Research in Astronomy and Astrophysics, 23:015018 (15pp), 2023 January https://doi.org/10.1088/1674-4527/aca649

Correction Factors of the Measurement Errors of the LAMOST-LRS Stellar Parameters

Shuhui Zhang[1,2], Guozhen Hu[1,2], Rongrong Liu[1,2], Cuiyun Pan[1,2], Lu Li[1,2,3], and Zhengyi Shao[1,4]
[1] Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
zyshao@shao.ac.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Centre for Astrophysics and Planetary Science, Racah Institute of Physics, The Hebrew University, Jerusalem, 91904, Israel
[4] Key Lab for Astrophysics, Shanghai 200234, China
Received 2022 July 26; revised 2022 November 7; accepted 2022 November 18; published 2023 January 11

Abstract

We aim to investigate the propriety of stellar parameter errors of the official data release of the LAMOST low-resolution spectroscopy (LRS) survey. We diagnose the errors of radial velocity (RV), atmospheric parameters ([Fe/H], $T_{\rm eff}$, $\log g$) and α-enhancement ([α/M]) for the latest data release version of DR7, including 6,079,235 effective spectra of 4,546,803 stars. Based on the duplicate observational sample and comparing the deviation of multiple measurements to their given errors, we find that, in general, the error of [α/M] is largely underestimated, and the error of RV is slightly overestimated. We define a correction factor k to quantify these misestimations and correct the errors to be expressed as proper internal uncertainties. Using this self-calibration technique, we find that the k-factors significantly vary with the stellar spectral types and the spectral signal-to-noise ratio (S/N). Particularly, we reveal a strange but evident trend between k-factors and error themselves for all five stellar parameters. Larger errors tend to have smaller k-factor values, i.e., they were more overestimated. After the correction, we recreate and quantify the tight correlations between S/N and errors, for all five parameters, while these correlations have dependence on spectral types. It also suggests that the parameter errors from each spectrum should be corrected individually. Finally, we provide the error correction factors of each derived parameter of each spectrum for the entire LAMOST-LRS DR7 and plan to update them for the later data releases.

Key words: catalogs – methods: data analysis – stars: fundamental parameters

1. Introduction

Error measurements of stellar astrophysical parameters are equally important with the parameter's estimation themselves. In dealing with the vast amount of observational data, modern statistical approaches, such as the Bayesian Inference, require well defined and quantified parameter uncertainties in order to establish a fully Bayesian framework in subsequent investigations on the physical properties of targets.

In the field of deriving fundamental stellar parameters from a large spectroscopic survey, there are many works have also discussed the issues of the parameter errors. For instance, Ting et al. (2017) have theoretically analyzed the resource of the parameter uncertainties from the low resolution spectrum, and also reminded the dependence on the spectral type. Zhang et al. (2020) and Wang et al. (2022) also provide clear description of the uncertainties as a function of signal-to-noise ratio (S/N) in their works on the LAMOST spectra. Besides, Jofré et al. (2019) have summarized the latest human efforts to assess the accuracy and precision of industrial abundances by providing insights into the steps and uncertainties associated with the process of determining stellar abundances. In that review, they have emphasized that the parameter uncertainties need to be disentangled into different budgets: random uncertainty, systematic uncertainty, and systematic bias, and they could be considered separately or simultaneously.

The internal (random) uncertainty is one of the main indicators that evaluate the precision of parameter measurements. It is usually provided by the data reduction pipeline of astronomical surveys and is often considered as a key feature in qualifying a survey program, as higher precision will lead to a more reasonable understanding of the intrinsic properties of targets. In this sense, the correctness of the error estimation is another critical issue of the survey. This is because either underestimation or overestimation will significantly affect the measurement of intrinsic scatters of the physical properties of interest, especially in cases where the error is similar to or even larger than the scatter value.

There are statistical methods to assess whether the error measurements of a survey are overestimated or underestimated. They are based on the principle that the measurement error, which is the internal uncertainty, should have the same level of the deviation of the measured parameter to its true or expected value. One method utilizes the selected targets with definite

true values. For example, the QSOs are theoretically expected to have zero parallaxes and zero proper motions. So the Gaia data reduction procedure can use the comparison of the QSO's observational data deviation from zero with their error distribution to estimate their correction factors of the parallax and proper motion errors and then apply them to the entire sample of Gaia (Lindegren et al. 2018). Alternatively, in the cases where we have duplicate observations of a given target, it is also possible to use the average observational value instead of the true or theoretical value and then compare the standard deviation with their given errors. For example, using this method, Tsantaki et al. (2022) assess the radial velocity (RV) errors of multiple recent surveys and estimate the correction factors of RV errors for each catalog.

The Large sky Area Multi-Object fiber Spectroscopic Telescope (hereafter, LAMOST) is a Chinese national scientific research facility operated by the National Astronomical Observatories, Chinese Academy of Sciences. It is a special quasi-meridian reflective Schmidt telescope (Wang et al. 1996; Su & Cui 2004; Zhao et al. 2006; Cui et al. 2012; Luo et al. 2012; Zhao et al. 2012) with both a large aperture of 4 m and a large field of view (FOV) of 5°, which enables it to observe up to 4000 targets per exposure simultaneously. Up to now, it has observed more than 10 million spectra.

The large volume of spectroscopic observations is posing great challenges for data analysis. Besides the official data releases of LAMOST (Luo et al. 2015, 2022), there are many works of deriving the stellar labels of LAMOST spectra based on different strategies. For example, Xiang et al. (2015) have established a stellar parameter pipeline at Peking University (LSP3) to determine radial velocity and stellar atmospheric parameters for the LAMOST Spectroscopic Survey of the Galactic Anti-center (LSS-GAC); Ho et al. (2017) have discussed the difference of precision between their data-driven approach (Cannon) and the LAMOST official pipeline for red giant stars; Xiang et al. (2017) employed the Kernel-based principal component analysis (KPCA) in dealing with the LAMOST spectra, and also used it for Red-Clump stars. More recently, the Stellar LAbel Machine (SLAM) method (Zhang et al. 2020, hereafter ZL20), and the Neural Network method (Wang et al. 2022, hereafter WH22) have been introduced in deriving stellar parameters from the LAMOST spectra.

Nevertheless, the public data release from the official LAMOST team is still the most widely used data product, followed by abundant scientific research works. According to the illustration of the LAMOST stellar parameter pipeline (LASP), the stellar parameters are derived based on the χ^2 fitting technique, and their "nominal" errors are estimated through an empirical approach, which is quite complex and indirect (see Section 4.4.5 of Luo et al. 2015 for details). Since the precision of stellar parameters measured from the spectra could be affected by many aspects, such as the spectral range, spectral resolution, wavelength calibration, stellar spectral type, and the measurement methods (Bouchy et al. 2001; Wang & Luo 2012; Wang et al. 2014, 2019), it probably has more or less misestimation of the parameter errors. So it is necessary to make a rigorous statistical assessment of the LAMOST parameter errors in order to carry out further in-depth research in investigating intrinsic stellar properties.

Fortunately, there are quite a large amount of duplicate observed targets in the LAMOST survey. Most of them are aimed at time-domain research programs, while some are due to the fiber-pointing restriction in the low star-number-density region of the multiply covered survey fields (Zhang et al. 2013; Liu et al. 2014; Zhang et al. 2014; Yuan et al. 2015). For example, in the Galactic Anti-center (LSS-GAC) survey, ∼23% of observed stars are actually targeted more than once (Liu et al. 2014). Therefore, these duplicate observational spectra construct a natural sub-sample to investigate the appropriation of the errors of parameters measured from LAMOST spectra. In this paper, we focus on the low-resolution spectrograph (LRS) catalog of the latest public data release (DR7) and use the duplicate sample to assess the parameter errors, by estimating their correction factors as functions of stellar type, signal-to-noise ratio (S/N) and the error itself. Then we will suggest the correction of errors for the entire LAMOST-LRS sample.

This paper is organized as follows. Section 2 describes the method to estimate the correction factors of parameter errors using the duplicate observational sources. In Section 3, we describe the LAMOST sample that will be assessed. The error correction factors and their dependence are discussed in Section 4. In Section 5, we calculate the error correction factors for the duplicate observed sample, quantify the correlations between the corrected errors and S/N, and then estimate the correction factors of each stellar parameter of each spectrum for the entire LAMOST-LRS DR7. Finally, a brief summary is presented in Section 6.

2. Method

For a specific stellar parameter (x), e.g., the radial velocity or the metallicity of a star with repeated spectral observations ($n_{\rm dup}$). Assume that each measurement (x_i, $i = 1,\ldots,n_{\rm dup}$) is randomly centered on the true value, it will be expected to follow a Gaussian distribution with standard deviation characterized by its observational error. Therefore, we define the normalized difference (U) of the ith measurement as:

$$U_i(x) = \sqrt{\frac{n_{\rm dup}}{n_{\rm dup}-1}}\frac{x_i-\bar{x}}{e_i}, \quad (1)$$

where e_i is the error of the ith measurement and $\bar{x}$ is the mean value of $n_{\rm dup}$ measurements of this star.

In the ideal case, U_i should follow a Gaussian distribution with zero mean and unit standard deviation, $\mathcal{N}(0, 1)$. We have to emphasize that this $\mathcal{N}(0, 1)$ assumption is the most

fundamental statistical principle that should be suitable for any sub-samples of the data set, whether for a specific stellar type or a sub-sample with low or high S/N, or even for a randomly select group of targets. Moreover, this feature should be appropriate for the whole data set of the survey. That means, if we totally have $N_{\rm spec}$ spectra of a set of $N_{\rm star}$ duplicated observed stars. Then, when we calculate the U values of $n_{\rm dup}$ measurements for each given star, the U distribution of totally $N_{\rm spec}$ values is also expected to follow $\mathcal{N}(0, 1)$. Usually, for parameters of a real survey, the U distributions may differ from the Gaussian shape and/or have the standard deviation unequal to one. That is why the error correction factors are often required.

In this paper, we define a dispersion parameter k to be the half of 16%–84% width of the U distribution. If $k > 1$, that means the difference of a measurement to its average value is generally larger than what the error expressed. That means, the error of this parameter is underestimated, and vice-versa. Therefore, k could be regarded as a correction factor of the error, with the corrected error to be ke_i. There is an advantage of using this definition rather than using the standard deviation (σ_U). Because of the data set of a real survey, it surely includes some (usually less than one percent) variable stars, which may extend the tails of U distribution. So in this case, the width of percentage range is much more robust than the standard deviation in characterizing the dispersion.

In the following sections, we will diagnose the k values for specific sub-samples of LAMOST stellar parameters to investigate the correctness of the errors and subsequently correct the errors for the entire LAMOST-LRS sample.

3. Sample of the LAMOST Low-resolution Spectra Survey

The seventh data release of LAMOST (DR7 v2, Luo et al. 2022) contains 10, 431, 197 low-resolution spectra, which can be available from the website.[5] These spectra have a resolution of $R \sim 1800$ at 5500 Å and a wavelength coverage of 3700 Å $\leqslant \lambda \leqslant$ 9000 Å.

The LAMOST spectral analysis pipeline (also called the 1D pipeline) is used to perform spectral classification. By using a cross-correlation method, the pipeline recognizes the spectral classes, e.g., galaxies, different spectral types of stars, QSOs and other small amounts of sub-classes (Luo et al. 2015). In the meantime, it determines the initial value of redshifts or radial velocities from the best fit correlation function.

LAMOST-LRS DR7 has a stellar parameter catalog of A, F, G and K spectroscopy types (classified by the LAMOST 1D pipeline). It contains effective stellar parameter (SP) results for 6, 079, 235 spectra after excluding spectra with invalid signal-to-noise ratios (S/N) or invalid parameter errors. We denote it as the SP-sample. It provides the measurements of radial velocity (RV) and three stellar atmospheric parameters, including the effective temperature ($T_{\rm eff}$), surface gravity ($\log g$) and metallicity ([Fe/H]). For the first time, LAMOST DR7 also provides the α-enhancement measurement ([α/M]) for about 60% spectra with the g-band S/N larger than 20. We further denote this sub-sample as the α-sample. All these parameters were automatically measured by the LAMOST Stellar Parameter pipeline (LASP) (Wu et al. 2011; Luo et al. 2015).

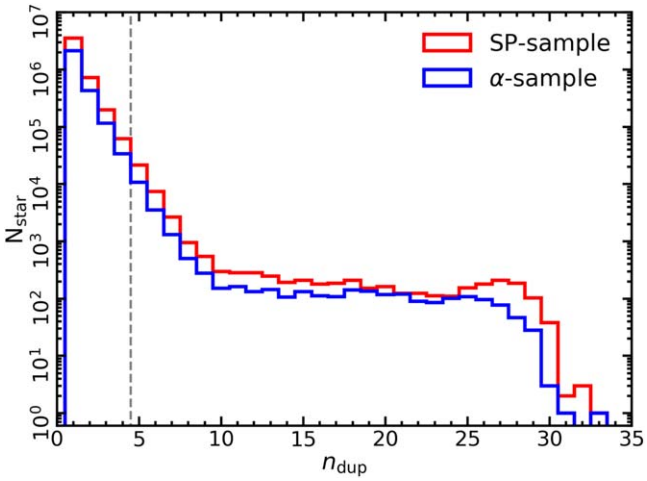

Figure 1. Distribution of the repeatedly observed spectra numbers ($n_{\rm dup}$) in LAMOST-LRS DR7. The red and blue histograms represent the SP-sample and the α-sample, respectively. Right side of the gray dashed vertical line is the defined duplicate SP-sample and α-sample of this paper.

Within the SP-sample, LAMOST DR7 also marks the duplicate spectral observations identified by a cross-match within 3″ in coordination. The histogram of duplicate observation numbers ($n_{\rm dup}$) of stars is plotted in Figure 1. In this work, we pick up a sub-sample of stars with $n_{\rm dup} \geqslant 5$ to assess the errors of LAMOST-LRS. It totally contains 36,696 stars with 251,346 spectra, called the duplicate SP-sample. The corresponding duplicate α-sample contains 18,570 stars with 130,535 spectra. The total numbers of stars, spectra, and the numbers of spectra of different stellar spectral types are listed in Table 1 for different definitions of sub-samples separately.

We can employ the duplicate sample to assess the measurement errors of stellar parameters derived from the LAMOST-LRS. For each spectrum in the duplicate sample, we calculate its U value based on Equation (1) for a given parameter. Then the U values will be used to determine the error correction factor k for this parameter and also discuss the k values of specific sub-samples, e.g., for different spectral types or S/N.

Figure 2 shows the number density distributions of spectra in the error-S/N plane for each parameter. Generally, according to the distribution shapes, it can be found that there is a

[5] http://dr7.lamost.org/v2.0/catalog

Table 1
Numbers of Assorted Samples

Spectral Type	SP-sample	α-sample	SP-sample$_{dup}$	α-sample$_{dup}$
A	94,262	26,965	4758	1092
F	1,881,799	1,422,634	92,243	62,382
G	3,065,809	2,019,558	122,326	64,730
K	1,037,365	135,566	32,019	2,331
N_{spec}	6,079,235	3,604,723	251,346	130,535
N_{star}	4,546,803	2,730,053	36,696	18,570

correlation between S/N and errors, where larger S/N leads to smaller error values. We also can find that the number density shapes of three stellar atmospheric parameters ([Fe/H], $T_{\rm eff}$, $\log g$) are similar. Possibly, it is because these three parameters were simultaneously estimated by matching the template based on the ELODIE library (Prugniel & Soubiran 2001, 2004; Prugniel et al. 2007), while the α-enhancement ([α/M]) was estimated based on the MARCS synthetic spectra (Gustafsson et al. 2008).

4. Variations of Correction Factors

4.1. The Distribution of Normalized Difference Value U

Using Equation (1), we calculate the normalized difference value U of each error of all five stellar parameters from each spectrum. The distributions of U of these parameters are shown as histograms in the left panels of Figure 3 for the duplicate samples. As expected, all five U distributions are centered on the zero-point. But they all significantly differ from the Gaussian shape, where they all have larger peak values and more extended wings on both sides. That means the error estimation of the LAMOST-LRS is not the perfect one and may have some complicated factors or conditions during the data reduction process. According to the overall dispersion values of U distributions, one may find that the radial velocity is slightly overestimated with $k_{\rm RV} = 0.83$, and the α-enhancement is significantly underestimated with $k_{[\alpha/{\rm M}]} = 1.79$.

On the other hand, considering the sub-samples that are constrained within a small area in the error-S/N plane, one may find that they are more likely to have a Gaussian shape. For example, we plot the U distributions of two sub-samples of F-type spectra in the right panels of Figure 3. These sub-samples are taken from the local areas of two given points (the black symbols in Figure 3) in the error-S/N planes, respectively. Both of them contain the nearest 500 spectra for [α/M] and 600 spectra for the other parameters. Clearly, the U shapes of sub-sample follow the Gaussian profile quite well but with significantly different dispersions. Therefore, we may understand that the non-Gaussian distribution of the overall sample is probably caused by the summation of multiple sub-samples with different dispersions.

4.2. Correction Factor k and its Dependence

Since the dispersion of U is presumed to be the correction factor (k) of error, the non-Gaussian shapes of U imply the complication of the k-factors. To make the diagnosis, we split the duplicate spectral sample by different conditions, i.e., different spectral types, S/N, and measurement error themselves. Figure 4 shows the k values as functions of these conditional features for all five parameters. Obviously, most of the k values differ from 1 and also have significant variations across these features.

Variation with Spectral Type: The variations of k-factor with stellar spectral type (including A, F, G, and K types) of the duplicate SP-sample and α-sample are shown in the left panel of Figure 4. We can see that $k_{[\alpha/{\rm M}]}$ has the most significant decreasing trend with the spectral type, from A-type spectra having a huge underestimation of error to K-type spectra having a slight overestimation. The values of $k_{\rm RV}$, $k_{T_{\rm eff}}$ and $k_{[{\rm Fe/H}]}$ also significantly decrease from A-type to K-type spectra. The stellar type dependence on $k_{\log g}$ is not obvious.

Variation with S/N: The variation of k-factor with the g-band signal-to-noise ratio (S/N$_g$) is shown in the middle panel of Figure 4. Clearly, the values of $k_{[\alpha/{\rm M}]}$ are all larger than 1. They have the most significant trend of decreasing k values with an increase in S/N. For the other three atmospheric parameters ([Fe/H], $T_{\rm eff}$, and $\log g$), the k values vary around 1 but without a monotonous trend like that of the [α/M]. The common feature is that they all have the smallest k values at S/N$_g \sim 20$ and $\sim$120. All of the $k_{\rm RV}$ values are less than 1 and have a similar variation trend to those of the atmospheric parameters. In brief, it can be concluded that, except for the [α/M], all S/N dependence on the k-factors are weak.

Variation with observational error: More importantly, we find there are significant correlations between k-factors and the error themselves for all five stellar parameters, whether they are generally overestimated or underestimated. As shown in the right panel of Figure 4, smaller errors have larger k values. That means the LAMOST-LRS measurements have systematic trends of the error estimations, with smaller errors being relatively underestimated and larger errors being relatively overestimated. All five parameters present monotonous decreasing trends but with different amplitude. The [α/M] and RV present substantial decreases while the other three parameters are relatively weaker.

In summary, errors of [α/M] are greatly underestimated, and $k_{[\alpha/{\rm M}]}$ is strongly correlated with the spectral types, S/N, and the errors themselves. The errors of RV are slightly overestimated, and $k_{\rm RV}$ has significant correlations with the spectral types and also with the errors. The other three atmospheric parameters, [Fe/H], $T_{\rm eff}$, and $\log g$, are neither strongly overestimated nor

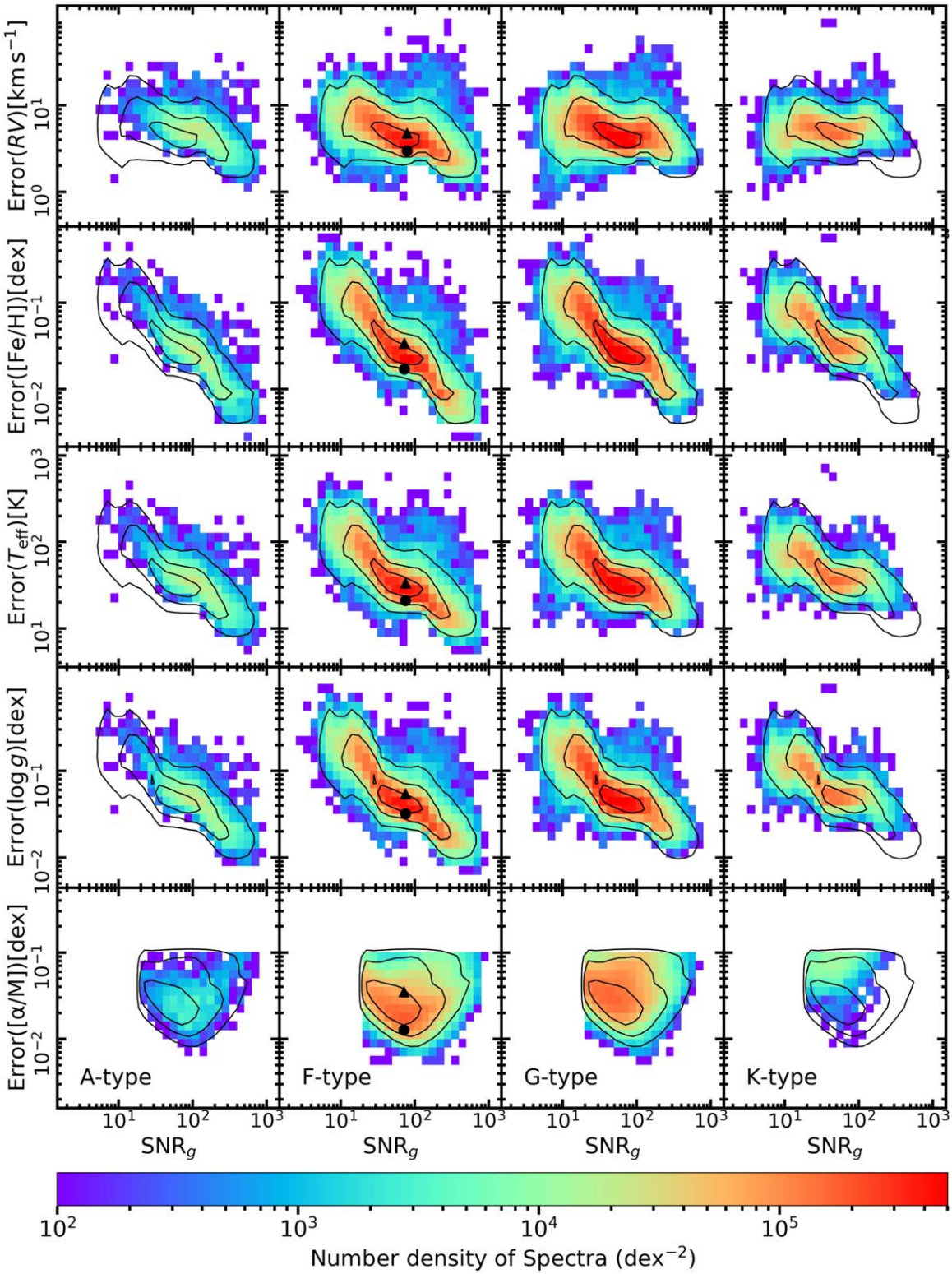

Figure 2. Number density distributions of spectra in the error-S/N planes of different spectral types of A, F, G and K for duplicate samples. The number densities are shown in a logarithmic scale by colors. The black lines are the 1σ, 2σ and 3σ contours of the number density of the entire SP-sample and α-sample. The black filled circle (sub1) and triangle (sub2) symbols point the positions of two "local" sub-samples of F-type spectra that described in Section 4.1.

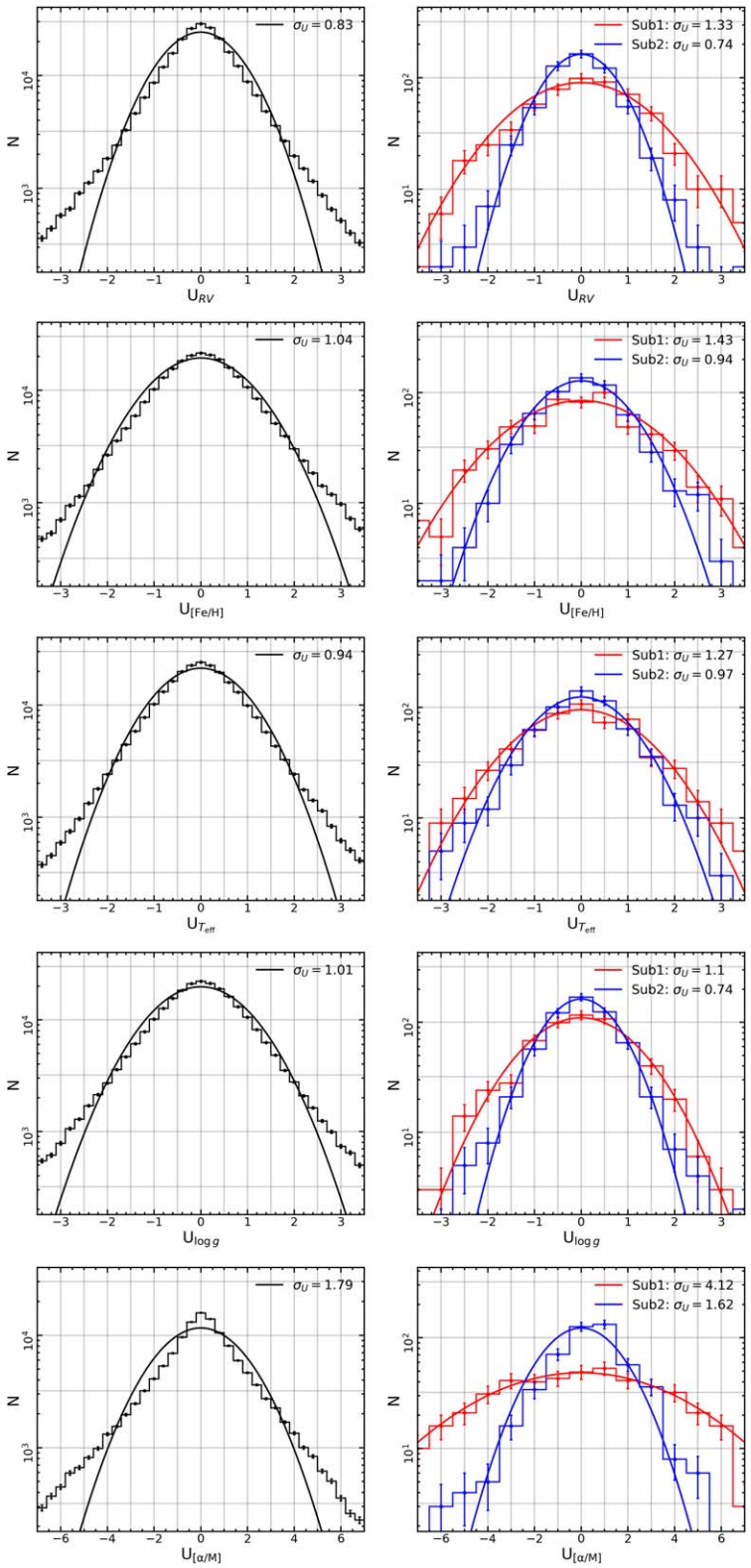

Figure 3. Distribution of U values for five stellar parameters. (Left) The entire duplicate samples. (Right) Two sub-samples of F-type spectra, with sub1 and sub2 being located at the circle and triangle symbols in Figure 2, respectively. The solid curves represent the Gaussian profiles with zero mean and standard deviation of $\sigma_U = k$. The error bars for individual bins are the Poisson fluctuation values, $N_{\rm bin}^{1/2}$.

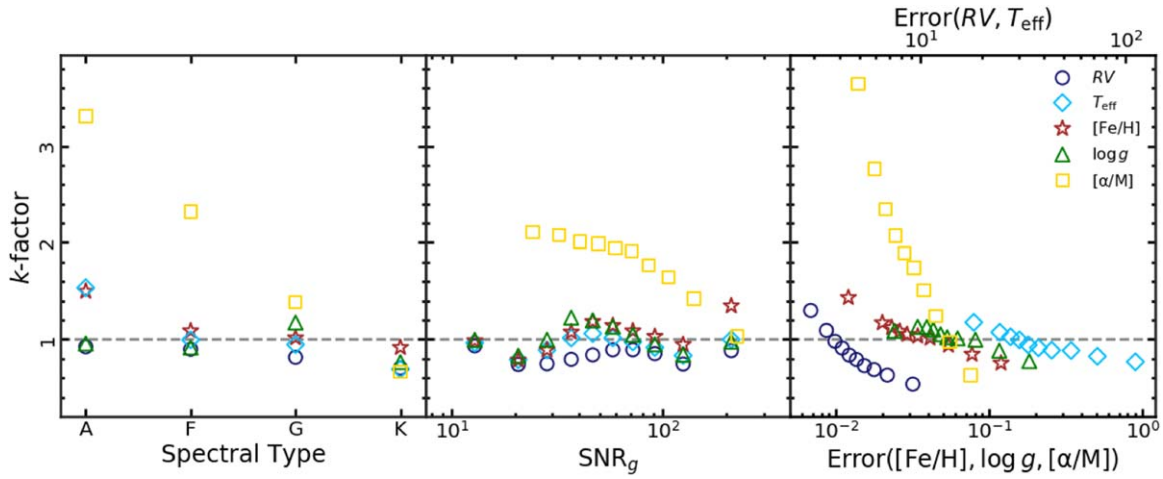

Figure 4. The error correction factors (k) as functions of the spectral type (left), S/N (middle) and the corresponding error themselves (right) for all five parameters. Different symbols represent different parameters, with blue circles for radial velocity (RV), cyan diamonds for effective temperature ($T_{\rm eff}$), red stars for metallicity ([Fe/H]), green triangles for surface gravity ($\log g$) and yellow squares for $\alpha-$enhancement ([α/M]). The gray dashed lines indicate the expected value k-factor $= 1$.

underestimated but still have significant variations of k-factors according to their spectral type, S/N, and error.

So, there are two unexpected matters that have to be corrected. First, according to the definition of U, its dispersion should be kept in constant across the whole data set, whether what kind of specific sub-samples are detected. Second, since Luo et al. (2015) claimed that the parameter errors from LASP may include the external uncertainty components, so it should be generally slightly larger than the internal (random) errors. But, as we find in the current data set, about half of the errors are underestimated, which reinforces the necessity of the error correction.

5. Correction Factors of the LAMOST-LRS

Ideally, the correction factor k of stellar parameters should be independent of any conditional features. Unfortunately, the current version of LAMOST-LRS (DR7) has non-negligible variations related to some features, i.e., the spectral type, S/N and corresponding parameter error. Therefore, the corrections of observational errors should be considered as functions of these features.

5.1. Correction of the Duplicate Sample

For the duplicate SP-sample and α-sample, we have calculated the U values of each spectrum for each parameter. As shown in the right panels of Figure 3, in a local area in the error-S/N plane, the distributions of U are more likely to have a Gaussian profile. Thus, we can suppose the dispersion of the "local" U profile to be the correction factor of the central spectrum.

First, a duplicate sample is split into four sub-samples by their spectral types, A, F, G, and K. Then, for a given spectral type and each derived parameter, we define its local area that includes the nearest $N_{\rm sub}^{1/2}$ spectra in the error-S/N plane, where $N_{\rm sub}$ is the total number of spectra of the sub-sample of a given spectral type (as listed in Table 1). Next, we take the dispersion value of U of this local area to be the k-factor of this parameter derived by this spectrum.

In Figures 5–8, we plot the k-factor results in the left panels. Each dot represents a spectrum and is colored by its k value of the specific parameter. It is clear that the variations of k are large, usually over several times, especially for the errors of [α/M] and RV. The dependence on both S/N$_g$ and error are obvious and smooth. Generally speaking, parameters with smaller S/N$_g$ and/or error have larger k values, which is consistent with the trends shown in Figure 4.

We then correct each parameter error by its corresponding k-factor for the duplicate SP-sample and α-sample. As a comparison, we repeat the above procedure for the corrected errors, i.e., calculate the U value for each parameter for each spectrum, estimate the U dispersion in the local area and plot the updated k-factors in the middle panels of Figures 5–8. We then find that the corrected errors have almost the same k values (with the same color), which are all approximate to 1, whether for different spectral types and parameters. It is the ideal phenomenon that we expect.

It should be noted that after this correction, the parameter errors of the duplicate samples are re-calibrated to the formal internal (random) uncertainties for the LASP.

5.2. Correlations between Parameter Errors and S/N of Spectrum

After the correction, there is another improvement. We can find that the number density distribution (contours in the

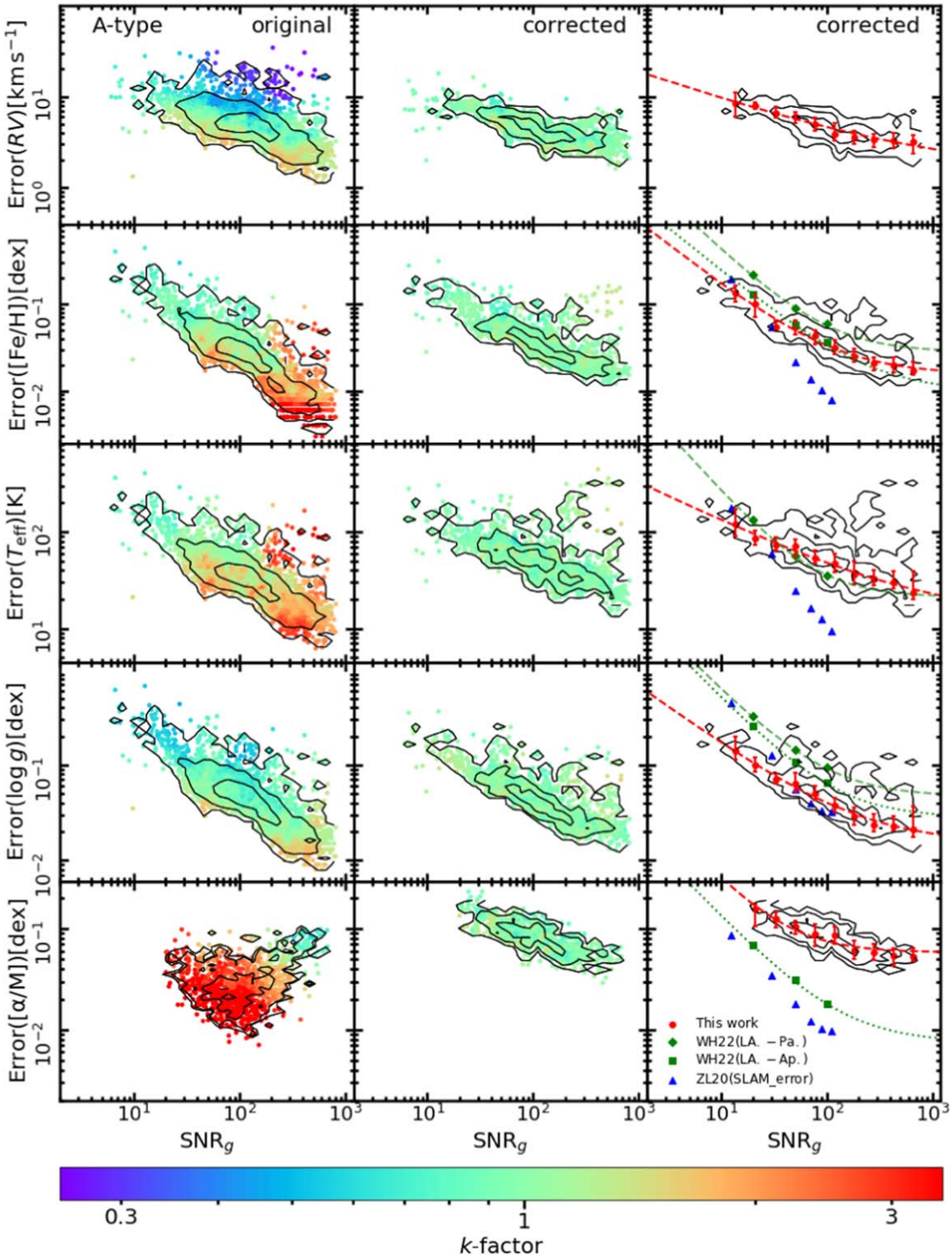

Figure 5. Distributions of correction factors (k) of original errors (left) and corrected errors (middle and right) of A-type spectra in the error-S/N plane. Each dot represents a spectrum in the duplicate SP-sample and α-sample and is colored by its k value of the specific parameter. The black lines are the 1 σ, 2σ and 3σ contours of the spectral number density. In the right panels, for each parameter, red symbols with error bars represent the median values and 16%–84% percentage ranges of the corrected errors in given S/N$_g$ bins. Red dashed lines correspond to their best fitting. Green diamonds and squares present the typical random errors of stellar parameters estimated by WH22 for their LAMOST-PASTEL (LA.-Pa.) and LAMOST-APOGEE(LA.-Ap.) training samples, with the green lines for the fitting curves. Blue triangles present the random errors (SLAM-error) shown in ZL20.

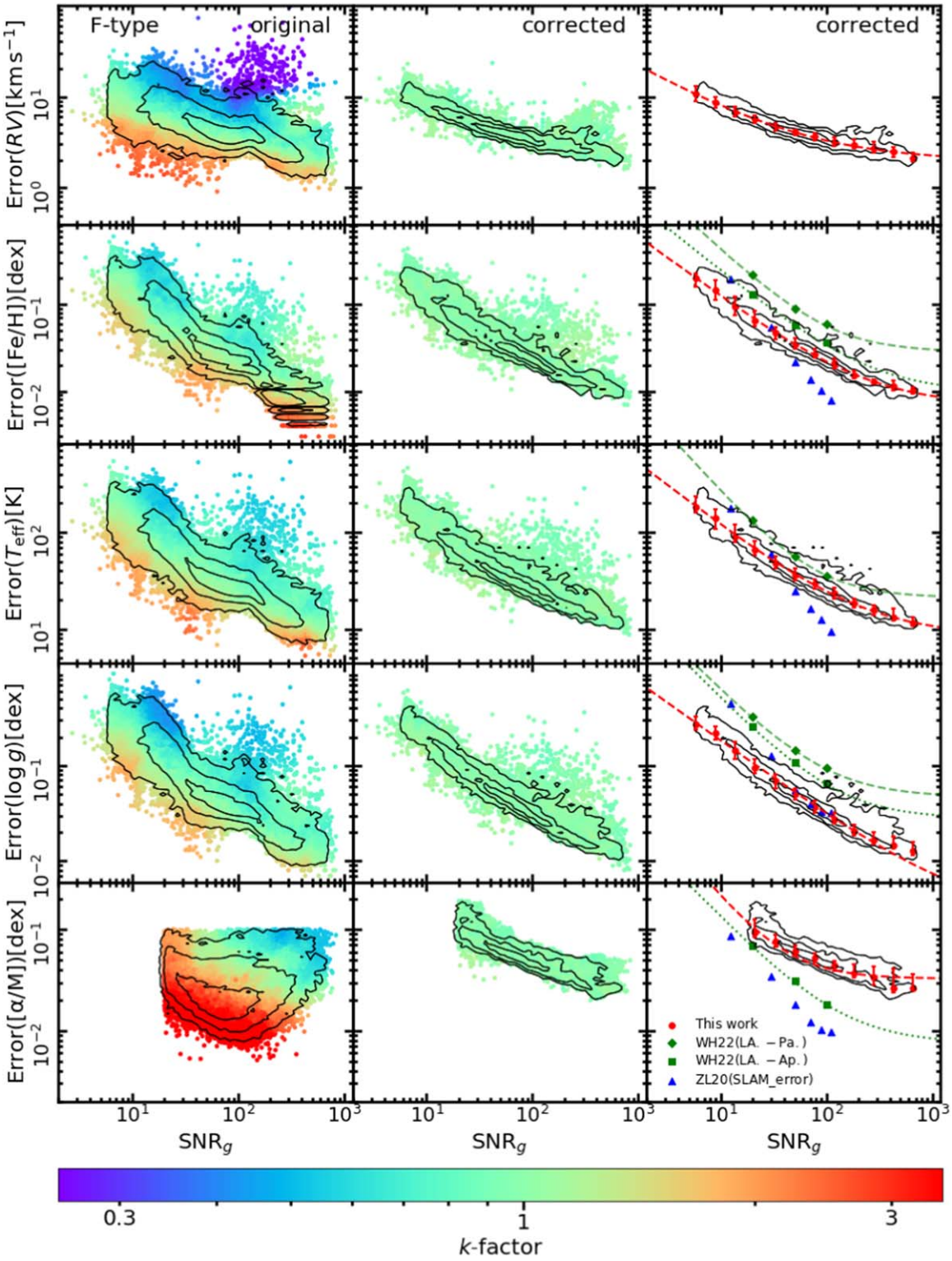

Figure 6. Same as Figure 5, but for the F-type spectra.

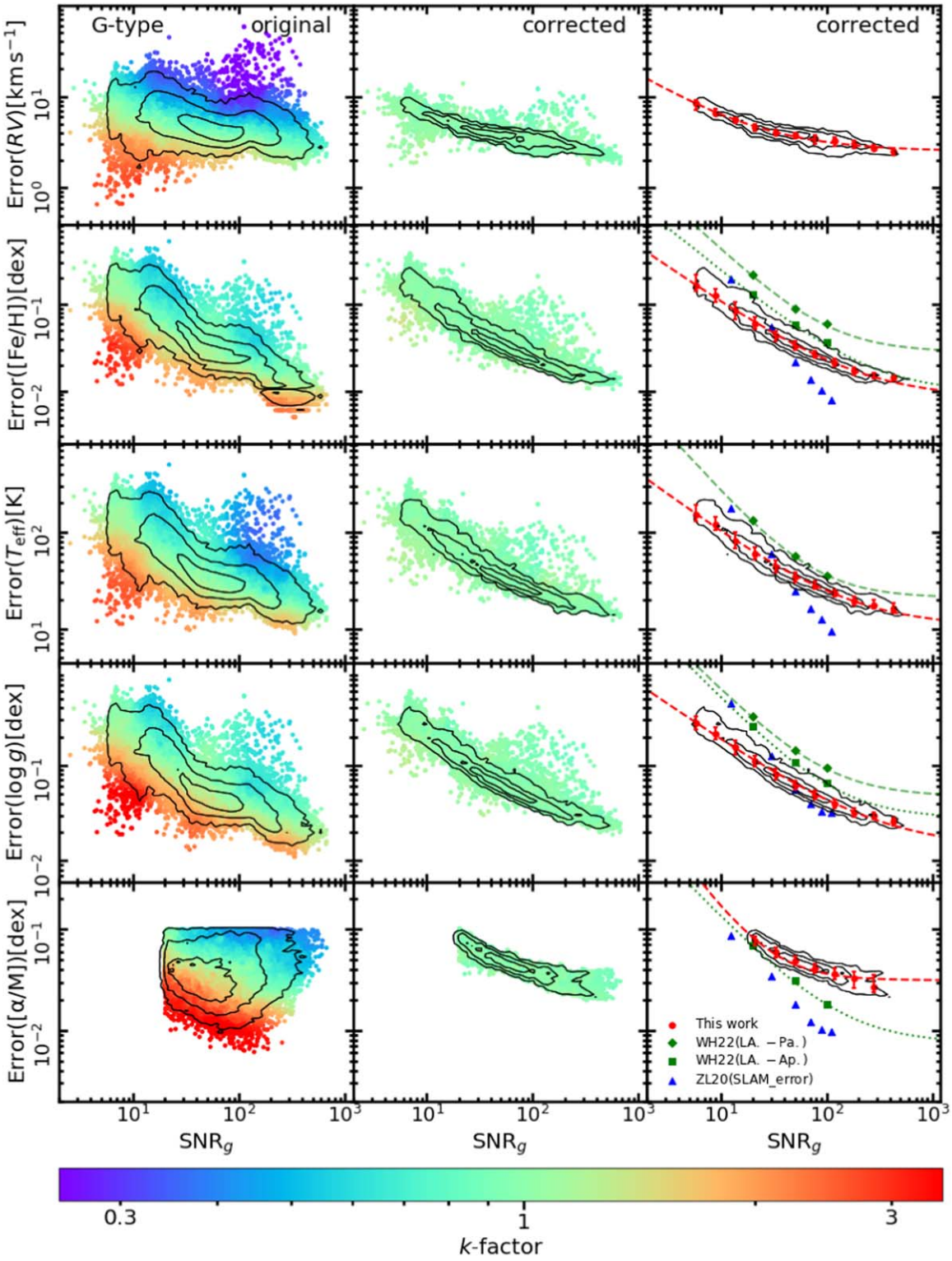

Figure 7. Same as Figure 5, but for the G-type spectra.

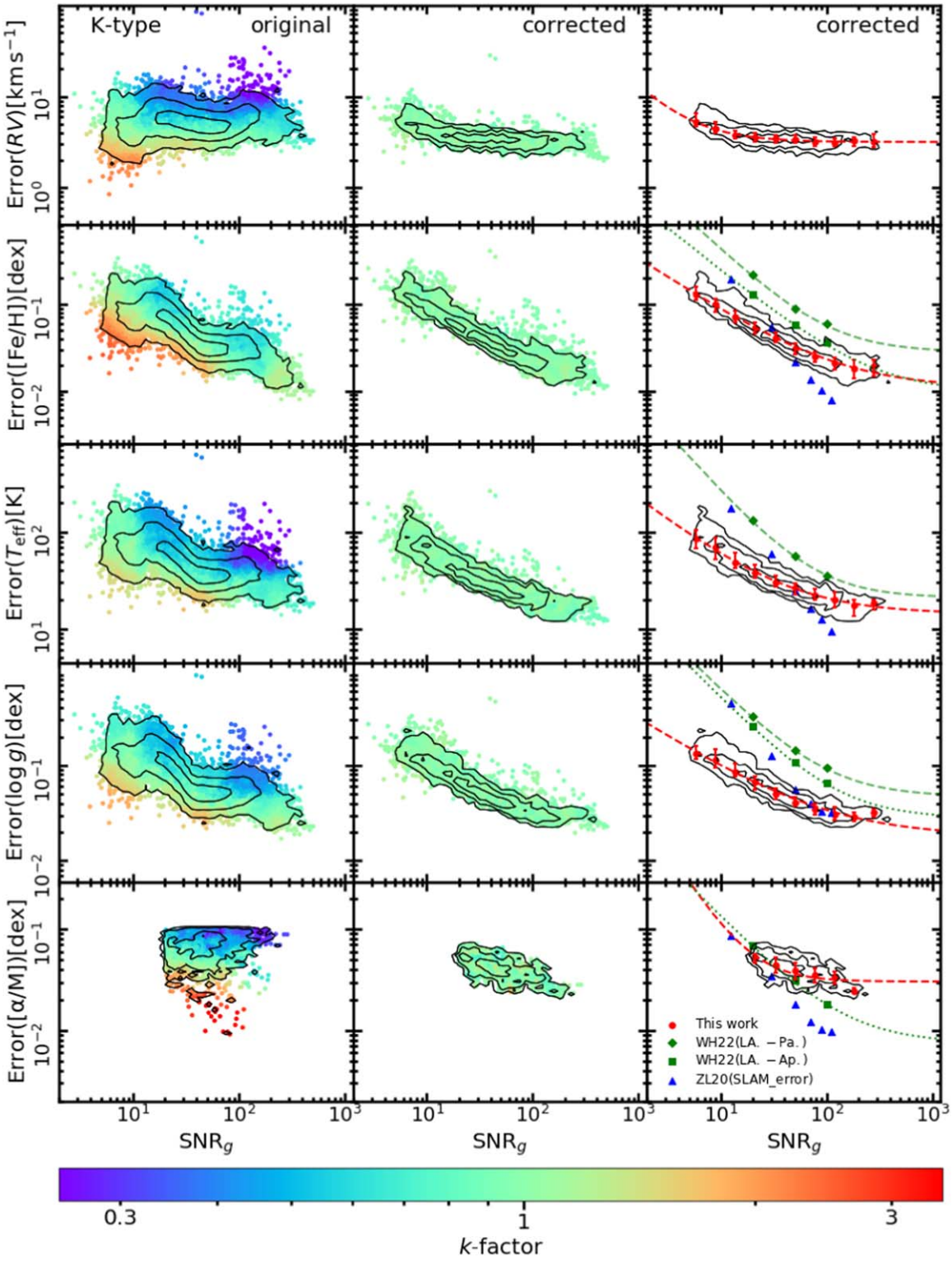

Figure 8. Same as Figure 5, but for the K-type spectra.

Table 2
The Fitting Coefficients of the Parameter Error as Functions of S/N_g

Spectral Type	e_{RV}(km s^{-1})			$e_{[Fe/H]}$(dex)			$e_{T_{eff}}$(K)			$e_{\log g}$(dex)			$e_{[\alpha/M]}$(dex)		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
A	1.5	0.415	21.5	0.015	0.950	1.416	12.4	0.533	414.4	0.015	0.798	0.997	0.058	1.178	4.001
F	2.0	0.664	28.0	0.007	0.901	0.947	8.5	0.849	800.4	0.003	0.796	1.134	0.033	1.329	3.988
G	2.5	0.765	22.6	0.008	0.839	0.692	10.4	0.810	604.9	0.014	0.768	1.037	0.031	1.450	3.978
K	3.2	1.274	19.4	0.011	0.798	0.489	14.3	0.851	328.7	0.018	0.712	0.427	0.030	1.669	3.958

middle and right panels of Figures 5–8) have more regular patterns than the uncorrected ones (in the left panels) and also present tight error-S/N relationships.

Red symbols with error bars in the right panels represent the median values and the 16%–84% percentage ranges of the corrected errors of given S/N_g bins. We follow WH22 to assume that the parameter error (e) as an empirical function of S/N:

$$e = a + \frac{c}{(S/N_g)^b} \quad (2)$$

The best fitting coefficients (a, b, c) of each stellar parameter for each spectral type are summarized in Table 2. As shown as the red dashed lines in the right panels of Figures 5–8, this function is definitely satisfied for the error-S/N correlations for all stellar parameters. We also plot the internal (random) errors from ZL20 and WH22 for comparison. We can find that the trends of the error dependence on S/N are similar, but with significant offsets between different stellar measurement approaches. The errors from SLAM (ZL20) are generally smaller and those from the Neural Network method (WH22) are larger. It reveals the fact that even based on the same (or similar) observational data, the different data reduction procedures may lead to different internal uncertainty levels for the derived parameters.

5.3. Correction of the Entire LAMOST-LRS DR7

For the entire LAMOST-LRS DR7 sample, most stars have only one spectrum, so we cannot calculate their U values to derive the dispersion. However, we can reasonably assume that the entire sample's error corrections follow the same functions of conditional feature as the duplicate samples. So there are two methods to correct, or re-estimate their errors. One is using the empirical formula of Equation (2) and the coefficients in Table 2 to directly calculate the corresponding errors for each spectrum according to its spectral type and S/N_g. In this work, we employ an alternative method. For each spectrum of each parameter, according to its S/N_g and original error, we can also define its local area in the error-S/N plane, which contains the nearest $N_{sub}^{1/2}$ spectra from the duplicate sample. Then, the k value from the local spectra of the duplicate sample is regarded as the k-factor of this spectrum from the entire sample. Using this method, we have derived the correction factors for all five parameters of each spectrum of the LAMOST-LRS DR7. The k values are listed in Table 3. In Table 4, we also list the typical errors of each parameter for different spectral types at $S/N_g = (20, 50, 100, 200)$ separately. It indicates that the typical errors are not only related to the spectral S/N, but also vary with different spectral types. Briefly, the later spectral type has the more precise measurement from LASP.

Figure 9 shows the distributions of k of five parameters of the entire LAMOSR-LRS DR7 sample. We can find that the majority of [α/M] errors are underestimated, with the median value of correction factor $k_{[\alpha/M]} = 1.81$. Most of the RV errors are overestimated, and the median value $k_{RV} = 0.85$. For the other three stellar atmospheric parameters, the k values are all approximately centered on 1 but still have very large scatters $\sim$0.3 dex. That means, for individual spectrum, the distinguishable corrections of the original errors are necessary.

There are two points that should be mentioned in using of Table 3. One is that the k-factors listed in the table are only for the LAMOST-DR7 (v2, Luo et al. 2022). Second, we have to emphasize that after the correction, the errors are updated to the formal internal (random) uncertainties of LASP. These corrected errors can be used in the cases of investigating the intrinsic stellar properties if only the LAMOST(LASP) parameters are employed. If one carries out an investigation that combines data sets from multiple surveys, or the same survey but with different data reduction approaches, the systematic (external) uncertainties among them should also be considered.

5.4. Previous Data Releases of LAMOST-LRS

In practice, we have used the same method to diagnose the parameter errors of earlier data releases of LAMOST-LRS, i.e., DR5 and DR6. It is found that the variations of k values are also significant. The dependence on the spectral type, S/N, and errors are all similar to those of the current DR7. More severely, the earlier data releases seem to have largely overestimated the parameters errors. The overestimation has also found by other works. For instance, the RV errors of DR5

Table 3
An Example of Error Correction Factors of Observational Parameters

Obsid	Starid	$n_{\rm dup}$	Spectral Type	S/N_g	$k_{\rm RV}$	$k_{\rm [Fe/H]}$	$k_{T_{\rm eff}}$	$k_{\log g}$	$k_{\rm [\alpha/M]}$
91206103	215619	5	K1	22.27	0.47	0.67	0.51	0.53	0.65
283206103	215619	5	G9	12.59	0.60	0.70	0.72	0.81	
401216192	215619	5	G5	6.07	1.89	1.83	1.93	2.05	
555506103	215619	5	G8	16.26	0.74	0.73	0.79	0.86	
619416192	215619	5	K0	136.11	0.66	0.65	0.64	0.66	
427003200	929881	3	G7	68.70	1.04	1.17	1.06	1.30	0.93
419109163	929881	3	G8	24.70	0.50	0.71	0.69	0.80	1.35
433003200	929881	3	G7	30.50	0.72	0.88	0.94	1.18	0.70
290013186	271836	4	F5	36.78	0.70	0.96	0.95	1.0	4.22
337113186	271836	4	F5	71.17	1.07	1.20	1.14	1.04	5.57
103514249	271836	4	F5	28.59	1.08	1.02	1.13	1.07	3.44
103614249	271836	4	F4	14.31	0.84	0.75	0.77	0.83	
92915065	1343734	1	F7	69.98	1.20	1.27	1.11	1.11	2.92
217508200	2069432	1	K1	69.56	1.07	1.17	1.13	1.32	0.48
420511220	3324356	1	G5	22.38	0.59	0.68	0.72	0.82	
734108151	4504813	1	F6	20.72	0.84	0.78	0.81	0.74	2.85
…	…	…	…	…	…	…	…	…	…

Note. Column 1 is the LAMOST IDs of the objects, columns 2–3 are the ID and duplicate observation numbers of stars in the SP-sample, columns 4–5 are the spectroscopy types classified by the LAMOST 1D pipeline and the LAMOST spectral S/N of g-band, columns 6–10 are the correction factors (k) of each parameter. This table is available in its entirety in CSV format and will be updated for the later LAMOST data releases (http://nadc.china-vo.org/res/r101178/.)

have been evaluated by Tsantaki et al. (2022), and they concluded that the correction factor is ∼0.4.

Actually, the error values are updated with different data releases of LAMOST. We have compared the errors of RV, [Fe/H], $T_{\rm eff}$ and $\log g$ of the same spectra of DR6 and DR7, and found that from DR6 to DR7, there already has been an overall (or systematic) correction by factors of ∼0.83, ∼ 0.45, ∼ 0.42 ∼ 0.42 for RV, [Fe/H], $T_{\rm eff}$ and $\log g$ respectively. So, in fairness, DR7 of LAMOST-LRS has so far the most improved estimations of the errors for all stellar parameters except [α/M], with average (or typical) k values not far away from 1. However, the problem of the spectral type, S/N and error dependence on k-factors still remain, and the wide distributions of the k values (see Figure 9) suggest that the parameter errors should be corrected individually.

6. Summary

This work aims to investigate whether the measurement errors of stellar parameters from the official data release of LAMOST-LRS are overestimated or underestimated, and dedicate to correct them to the proper internal uncertainties, which obey the hypothesis that the parameter's deviation and uncertainty are keeping in the equal level.

We define the dispersion of the normalized difference U of repeated measurements to represent the correction factor of parameter errors. For five derived parameters of LAMOST-LRS, RV, [Fe/H], $T_{\rm eff}$, $\log g$, and [α/M], we find that the k values have significant variations with the spectral type, S/N, and the error themselves. That means the correction factor of errors should be a function of these three conditional features. Generally, earlier type spectra have relative underestimations of the errors. Another very significant trend is found for the parameter error themselves. That is, smaller errors have larger underestimations, and larger errors have more overestimations.

Using the duplicate spectral samples, we calculate correction factors as functions of spectral type, S/N_g and error. After the correction, we quantify the tight correlations of corrected errors with S/N_g for all five parameters. These correlations are also spectral types dependent.

We further calculate correction factors of all five observational parameters for the entire LAMOST-LRS DR7 catalog. The majority of the [α/M] errors are largely underestimated, and most of the RV errors are overestimated. All five parameters have significantly wide dispersions that due to the dependence on the spectral type, S/N, and error values, suggests that the errors should be corrected individually.

Additionally, we have to mention that, in this work, we have only analyzed and corrected the mono-parameter error. For the three simultaneously derived atmosphere parameters, the similar structure of their distributions in the error-S/N planes, either for the original or the corrected errors, implies that the measurements of these parameters are associated, so the covariance among them should also be considered if they were provided. That means there is still plenty of room for improvement of the LAMOST data reduction pipeline.

Table 4
The Typical Parameter Errors at Different S/N_g

Spectral Type	e_{RV}(km s^{-1})				$e_{[Fe/H]}$(dex)				$e_{T_{eff}}$(K)				$e_{\log g}$(dex)				$e_{[\alpha/M]}$(dex)			
A	7.4	6.2	4.9	3.6	0.099	0.051	0.033	0.025	92	65	47	32	0.10	0.06	0.04	0.03	0.140	0.109	0.087	0.065
F	5.9	4.1	3.4	2.9	0.068	0.033	0.023	0.015	69	36	25	17	0.10	0.05	0.03	0.02	0.096	0.062	0.047	0.036
G	4.7	3.8	3.4	3.0	0.061	0.033	0.023	0.016	61	34	25	19	0.11	0.06	0.04	0.03	0.078	0.048	0.038	0.032
K	3.6	3.5	3.1	3.3	0.056	0.031	0.022	0.019	39	26	21	18	0.07	0.04	0.03	0.03	0.052	0.038	0.034	0.026

Note. The typical errors of each parameter for different spectral types at $S/N_g = 20$, 50, 100, 200 are respectively shown from left to right.

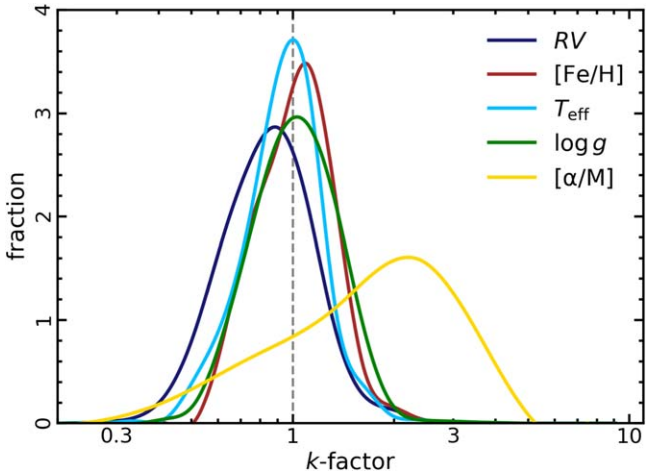

Figure 9. Normalized probability density distributions of correction factor k for five parameters of the entire SP-sample or α-sample.

Acknowledgments

We sincerely thank the anonymous referee for valuable comments and constructive suggestions. We thank Jing Zhong, Ruixiang Chang, Rui Wang, and Yunliang Zheng for helpful discussions. This work is supported by the National Natural Science Foundation of China (NSFC) under grants U2031139 and 12273091, the National Key R&D Program of China No. 2019YFA0405501, and the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A08. Lu Li thanks the support of the UCAS Joint PHD Training Program. Guo Shou Jing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

ORCID iDs

Shuhui Zhang https://orcid.org/0000-0002-1313-5647
Guozhen Hu https://orcid.org/0000-0003-1828-5318
Rongrong Liu https://orcid.org/0000-0001-5968-1144
Lu Li https://orcid.org/0000-0002-0880-3380
Zhengyi Shao https://orcid.org/0000-0001-8611-2465

References

Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2017, ApJ, 836, 5
Jofré, P., Heiter, U., & Soubiran, C. 2019, ARA&A, 57, 571
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al. 2014, in Setting the Scene for Gaia and LAMOST, IAU Symp., Vol. 298, ed. S. Feltzing et al., 310
Luo, A.-L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, RAA, 12, 1243
Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095
Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2022, V/156 VizieR Online Data Catalog, V/156
Prugniel, P., & Soubiran, C. 2001, A&A, 369, 1048
Prugniel, P., & Soubiran, C. 2004, arXiv:astro-ph/0409214
Prugniel, P., Soubiran, C., Koleva, M., et al. 2007, arXiv:astro-ph/0703658
Su, D.-Q., & Cui, X.-Q. 2004, ChJAA, 4, 1
Ting, Y.-S., Conroy, C., Rix, H.-W., et al. 2017, ApJ, 843, 32
Tsantaki, M., Pancino, E., Marrese, P., et al. 2022, A&A, 659, A95
Wang, C., Huang, Y., Yuan, H., et al. 2022, ApJS, 259, 51
Wang, F., Luo, A., & Zhang, H. 2014, in Setting the Scene for Gaia and LAMOST, IAU Symp., Vol. 298, ed. S. Feltzing et al., 444
Wang, F. F., & Luo, A. L. 2012, in International Workshop on Stellar Libraries, Astronomical Society of India Conf., Vol. 6 (University of Delhi, India) ed. Philippe Prugniel & Harinder P. Singh, 253
Wang, R., Luo, A.-L., Chen, J.-J., et al. 2019, ApJS, 244, 27
Wang, S.-G., Su, D.-Q., Chu, Y.-Q., et al. 1996, ApOpt, 35, 5155
Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924
Xiang, M.-S., Liu, X.-W., Shi, J.-R., et al. 2017, MNRAS, 464, 3657
Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, MNRAS, 448, 822
Yuan, H.-B., Liu, X.-W., Huo, Z.-Y., et al. 2015, MNRAS, 448, 855
Zhang, B., Liu, C., & Deng, L.-C. 2020, ApJS, 246, 9
Zhang, H.-H., Liu, X.-W., Yuan, H.-B., et al. 2013, RAA, 13, 490
Zhang, H.-H., Liu, X.-W., Yuan, H.-B., et al. 2014, RAA, 14, 456
Zhao, G., Chen, Y.-Q., Shi, J.-R., et al. 2006, ChJAA, 6, 265
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 723