our sample stars are uniformly targeted without significant bias against any of these stars. Figure 17 shows that at the cool temperature side ($T_{\rm eff} \lesssim 10\,000$ K), there is a large fraction of stars with super-solar metallicity, forming the high-[Fe/H] tail of the [Fe/H] distribution. Many of these super-solar metallicity stars are probably AmFm stars, which have been demonstrated to contributed a large fraction ($\sim$40%) of the A/F stars with mass higher than 1.4 $M_\odot$ in the LAMOST sample (Xiang et al. 2020).

On the metal-poor side of the $T_{\rm eff}$–[Fe/H] plane, there is a significant number of metal-poor stars with [Fe/H] $\lesssim -1.0$ dex, with a lower [Fe/H] border increasing from ~ -2.5 dex at $T_{\rm eff} = 7000$ K to ~ -1.5 dex at $T_{\rm eff} = 11\,000$ K. These metal-poor stars are mostly BHB stars and blue straggler stars of Galactic halo and thick disk populations; we expect there are very few single main-sequence stars given their high temperatures. As a demonstration, Fig. 18 shows that stars with $|b| > 15°$

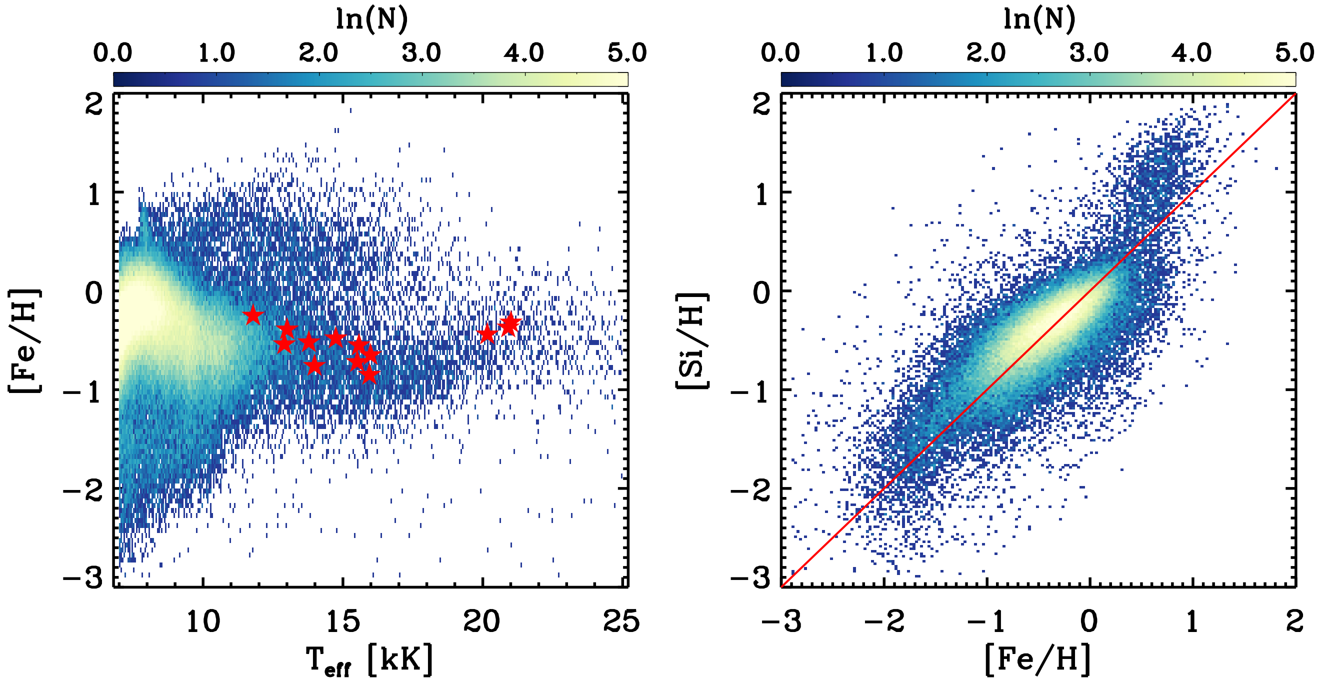

Fig. 17. Density distribution of our sample stars with $S/N > 50$ in the $T_{\rm eff}$ – [Fe/H] (*left*) and [Fe/H] – [Si/H] (*right*) planes. In the *left panel*, the stars in red show the SPB stars of Gebruers et al. (2021), with [Fe/H] from high-resolution spectroscopy (LTE-based). While the $T_{\rm eff}$-dependent trend of the [Fe/H] distribution is likely due to NLTE effects, the figure clearly reveals the presence of chemically peculiar stars with super-solar [Fe/H] values, as well as halo BHB stars at the metal-poor end. The *right panel* shows good consistency between [Si/H] and [Fe/H], verifying the robustness of the determinations (regardless of the systematics due to NLTE). The group of stars with [Fe/H] $\sim$ 0.7 dex and [Si/H] $\gtrsim$ 1 dex are magnetic ApBp stars, as indicated by their signature enhancement in Si (e.g., Michaud et al. 2015). In the *right panel*, we omit stars with $T_{\rm eff}$ < 8000 K because their [Si/H] estimates are found to suffer from larger systematics.

exhibit a [Fe/H] peak at around ~ -1.6 dex, which is consistent with previous [Fe/H] estimates of the nearby halo stars (e.g., Carollo et al. 2007; Beers et al. 2012; An et al. 2013; Zuo et al. 2017; Youakim et al. 2020). It is expected that the nearby halo stars dominate our metal-poor sample as the majority of them are brighter than 14 mag (Fig. 2). Figure 18 shows that the BHB stars can be well separated from the blue stragglers in the $T_{\rm eff}$–$\log g$ diagram. The BHB stars in our sample can have a temperature higher than 10 000 K stars, and their distribution in the $T_{\rm eff}$–$\log g$ diagram is consistent with the prediction of the MIST evolutionary tracks. Whereas, blue stragglers, the product of binary mass transfer (Chen & Han 2008a,b), occupy a different locus in the $T_{\rm eff}$–$\log g$ plane as they have larger $\log g$ values than (and therefore are separable from) BHB. This makes it possible to identify a large set (~5000) of BHB stars (and blue straggler stars, as well), which have been widely used as tracers to study the structure and dynamics of the Milky Way halo (e.g., Sirko et al. 2004b,a; Kinman et al. 2007; Xue et al. 2008; De Propris et al. 2010; Ruhland et al. 2011; Kafle et al. 2012; Yang et al. 2019; Starkenburg et al. 2019; Bird et al. 2021), from the current sample.

Our results also reveal that most of these metal-poor stars are slow rotators with $v\sin i \lesssim 150\,{\rm km\,s^{-1}}$ (Fig. 16). Among them, we find that the BHB stars have a $v\sin i$ distribution of $49 \pm 36\,{\rm km\,s^{-1}}$, which is qualitatively consistent with literature values (Peterson et al. 1995; Cohen & McCarthy 1997; Behr 2003; Recio-Blanco et al. 2004; Cortés et al. 2009), taking into account the relatively large uncertainty in our $v\sin i$ estimates.

The right panel of Fig. 17 illustrates that for the majority of stars of $T_{\rm eff}$ > 8000 K, the [Si/H] estimates are consistent with the [Fe/H] estimates, while at the metal-poor side (e.g., [Fe/H] < −1.0 dex), the [Si/H] is about 0.2 dex higher than the [Fe/H], which means that the metal-poor halo stars have [Si/Fe] $\sim$ 0.2 dex. This is consistent with the [α/Fe] values of the Galactic halo stars in previous work (e.g., Hayes et al. 2018; Conroy et al. 2019). We caution about the [Si/H] estimates for stars with $T_{\rm eff}$ < 8000 K, which we find much larger uncertainties, probably due to the weak features of Si in those relatively cool stars.

5.6. The parameter catalog

Our stellar parameter catalog contains $T_{\rm eff}$, $\log g$, $v\sin i$, [Fe/H], and [Si/H] for 332 172 unique stars with $T_{\rm eff}$ > 7000 K and $S/N > 5$ from 454 693 spectra of LAMOST DR6. Among them, 59% are from both the LAMOST classification and the Zari et al. (2021) selection method; 29% are from the LAMOST classification only, and 11% are from the Zari et al. (2021) selection method only. Table 6 presents a description of the columns in our catalog. Here we summarize a few notes of caution when using the catalog.

First, for stars with $\log g > 5$, which are mostly hot subdwarfs, the labels should be used with cautious as they are determined with model spectra extrapolated by HotPayne. Second, we assigned a $v\sin i$ value of $0\,{\rm km\,s^{-1}}$ for stars for which our measurement gives negative values due to extrapolation. We also caution that although the $v\sin i$ estimates should be statistically robust, the uncertainty is large. An inspection of the spectra is necessary when focusing on the $v\sin i$ of individual stars, and this is particularly true for stars with unexpectedly large $v\sin i$ values as many of them have been found to be artifacts due to either binary effects or data imperfection.

Third, the labels are deduced based on LTE models. They may suffer systematics due to NLTE effects as well as stellar

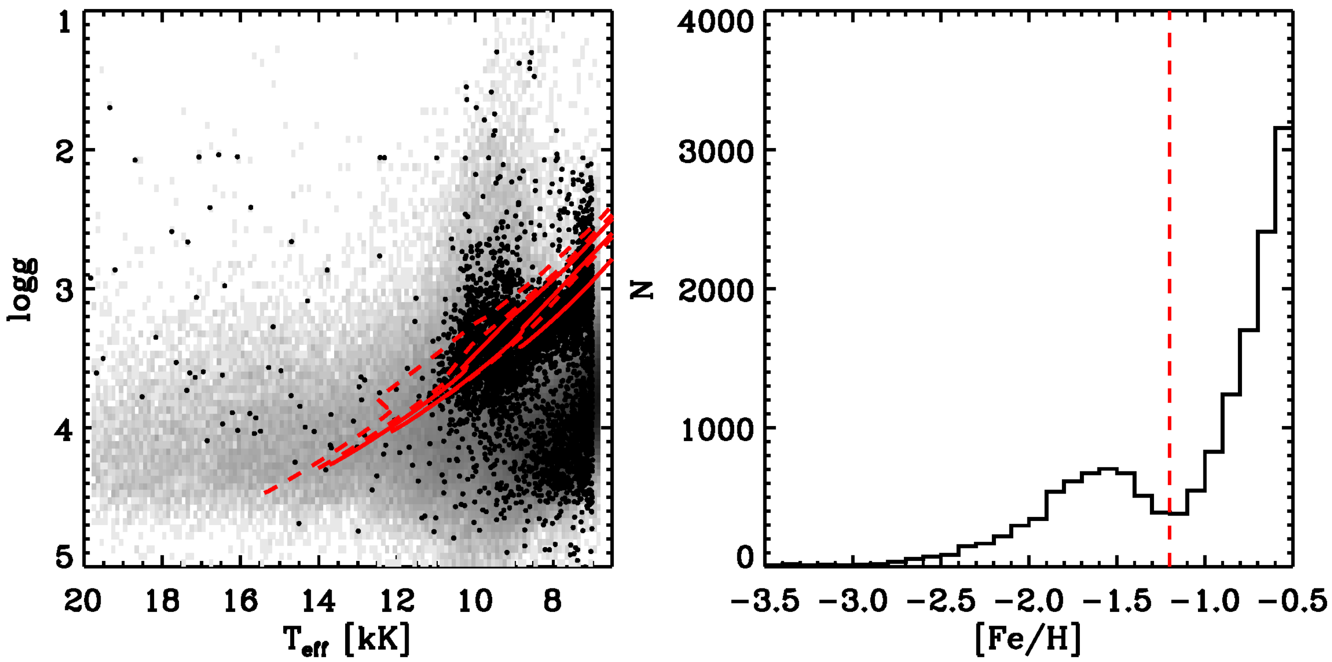

Fig. 18. Illustration of halo BHB stars revealed in the $T_{\rm eff}$–$\log g$ diagram. *Left*: Background in gray scale showing the stellar density (in logarithmic scale) of our sample stars with $S/N > 50$. The black dots represent the stars with $|b| > 15°$ and $[Fe/H] < -1.2$ dex. There are two sequences separated in $\log g$: the upper sequence with lower $\log g$ values is composed of halo BHB stars, and the lower sequence with larger $\log g$ values is presumed to comprise blue straggler stars. The solid lines in red are MIST evolutionary tracks of He-burning stars with $[Fe/H] = -1.5$, $V/V_c = 0.4$, and masses of 0.65, 0.60, and 0.58 $M_\odot$, respectively, from upper to lower. The dashed lines in red show similar tracks but for $[Fe/H] = -2.0$. *Right*: [Fe/H] distribution of our sample stars with $S/N > 50$ and $|b| > 15°$. Only the metal-poor side of $[Fe/H] < -0.5$ dex is shown. The vertical dashed line indicates a constant [Fe/H] of -1.2 dex, the sample below which mostly comprises the metal-poor halo BHB stars.

wind effects, particularly for stars with $T_{\rm eff} > 25\,000$ K and for stars with low $\log g$ values.

Finally, the label uncertainties presented in the catalog are the formal errors of the fitting and have been scaled to match the dispersion of repeat observations as a function of $T_{\rm eff}$ and S/N with a three-order polynomial, similar to in Xiang et al. (2019). The uncertainties provided here only concern the internal precision: they do not account for the zero-point offset induced by imperfections of the adopted model spectra.

5.7. Future improvements

This is our first attempt at deriving stellar labels for hot stars from a vast set of low-resolution $R \sim 1800$ spectra, for which most of the spectral features are very blended and weak, with an automated full-spectral fitting technique. Although we have demonstrated that the $R \sim 1800$ optical spectra do contain a wealth of information of hot stars that in principle can give us robust estimates of stellar parameters as well as abundances for many (>5) elements, this work contains a few imperfections that need to be improved in future work.

First, it is well known that NLTE effects (and wind effects) play an important role for hot-star spectra. To obtain accurate labels, particularly for elemental abundances, a large grid of NLTE spectral models are urgently needed. Secondly, it is well known that a large fraction of hot stars are in binary systems (e.g., Sana et al. 2012, 2013; Xiang et al. 2021), future spectroscopy should take the binary effect into account to derive reliable label estimates for binary systems. Similarly, many of the LAMOST O- and B-type stars (>10%) have emission lines in their spectra (e.g., Xiang et al. 2021). In the current work, we have neglected effects due to the emission lines. Therefore, for such systems, the stellar parameters might be problematic in unknown ways, adding uncertainty to stars hotter than 25 000 K. Thirdly, though in this work we have adopted the mean LSF of the LAMOST spectra, it is found that the fiber-to-fiber and temporal variations in the LAMOST LSF do have a significant impact on the $v\sin i$ estimates. Incorporating the exact LSF for individual stars should improve the $v\sin i$ measurements. Finally, we have adopted the stellar line-of-sight velocity from the LAMOST pipeline to correct the spectra into rest frame. However, we do find that, at times, the LAMOST pipeline provides erroneous line-of-sight velocities for some OB stars, particularly those with emission lines in their spectra. In future work, we may derive the stellar labels together with line-of-sight velocity simultaneously from the spectra.

6. Summary

In this study we have demonstrated that stellar atmospheric parameters and multiple elemental abundances of hot stars can be deduced from the low-resolution ($R \sim 1800$) optical spectra. This is first examined with the CR bound calculation, which suggests that for a typical B star with a spectral S/N of ∼100, the theoretical limit of the precision one could attain is 300 K in $T_{\rm eff}$, 0.03 cm s^2 in $\log g$, 1.0 km s^{-1} in $v_{\rm mic}$, 10 km s^{-1} in $v\sin i$, 0.01 in $N_{\rm He}/N_{\rm tot}$, ∼0.1 dex in abundances of C, O, Si, S, and Fe, and ∼0.2 dex in abundances of N and Mg.

We adopted PAYNE to derive 11 labels – $T_{\rm eff}$, $\log g$, $v_{\rm mic}$, $v\sin i$, $N_{\rm He}/N_{\rm tot}$, [C/H], [N/H], [O/H], [Si/H], [S/H], and [Fe/H] – from a vast set of LAMOST spectra, taking the LTE-based Kurucz synthetic spectra as the underlying model spectra. Testing with synthetic spectra and examining with duplicate

Table 6. Descriptions for the stellar label catalog of LAMOST DR6 hot stars.

Field	Description
Specid	LAMOST spectra ID in the format of "date-planid-spid-fiberid"
Ra	Right ascension from the LAMOST DR6 catalog (J2000; deg)
Dec	Declination from the LAMOST DR6 catalog (J2000; deg)
Uniqflag	Flag to indicate repeat visits; uniqflag = 1 means unique star, uniqflag = 2, 3, ..., n indicates the nth 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 fits header (km s^{-1})
rv_err	Uncertainty in radial velocity (km s^{-1})
T_{eff}	Effective temperature
T_{eff}_err	Uncertainty in T_{eff}
$\log g$	Surface gravity
$\log g$_err	Uncertainty in $\log g$
$v\sin i$	Projected rotation velocity
$v\sin i$_err	Uncertainty in $v\sin i$
[Fe/H]	Iron abundance
[Fe/H]_err	Uncertainty in [Fe/H]
[Si/H]	Silicon abundance
[Si/H]_err	Uncertainty in [Si/H]
corr_T_{eff}_$\log g$	Correlation coefficient between T_{eff} and $\log g$, derived from the covariance array of the fit
corr_T_{eff}_$v\sin i$	Correlation coefficient between T_{eff} and $v\sin i$
corr_T_{eff}_[Si/H]	Correlation coefficient between T_{eff} and [Si/H]
corr_T_{eff}_[Fe/H]	Correlation coefficient between T_{eff} and [Fe/H]
corr_$\log g$_$v\sin i$	Correlation coefficient between $\log g$ and $v\sin i$
corr_$\log g$_[Si/H]	Correlation coefficient between $\log g$ and [Si/H]
corr_$\log g$_[Fe/H]	Correlation coefficient between $\log g$ and [Fe/H]
corr_$v\sin i$_[Si/H]	Correlation coefficient between $v\sin i$ and [Si/H]
corr_$v\sin i$_[Fe/H]	Correlation coefficient between $v\sin i$ and [Fe/H]
corr_[Si/H]_[Fe/H]	Correlation coefficient between [Si/H] and [Fe/H]
chisq_red	Reduced χ^2 (i.e., χ^2/dof) of the fit
chi2ratio	Deviation of the chisq_red from the median value of chisq_red for stars of similar S/N and T_{eff}, divided by the dispersion of the chisq_red
source_dr6oba	Target from the LAMOST classification for OBA stars? (1 = 'YES', 0 = 'NO')
source_Zari21	Target from the photometry-based identification of Zari et al. (2021)? (1 = 'YES', 0 = 'NO')
source_Liu19	Target from the OB star catalog of Liu et al. (2019)? (1 = 'YES', 0 = 'NO')
gaia_id	Source ID of *Gaia* eDR3, matched with the CDS-Xmatch service (Boch et al. 2012) using 3″ criterion
parallax	*Gaia* eDR3 parallax
parallax_error	Uncertainty in parallax
pmra	*Gaia* eDR3 proper motion in RA direction
pmra_error	Uncertainty in pmra
pmdec	*Gaia* eDR3 proper motion in Dec direction
pmdec_error	Uncertainty in pmdec
RUWE	*Gaia* eDR3 RUWE
r_geo	Geometric distance derived from *Gaia* parallax by Bailer-Jones et al. (2021)
r_geo_low	Distance at 16th percentile of the probability distribution
r_geo_high	Distance at 84th percentile of the probability distribution
fidelity	Value of fidelity from Rybizki et al. (2022), an identifier for spurious astrometric solutions
ebv_bayes19	$E(B-V)$ interpolated from the 3D map of Green et al. (2019) using r_geo
ebv_pair	$E(B-V)$ estimated using intrinsic color from synthetic spectra of HOTPAYNE stellar parameters
ebv_pair_err	Uncertainty in ebv_pair
G/BP/RP	*Gaia* eDR3 magnitudes
J/H/K	2MASS photometric magnitudes
qual_2MASS	2MASS photometric quality flag
W1/W2/W3/W4	WISE photometric magnitudes
qual_wise	WISE photometric quality flag
var_wise	WISE varibility flag

Notes. The catalog is only available at the CDS via anonymous ftp to `cdsarc.u-strasbg.fr` (`130.79.128.5`) or via `http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/662/A66`

LAMOST observations confirms the robustness of the label estimates. We also show through the repeated spectra from the same objects that we can attain an internal precision that is consistent with the CR bound.

Despite the excellent internal precision and consistency, external examinations with literature values suggest that the abundance estimates may suffer non-negligible zero-point systematics, likely consequences of NLTE effects. Nonetheless, the current estimates of $T_{\rm eff}$, $\log g$, and $v\sin i$ are likely accurate, except for stars with $T_{\rm eff} > 25\,000$ K or supergiants with $\log g \lesssim 2$. Considering these nontrivial systematics, we present five stellar labels in our catalog, $T_{\rm eff}$, $\log g$, $v\sin i$, [Si/H], and [Fe/H], and can make other stellar labels available upon request.

In the $T_{\rm eff}$–$\log g$ diagram, our LAMOST sample stars cover a broad range of parameter space that corresponds to a stellar mass range from 1.5 $M_\odot$ to larger than 50 $M_\odot$. The $v\sin i$ distribution decreases roughly as a power law from a peak at ∼30 km s^{-1} to ∼300 km s^{-1}, above which the stellar density exhibits a cutoff. This terminate $v\sin i$ increases with $T_{\rm eff}$, from ∼300 km s^{-1} at 7000 K to ∼600 km s^{-1} at 40 000 K. This $T_{\rm eff}$-dependent cutoff is consistent with the theoretical prediction of the critical rotation velocity. The $v\sin i$ distribution also suggests that at all temperature from 7000 to 25 000 K, the stars are dominated by slow rotators with $v\sin i < 100$ km s^{-1}.

Our results reveal a sample of ApBp stars with significantly enhanced [Fe/H] ([Fe/H] > 0.2 dex) in the range of $7000 < T_{\rm eff} < 20\,000$ K, and many of them also have significantly enhanced [Si/H] ([Si/H] > 0.4 dex). These ApBp stars have rotation velocities lower than 150 km s^{-1}, which is consistent with previous studies. The label estimates also allow the identification of a large sample of BHB stars in the $T_{\rm eff}$–$\log g$ plane. Our catalog of $T_{\rm eff}$, $\log g$, $v\sin i$, [Si/H], and [Fe/H] for a sample of 332 172 stars with $T_{\rm eff} > 7000$ K from LAMOST DR6, including more than 64 000 OB stars with $T_{\rm eff} > 10\,000$ K, are publicly accessible.

Our analysis method and conclusion are generic and can be applied to other surveys, such as the SDSS-V survey. Future works based on NLTE analysis are desirable to improve upon the current results, particularly to obtain accurate stellar abundances from this vast set of LAMOST spectra, as well as the coming SDSS-V spectra.

Acknowledgements. The authors thank the reviewer, Prof. Artemio Herrero, for his carefully reading and valuable comments on the manuscripts. M. Xiang also thanks Prof. H.-W. Zhang for helpful discussion. 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. R.P.K. acknowledges support by the Munich Excellence Cluster Origins funded under germany's Excellence Strategy EXC-2094 390783311. S.G. gratefully acknowledges support from the Research Foundation Flanders (FWO) by means of a PhD Aspirant mandate under contract No. 11E5620N. The research leading to these results has (partially) received funding from the KU Leuven Research Council (grant C16/18/005: PARADISE), from the Research Foundation Flanders (FWO) under grant agreement G0H5416N (ERC Runner Up Project), and from the BELgian federal Science Policy Office (BELSPO) through PRODEX grant PLATO. This work has used data products from the Guoshoujing Telescope where the LAMOST survey was executed. 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 used data products from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* MultiLateral Agreement. The *Gaia* mission website is `https://www.cosmos.esa.int/gaia`. The *Gaia* archive website is `https://archives.esac.esa.int/gaia`. This work has used the ESO archive data, with the following project ids: 60.A-9022, 70.D-0191, 072.C-0488, 076.B-0055, 077.C-0379, 084.B-0869, 095.D-0234, 096.D-0008, 0101.C-0360, 0102.C-0124, 183.C-0972, 194.C-0833. This work has used data products from the 2MASS, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. This work has used the CDS-Xmatch service (`http://cdsxmatch.u-strasbg.fr/xmatch`). This work has also used the TOPCAT, IDL and Python software infrastructure.

References

Abt, H. A., Chaffee, F. H., & Suffolk, G. 1972, ApJ, 175, 779
Abt, H. A., Levato, H., & Grosso, M. 2002, ApJ, 573, 359
Aerts, C. 2021, Rev. Mod. Phys., 93, 015001
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
An, D., Beers, T. C., Johnson, J. A., et al. 2013, ApJ, 763, 65
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Auer, L. H., & Mihalas, D. 1972, ApJS, 24, 193
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
Beers, T. C., Carollo, D., Ivezić, Ž., et al. 2012, ApJ, 746, 34
Behr, B. B. 2003, ApJS, 149, 101
Berger, T. A., Kudritzki, R.-P., Urbaneja, M. A., et al. 2018, ApJ, 860, 130
Bird, S. A., Xue, X.-X., Liu, C., et al. 2021, ApJ, 919, 66
Boch, T., Pineau, F., & Derriere, S. 2012, ASP Conf. Ser., 461, 291
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977
Bragança, G. A., Daflon, S., Cunha, K., et al. 2012, AJ, 144, 130
Buder, S., Sharma, S., Kos, J., et al. 2021, MNRAS, 506, 150
Carlin, J. L., Lépine, S., Newberg, H. J., et al. 2012, Res. Astron. Astrophys., 12, 755
Carollo, D., Beers, T. C., Lee, Y. S., et al. 2007, Nature, 450, 1020
Casey, A. R., Hawkins, K., Hogg, D. W., et al. 2017, ApJ, 840, 59
Chen, X., & Han, Z. 2008a, MNRAS, 384, 1263
Chen, X., & Han, Z. 2008b, MNRAS, 387, 1416
Chen, B. Q., Liu, X. W., Yuan, H. B., et al. 2018, MNRAS, 476, 3278
Chen, B. Q., Huang, Y., Yuan, H. B., et al. 2019, MNRAS, 483, 4277
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Chojnowski, S. D., Hubrig, S., Hasselquist, S., et al. 2020, MNRAS, 496, 832
Cohen, J. G., & McCarthy, J. K. 1997, AJ, 113, 1353
Conroy, C., Naidu, R. P., Zaritsky, D., et al. 2019, ApJ, 887, 237
Cortés, C., Silva, J. R. P., Recio-Blanco, A., et al. 2009, ApJ, 704, 750
Cramér, H. 1945, Mathematical Methods of Statistics (Princeton: Princeton University Press)
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, Res. Astron. Astrophys., 12, 1197
De Cat, P., Fu, J. N., Ren, A. B., et al. 2015, ApJS, 220, 19
de Mink, S. E., Langer, N., Izzard, R. G., Sana, H., & de Koter, A. 2013, ApJ, 764, 166
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, Res. Astron. Astrophys., 12, 735
De Propris, R., Harrison, C. D., & Mares, P. J. 2010, ApJ, 719, 1582
Dotter, A. 2016, ApJS, 222, 8
Dufton, P. L., Langer, N., Dunstall, P. R., et al. 2013, A&A, 550, A109
Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, A&A, 537, A146
Fu, J.-N., Cat, P. D., Zong, W., et al. 2020, Res. Astron. Astrophys., 20, 167
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144
Gebran, M., Monier, R., Royer, F., Lobel, A., & Blomme, R. 2014, in Putting A Stars into Context: Evolution, Environment, and Related Stars, eds. G. Mathys, E. R. Griffin, O. Kochukhov, R. Monier, & G. M. Wahlgren, 193
Gebruers, S., Straumit, I., Tkachenko, A., et al. 2021, A&A, 650, A151
Gray, D. F. 1992, The Observation and Analysis of Stellar Photospheres (Cambridge: Cambridge University Press), 20
Gray, R. O., Corbally, C. J., De Cat, P., et al. 2016, AJ, 151, 13
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Hainich, R., Ramachandran, V., Shenar, T., et al. 2019, A&A, 621, A85
Han, Z., Podsiadlowski, P., Maxted, P. F. L., Marsh, T. R., & Ivanova, N. 2002, MNRAS, 336, 449
Han, Z., Podsiadlowski, P., Maxted, P. F. L., & Marsh, T. R. 2003, MNRAS, 341, 669
Haucke, M., Cidale, L. S., Venero, R. O. J., et al. 2018, A&A, 614, A91
Hayes, C. R., Majewski, S. R., Shetrone, M., et al. 2018, ApJ, 852, 49
Heber, U. 2009, ARA&A, 47, 211
Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2017, ApJ, 836, 5
Hoaglin, D. C., Mosteller, F., & Tukey, J. W. 1983, Understanding Robust and Exploratory Data Anlysis (Hoboken: John Wiley & Sons)
Hobbs, L. M., York, D. G., Snow, T. P., et al. 2008, ApJ, 680, 1256
Holgado, G., Simón-Díaz, S., Barbá, R. H., et al. 2018, A&A, 613, A65

Huang, W., Gies, D. R., & McSwain, M. V. 2010, ApJ, 722, 605
Hubeny, I., & Lanz, T. 1995, ApJ, 439, 875
Hubeny, I., & Lanz, T. 2011, Astrophysics Source Code Library [record ascl:1109.022]
Hümmerich, S., Paunzen, E., & Bernhard, K. 2020, A&A, 640, A40
Jorgensen, U. G., Larsson, M., Iwamae, A., & Yu, B. 1996, A&A, 315, 204
Justham, S., Podsiadlowski, P., & Han, Z. 2011, MNRAS, 410, 984
Kafle, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, ApJ, 761, 98
Kingma, D. P., & Ba, J. 2014, ArXiv e-prints [arXiv:1412.6980]
Kinman, T. D., Cacciari, C., Bragaglia, A., Buzzoni, A., & Spagna, A. 2007, MNRAS, 375, 1381
Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, ArXiv e-prints [arXiv:1711.03234]
Kudritzki, R. P. 1976, A&A, 52, 11
Kudritzki, R. P. 1979, Liege International Astrophysical Colloquia, 22, , 295
Kudritzki, R. P. 1998, in Stellar astrophysics for the local group: VIII Canary Islands Winter School of Astrophysics, eds. A. Aparicio, A. Herrero, & F. Sánchez, 149
Kudritzki, R.-P., & Puls, J. 2000, ARA&A, 38, 613
Kudritzki, R. P., Castro, N., Urbaneja, M. A., et al. 2016, ApJ, 829, 70
Kurucz, R. L. 1970, SAO Special Report, 309
Kurucz, R. L. 1993, SYNTHE Spectrum Synthesis Programs and Line Data (Cambridge, Mass.: Smithsonian Astrophysical Observatory)
Kurucz, R. L. 2005, Mem. Soc. Astron. It. Suppl., 8, 14
Lanz, T., & Hubeny, I. 2007, ApJS, 169, 83
Lanz, T., Brown, T. M., Sweigart, A. V., Hubeny, I., & Landsman, W. B. 2004, ApJ, 602, 342
Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, AJ, 136, 2022
Li, G.-W. 2020, ApJ, 892, L26
Liu, X. W., Yuan, H. B., Huo, Z. Y., et al. 2014, IAU Symp., 298, 310
Liu, Z., Cui, W., Liu, C., et al. 2019, ApJS, 241, 32
Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, Res. Astron. Astrophys., 15, 1095
Luo, Y.-P., Németh, P., Liu, C., Deng, L.-C., & Han, Z.-W. 2016, ApJ, 818, 202
Luo, Y., Németh, P., Deng, L., & Han, Z. 2019, ApJ, 881, 7
Luo, Y., Németh, P., & Li, Q. 2020, ApJ, 898, 64
Maeder, A. 2009, Physics, Formation and Evolution of Rotating Stars (Berlin: Springer)
Maeder, A., Przybilla, N., Nieva, M.-F., et al. 2014, A&A, 565, A39
Martins, F., Hervé, A., Bouret, J. C., et al. 2015, A&A, 575, A34
Masseron, T., Plez, B., Van Eck, S., et al. 2014, A&A, 571, A47
Mathys, G., Hubrig, S., Landstreet, J. D., Lanz, T., & Manfroid, J. 1997, A&AS, 123, 353
Michaud, G., Alecian, G., & Richer, J. 2015, Atomic Diffusion in Stars (Berlin: Springer)
Miller Bertolami, M. M., Althaus, L. G., Unglaub, K., & Weiss, A. 2008, A&A, 491, 253
Nieva, M. F., & Przybilla, N. 2012, A&A, 539, A143
Nieva, M.-F., & Przybilla, N. 2014, A&A, 566, A7
Paunzen, E., Hümmerich, S., & Bernhard, K. 2021, A&A, 645, A34
Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3
Pedersen, M. G., Aerts, C., Pápics, P. I., & Rogers, T. M. 2018, A&A, 614, A128
Peterson, R. C., Rood, R. T., & Crocker, D. A. 1995, ApJ, 453, 214
Preston, G. W. 1974, ARA&A, 12, 257
Przybilla, N., Butler, K., Becker, S. R., & Kudritzki, R. P. 2006, A&A, 445, 1099
Przybilla, N., Nieva, M.-F., & Butler, K. 2008, ApJ, 688, L103
Przybilla, N., Firnstein, M., Nieva, M. F., Meynet, G., & Maeder, A. 2010, A&A, 517, A38
Qin, L., Luo, A. L., Hou, W., et al. 2019, ApJS, 242, 13
Rao, C. R. 1945, Bull. Calcutta Math. Soc., 37, 81
Recio-Blanco, A., Piotto, G., Aparicio, A., & Renzini, A. 2004, A&A, 417, 597
Renson, P., & Manfroid, J. 2009, A&A, 498, 961
Ruhland, C., Bell, E. F., Rix, H.-W., & Xue, X.-X. 2011, ApJ, 731, 119
Rybizki, J., Green, G., Rix, H.-W., et al. 2022, MNRAS, 510, 2597
Sana, H., de Mink, S. E., de Koter, A., et al. 2012, Science, 337, 444
Sana, H., de Koter, A., de Mink, S. E., et al. 2013, A&A, 550, A107
Sandford, N. R., Weisz, D. R., & Ting, Y.-S. 2020, ApJS, 249, 24
Simón-Díaz, S., & Herrero, A. 2014, A&A, 562, A135
Sirko, E., Goodman, J., Knapp, G. R., et al. 2004a, AJ, 127, 899
Sirko, E., Goodman, J., Knapp, G. R., et al. 2004b, AJ, 127, 914
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smith, K. C. 1996, Ap&SS, 237, 77
Starkenburg, E., Youakim, K., Martin, N., et al. 2019, MNRAS, 490, 5757
Steinmetz, M., Guiglion, G., McMillan, P. J., et al. 2020, AJ, 160, 83
Sun, W., Duan, X.-W., Deng, L., et al. 2021, ApJS, 257, 22
Takeda, Y., Kawanomoto, S., & Ohishi, N. 2007, PASJ, 59, 245
Talon, S., Richard, O., & Michaud, G. 2006, ApJ, 645, 634
Ting, Y.-S., Conroy, C., & Rix, H.-W. 2016, ApJ, 826, 83
Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2017a, ApJ, 843, 32
Ting, Y.-S., Rix, H.-W., Conroy, C., Ho, A. Y. Q., & Lin, J. 2017b, ApJ, 849, L9
Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, ApJ, 879, 69
Vick, M., Michaud, G., Richer, J., & Richard, O. 2010, A&A, 521, A62
Webbink, R. F. 1984, ApJ, 277, 355
Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, MNRAS, 448, 822
Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2017, MNRAS, 467, 1890
Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, ApJS, 245, 34
Xiang, M.-S., Rix, H.-W., Ting, Y.-S., et al. 2020, ApJ, 898, 28
Xiang, M., Rix, H.-W., Ting, Y.-S., et al. 2021, ApJS, 253, 22
Xue, X. X., Rix, H. W., Zhao, G., et al. 2008, ApJ, 684, 1143
Yang, C., Xue, X.-X., Li, J., et al. 2019, ApJ, 886, 154
Youakim, K., Starkenburg, E., Martin, N. F., et al. 2020, MNRAS, 492, 4986
Yuan, H. B., Liu, X. W., & Xiang, M. S. 2013, MNRAS, 430, 2188
Yuan, H. B., Liu, X. W., Huo, Z. Y., et al. 2015, MNRAS, 448, 855
Zari, E., Rix, H. W., Frankel, N., et al. 2021, A&A, 650, A112
Zhang, X., & Jeffery, C. S. 2012, MNRAS, 419, 452
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, Res. Astron. Astrophys., 12, 723
Zorec, J., & Royer, F. 2012, A&A, 537, A120
Zuo, W., Du, C., Jing, Y., et al. 2017, ApJ, 841, 59

Appendix A: The gradient spectra of early A-type stars

The A0-type dwarfs have the strongest hydrogen lines in their spectra, which means that the gradient spectra $\partial f_\lambda/\partial\lambda$ of hydrogen lines reach a minimum for these stars. As an illustration, Fig. A.1 shows the gradient spectra of dwarf stars ($\log g = 4.5$, $[Fe/H] = 0$) at different temperatures. The gradient spectra of hydrogen lines have negative values at $T_{\rm eff} < 9500$ K, and become positive values at $T_{\rm eff} > 9500$ K. This reversal leads to the A0 "desert" shown in Fig. 9.

Appendix B: The Cramér-Rao bound of A stars

Figure B.1 shows the CR bound for an A-type dwarf star with $T_{\rm eff} = 9500$ K, $\log g = 4$, and $[Fe/H] = 0$, both from the full (unmasked) spectra and from the spectra after masking hydrogen lines as introduced in Sect. 4.5. Masking hydrogen lines loses part of the astrophysical information content, thereby increasing the CR bound. Nevertheless, ,the label estimates remain feasible even after applying the masking, with CR bounds of $\sim$100 K in $T_{\rm eff}$, 0.1 dex in $\log g$, 0.3 km s^{-1} in $v_{\rm mic}$, 15 km s^{-1} in $v\sin i$, and 0.1 dex in the abundance for a few elements (e.g., C, O, Mg, Fe) for a spectral S/N of 100.

Appendix C: Comparison between LTE and NLTE model spectra

Figures C.1 and C.2 present a comparison of the Kurucz LTE model spectra with NLTE spectra from the Potsdam Wolf-Rayet (PoWR) models (Hainich et al. 2019), for different temperatures. The NLTE effects (and stellar winds) strongly impact the helium lines at all temperatures, and have strong impacts on the hydrogen lines for stars with $T_{\rm eff} \gtrsim 25\,000$ K (see also, e.g., Auer & Mihalas 1972; Kudritzki 1976, 1979).

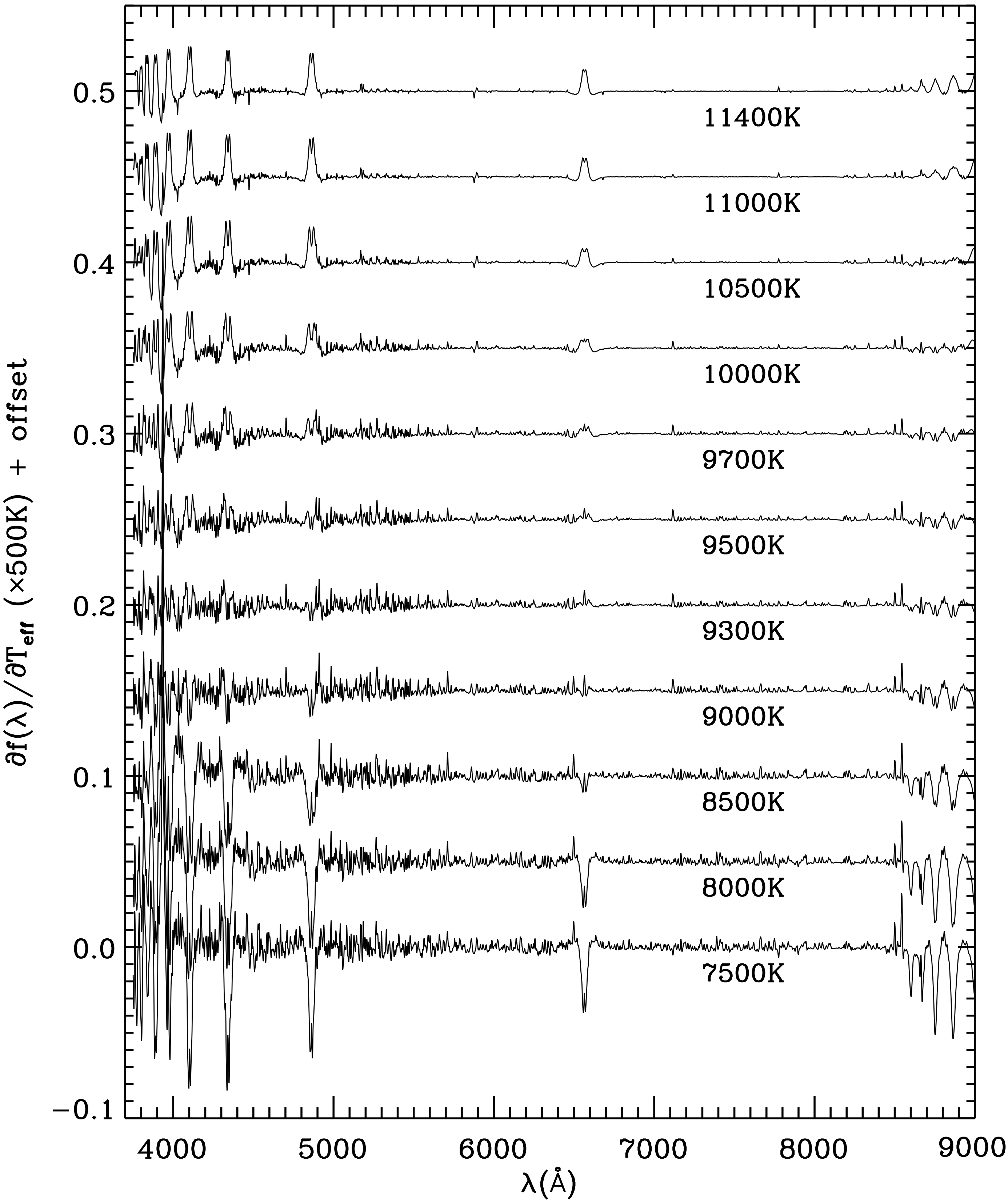

Fig. A.1. Gradient spectra of $T_{\rm eff}$ derived from the Kurucz model at different effective temperatures, as marked in the figure. We assume log $g = 4.5$ and [Fe/H] = 0. The A0 dwarf ($T_{\rm eff} = 9500$ K) has the minimal gradients in hydrogen lines (close to 0), and they form an inflection point in the gradient of the hydrogen lines as a function of $T_{\rm eff}$, which leads to the artificial A0 star desert of the stellar distribution in the $T_{\rm eff}$-log g diagram as seen in Fig. 9.

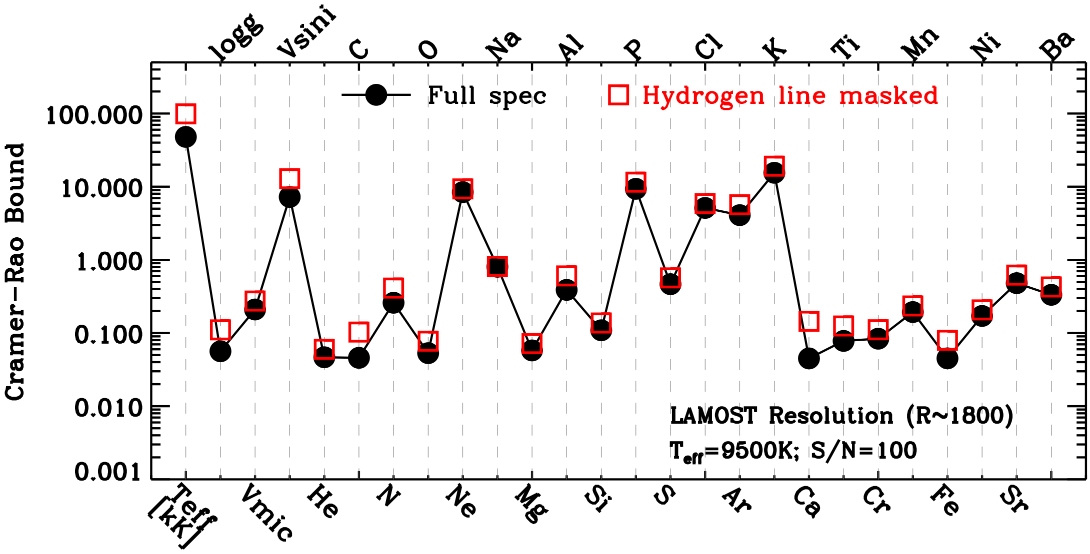

Fig. B.1. CR bound for an A-type star with $T_{\rm eff} = 9500$ K, $\log g = 4.5$, and [Fe/H] = 0. The black dots show the CR bound derived from the full spectra, and the red squares are derived from the spectra after masking the hydrogen lines as described in Sect. 4.5.

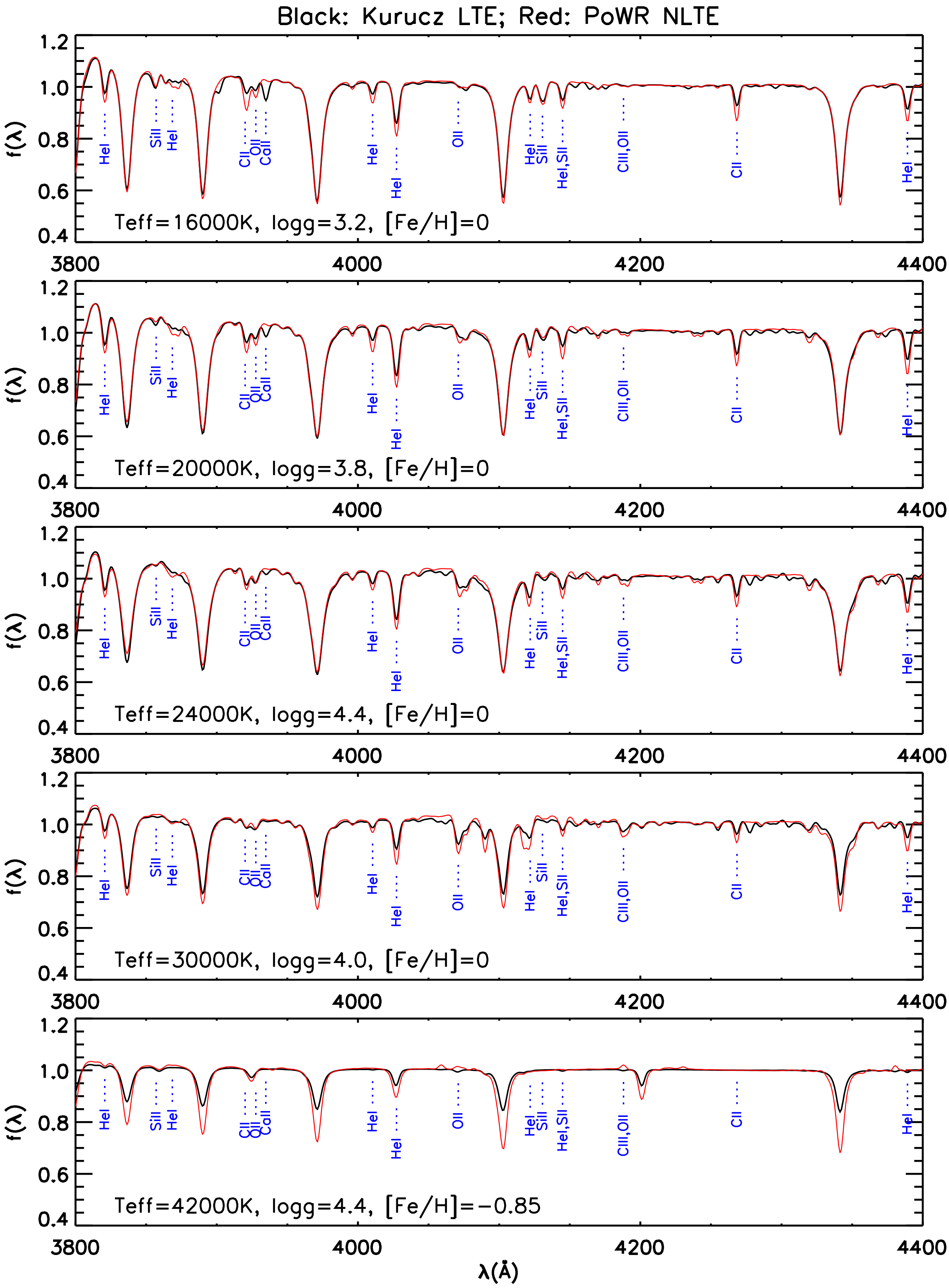

Fig. C.1. Comparison of Kurucz LTE model spectra (black) with the PoWR NLTE model spectra (red) for different parameters. The NLTE effect has a strong impact on the He lines and metal lines at all temperatures, and it also has a strong impact on the hydrogen lines for stars with $T_{\rm eff} \gtrsim 25\,000$ K.

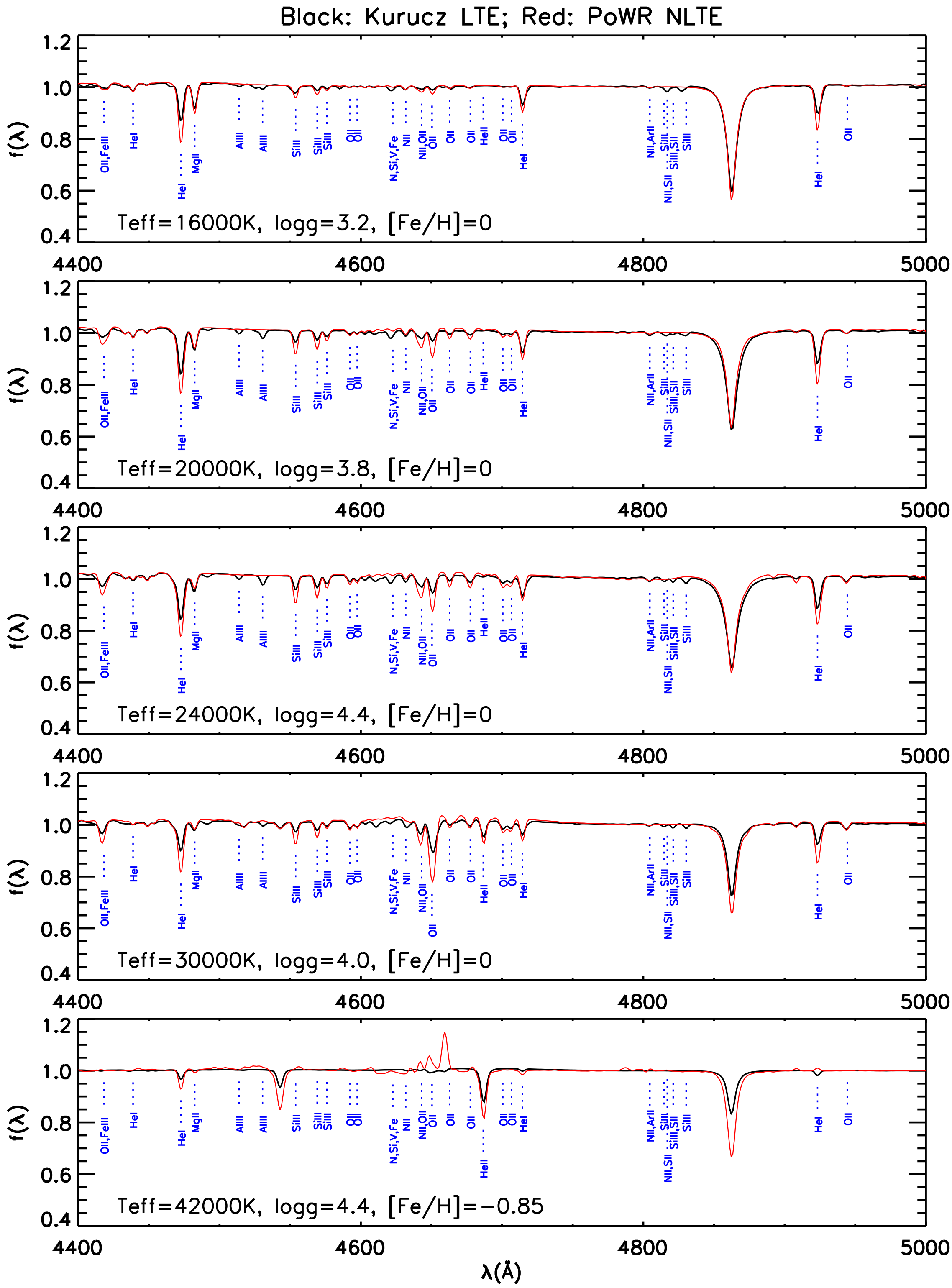

Fig. C.2. Same as Fig. C.1, but for the wavelength range λ4400–5500Å.