Correction to the Photometric Colors of the Gaia Data Release 2 with the Stellar Color Regression Method

Zexi Niu1,2, Haibo Yuan3, and Jifeng Liu1

1 National Astronomical Observatories, Chinese Academy of Sciences, No.20A Datun Road, Chaoyang District, Beijing, People’s Republic of China
2 University of Chinese Academy of Sciences, No.19(A) Yuquan Road, Shijingshan District, Beijing 100049, People’s Republic of China
3 Department of astronomy, Beijing Normal University, 19th Xinjiekouwai Street, Haidian District, Beijing, People’s Republic of China

Received 2020 November 5; revised 2021 January 5; accepted 2021 January 11; published 2021 March 4

Abstract

The second Gaia data release (DR2) delivers accurate and homogeneous photometry data of the whole sky of an exquisite quality, reaching down to the unprecedented millimagnitude (mmag) level for the \( G \), \( G_{\text{RP}} \), and \( G_{\text{BP}} \) passbands. However, the presence of magnitude-dependent systematic effects at the 10 mmag level limits its power in scientific exploitation. In this work, using about a half million stars in common with the LAMOST DR5, we apply the spectroscopy-based stellar color regression method to calibrate the Gaia \( G - G_{\text{RP}} \) and \( G_{\text{BP}} - G_{\text{RP}} \) colors. With an unprecedented precision of about 1 mmag, systematic trends with \( G \) magnitude are revealed for both colors in great detail, reflecting changes in instrument configurations. Color-dependent trends are found for the \( G_{\text{BP}} - G_{\text{RP}} \) color and for stars brighter than \( G \sim 11.5 \) mag. The maximum correction term of the calibration is about 20 mmag in general and varies by a few mmag/mag. A revised color–color diagram of Gaia DR2 is given, and some applications are briefly discussed.

Unified Astronomy Thesaurus concepts: Stellar photometry (1620); Astronomy data analysis (1858); Fundamental parameters of stars (555)

Supporting material: machine-readable table

1. Introduction

The second data release of the Gaia mission provides photometry data in the \( G \) band for approximately 1.7 billion sources and in the integrated \( G_{\text{BP}} \) and \( G_{\text{RP}} \) bands for approximately 1.4 billion sources calibrated to a consistent and homogeneous photometric system (Gaia Collaboration et al. 2016, 2018). The \( G \) magnitude is measured in the astrometric field and extracted by a point-spread-function fitting with a typical formal uncertainty of under 1 mmag for stars with \( G < 16 \) mag. The \( G_{\text{BP}} \) and \( G_{\text{RP}} \) magnitudes are obtained by the BP and RP photometer with larger formal uncertainties, about a few mmag for stars with \( G < 15 \) mag (Evans et al. 2018; Riello et al. 2018). With such an enormous data volume and exquisite data quality, Gaia presents a significant advance in all photometric investigations.

To make full use of the mmag-precision photometry yielded by the Gaia DR2, photometric calibration to mmag precision is required. However, systematic effects at the 10 mmag level or higher are shown by a number of photometric validation tests. Arenou et al. (2018) first showed a global increase (~2 mmag/mag) of the \( G - G_{\text{BP}} \) residuals with increasing \( G \) magnitude by subtracting fitted \( G - G_{\text{BP}} \) as a third-order polynomial function of \( G_{\text{BP}} - G_{\text{RP}} \). They also made comparisons with the external catalogs and found a similar global variation of up to 10 mmag/mag. Other studies, using well-calibrated, high-quality spectral libraries, such as Casagrande & VandenBerg (2018, hereafter CV18), Weiler (2018, hereafter WEI18), and Maíz Apellániz & Weiler (2018, hereafter MAW18) conducted synthetic photometry and compared it with the Gaia DR2 photometry. All of them detected a similar tendency in the \( G \) magnitude and proposed a linear correction for 3.4, 3.5, and 3.2 mmag/mag separately. Moreover, WEI18 and MAW18 found a significant color-dependent offset of 20 mmag between faint and bright stars in \( G_{\text{BP}} \). Nevertheless, since only hundreds of stars were available in their method, their linear corrections have an exact tendency yet lack fine details.

Taking advantage of the fact that millions of stellar spectra and their associated precise stellar atmospheric parameters are now available, Yuan et al. (2015c) has proposed a spectroscopy-based stellar color regression (SCR) method to achieve a mmag-precision color calibration of wide-field imaging surveys. The method was applied to the Sloan Digital Sky Survey (SDSS; York et al. 2000) Stripe 82 data, achieving a precision of ~2–5 mmag for different SDSS colors. It is straightforward and able to capture subtle structures thanks to the intensive database. Analogous to the SDSS database, the Large Sky Area Multi-object Fiber Spectroscopic Telescope (LAMOST; Deng et al. 2012; Zhao et al. 2012; Liu et al. 2014) spectroscopic sky survey is also appropriate for this method. It began in 2012, acquiring thousands of stellar spectra per night with the spectral resolution \( R \sim 1800 \) and accumulating more than 8 million stellar spectra in DR5. Effective temperature \( T_{\text{eff}} \), surface gravity \( \log g \), and metallicity [Fe/H] are delivered from the spectra through the LAMOST Stellar Parameter pipeline (LASP; Wu et al. 2011; Luo et al. 2015) with the precision of about 110 K, 0.2 dex, and 0.1 dex, respectively (Luo et al. 2015). In this work, combining the LAMOST DR5 spectroscopy and the Gaia DR2 photometry, we apply the SCR method and obtain new corrections to the Gaia photometry to an unprecedented precision of about 1 mmag.

The paper is organized as follows. We describe our data selection in Section 2. The method and result are demonstrated in Section 3. The result validation and the comparison with previous researches are discussed in Section 4. We summarize our paper in Section 5.
2. Data

Stars having high-precision photometric colors and well-determined stellar atmospheric parameters are required in our study. Here, we consider the following constraints to guarantee the data quality:

1. \texttt{phot_bp_mean_flux_over_error} > 100
2. \texttt{phot_g_mean_flux_over_error} > 100
3. \texttt{phot_rp_mean_flux_over_error} > 100
4. \texttt{phot_proc_mode} = 0
5. \texttt{duplicated_source} = False
6. \texttt{phot_bp_rp_excess_factor} (see text for limit)
7. \( E(B-V)_{\text{SFD}} < 0.05 \) mag
8. \(|\text{glat}| > 20\text{ deg}\)
9. \(|Z| > 0.2 \text{ kpc}\), where \( Z \) is vertical distance to the Galactic disk
10. the signal-to-noise ratios for the \( g \) band (\( S/N_g \)) of the LAMOST spectra are larger than 30
11. \( T_{\text{eff}} > 4500 \text{ K} \) for dwarf stars
12. remove variable stars.

Among them, criteria (1)–(6) are applied to the Gaia DR2 data. The term \texttt{phot_bp_rp_excess_factor} is defined as the excess of flux in the \( G_{\text{BP}} \) and \( G_{\text{RP}} \) with the respect to the \( G \) band, which is believed to be caused by background and contamination issues affecting the \( G_{\text{BP}} \) and \( G_{\text{RP}} \) data (Evans et al. 2018). We find that it has a significant systematic trend with apparent \( G \) magnitude and [Fe/H]. As shown in the Figure 1, the \texttt{phot_bp_rp_excess_factor} systematically increases with decreasing \( G \) magnitude (about 0.005 per magnitude) and decreases with increasing metallicity (about 0.001 per 0.5 dex) within a given \( G \) magnitude range. Therefore, we bin our sample in the \( G \) magnitude and [Fe/H] space at 1 mag and 0.5 dex intervals, respectively, and apply a 2\( \sigma \)-clipping on the \texttt{phot_bp_rp_excess_factor} to remove outliers in each bin (see Figure 1). About 4% of objects are excluded in this step. Note that due to the relatively large widths of the magnitude and [Fe/H] bins, the estimated \( \sigma \) values are larger than the scatter caused by the errors of the \texttt{phot_bp_rp_excess_factor}. Therefore, a 2\( \sigma \)-clipping is used, which is a bit excessive but secure.

We use the Schlegel, Finkbeiner & Davis (1998, hereafter SFD) dust reddening map and criteria (7)–(9) to select low extinction stars for reliable reddening correction.

The empirically determined temperature- and reddening-dependent reddening coefficients \( R(G-G_{\text{RP}}) \) and \( R(G_{\text{BP}}-G_{\text{RP}}) \) are given by the following functions (Y. Sun et al. 2021, in preparation):

\[
R(G_{\text{BP}}-G) = 0.901 - 0.225 \times E(B-V)_{\text{SFD}} + 0.178 \\
\times E(B-V)^2_{\text{SFD}} - 6.422 \times 10^{-5} \times T_{\text{eff}} - 3.982 \times 10^{-7} \\
\times T_{\text{eff}} \times E(B-V)_{\text{SFD}} + 3.541 \times 10^{-9} \times T_{\text{eff}}^2
\]

(1)

\[
R(G_{\text{BP}}-G_{\text{RP}}) = 1.060 - 0.253 \times E(B-V)_{\text{SFD}} + 0.264 \\
\times E(B-V)^2_{\text{SFD}} + 5.353 \times 10^{-6} \times T_{\text{eff}} - 4.505 \times 10^{-5} \\
\times T_{\text{eff}} \times E(B-V)_{\text{SFD}} + 6.64 \times 10^{-9} \times T_{\text{eff}}^2
\]

(2)

Note that all colors referred to hereafter are dereddened intrinsic ones.

Criteria (10) and (11) aim to select high-quality stellar parameters from the LAMOST data set. For stars with multiple observations, their mean values determined using inverse variance weighting are adopted.

For criterion (12), given that the Gaia satellite scans the sky repeatedly and publishes a mean magnitude for each object, variable stars could be less reliable and should be removed. A total of 112 objects identified as variable stars in the WISE (Chen et al. 2018b), ATLAS (Heinze et al. 2018), ASAS-SN (Jayasinghe et al. 2020), Catalina (Drake et al. 2017), and ZTF (Chen et al. 2020) surveys are removed in this step.

Finally, we divide the above sample into dwarfs and giants, considering that they probably suffer different systematic errors in effective temperature and metallicity. Dwarf stars are selected from the \( T_{\text{eff}} \) versus log \( g \) diagram as those below the cyan line in Figure 2. For giants, only red giant branch (RGB) stars are selected (with red clump stars removed) by cross-matching the above sample with a value-added RGB catalog of LAMOST DR4 (Wu et al. 2019). The final sample contains 433,845 dwarfs and 55,395 RGB stars, as displayed in Figure 2.

3. Method and Result

The key idea of the SCR method is that stars with the same atmospheric parameters should have the same intrinsic colors, and therefore plenty of stars with precisely determined
spectroscopic parameters can serve as excellent color standards to carry out very precise color calibrations. Yuan et al. (2015c) demonstrate the method with the SDSS Stripe 82 data. In their work, a well-calibrated control sample within a specific area is selected as a reference to define the intrinsic colors as a function of the stellar parameters, which in turn are used to map out the spatial variations of the color zero-point offset as functions of R.A. and decl., achieving a remarkable improvement in color calibration by a factor of two to three. Very recently, Huang et al. (2020) apply the SCR method to the second data release of the SkyMapper Southern Survey (Onken et al. 2019). A uniform calibration with a precision of better than 1% is achieved.

In this work, taking advantages of about a half million common stars selected in the previous section, we apply this method to the Gaia DR2 and obtain the empirical relations between the Gaia colors \((G - G_RP)\) and \((G - G_{RP})\) and the LAMOST DR5 stellar astrophysical parameters \((T_{\text{eff}}, [\text{Fe/H}], \log g)\). Since a systematic effect with the \(G\) magnitude is known, a control sample within a narrow magnitude range is needed as a reference to explore the magnitude-dependent variations. Gaia modifies its instrument configurations (Gates and Windows) according to the \(G\) magnitude. Even though an iterative calibration is performed to establish the internal photometric system, the convergence problems still exist among the different configurations, and stars around these joints may have relatively larger uncertainties (Evans et al. 2018). Considering the distribution of the Gates and Window Classes (Evans et al. 2018), as well as the number of stars, stars with \(13.3 < G < 13.7\) are selected as the control sample, which consists of 50,343 dwarfs and 6144 RGB stars.

Magnitude-dependent systematic effects in the LAMOST stellar parameters could cause false trends with \(G\) magnitudes in our results. We select multiply observed stars in the LAMOST DR5 to investigate such effects and find that there are no systematic effects with \(S/N_g\) until a very low \(S/N_g\) of about 15. The results are plotted in the top panels of Figure 3, which shows a comparison between observations of very different \(S/N_g\). To further verify this result, we cross-match the LAMOST DR5 with the APOGEE DR16 (Ahumada et al. 2020) and select a sample of solar-type stars to plot their stellar parameter differences as a function of LAMOST \(S/N_g\) in the bottom panels of Figure 3. A flat offset is seen for each stellar parameter at \(S/N_g > 15\), consistent with the test result from duplicated stars. Note the offsets are not exactly zero due to systematic errors in the LAMOST and APOGEE stellar parameters.

To precisely construct the relationships between the Gaia colors and the LAMOST stellar parameters, we divide our control sample into different bins as follows: \(T_{\text{eff}}\) in 300 K, \([\text{Fe/H}]\) in 0.2 dex, and \(\log g\) in 0.3 dex for dwarfs; \(T_{\text{eff}}\) in 300 K, \([\text{Fe/H}]\) in 0.5 dex, and \(\log g\) in 0.4 dex for RGB stars. A linear fitting formula is adopted for each grid: \(C = a_0 + a_1 \times T_{\text{eff}} + a_2 \times [\text{Fe/H}] + a_3 \times \log g\), where \(C\) can be \((G_{BP} - G_{RP})_0\) or \((G - G_{RP})_0\). We perform 2\(\sigma\)-clipping during the fitting process. Bins containing fewer than 50 stars are excluded to avoid poor fitting. The fitting residuals are plotted in Figures 4 and 5. The median residuals of both dwarfs and RGB giants are very close to zero. The typical standard deviations are about 7.5 mmag and 15 mmag for \(G - G_{RP}\) and \(G_{BP} - G_{RP}\), respectively.

Applying the empirical relations derived above to the whole sample, color residuals/corrections (fitted colors – observed colors) are obtained. Their distributions in the \(T_{\text{eff}}\) and \((G_{BP} - G_{RP})_0\) plane are shown in Figures 6 and 7. At a given effective temperature, stars with redder intrinsic colors have typically smaller (negative) residuals. To examine the color dependence, both dwarfs and RGB stars are divided into subsamples by the black solid lines in Figures 6 and 7. Four subsamples are adopted for dwarfs (main sequence, hereafter MS), with typical \((G_{BP} - G_{RP})_0\) colors of 1.06, 0.90, 0.77, and 0.66 mag, respectively. Two subsamples are adopted for RGB stars, with typical \((G_{BP} - G_{RP})_0\) colors of 1.13 and 1.01 mag, respectively. The sample numbers from red to blue are 32,655, 10,2441, 16,3325, and 68,493 for dwarfs, and 24,697 and 14,601 for giants. The gray points are discarded because stars there are partly dropped due to the requirement \(N > 50\), and the dramatic change in numbers at the margin can cause a bias in the color residuals.

For each subsample, the distributions of its color residuals against \(G\) magnitudes are plotted in Figures 8 and 9. The points are grouped into bins along the \(G\) magnitude for every 0.05 mag. The median and standard deviation values are overplotted in orange lines. Typical standard deviations are \(\sim 7–9\) mmag for \(G - G_{RP}\) and \(\sim 12–17\) mmag for \(G_{BP} - G_{RP}\). Magnitude-dependent offsets are clearly seen for all subsamples. The offsets (median of the color residuals) are smoothed by the locally estimated scatterplot smoothing (LOWESS) and taken as the calibration curves. The algorithm works by taking the frac \(\times N\) closest points to each data point and estimating smoothed \(y\) values using a weighted linear regression based on their \(x\) distances, where \(N\) is the total number of points and frac varies for subsamples with a typical value of 0.05. Note that the curves for different subsamples may have different magnitude ranges. In places with few sources at the very bright or faint ends, linear extrapolation is performed. Specifically, corrections can be expressed as:

\[
C' = C + \Delta C
\]
where $C'$ is the corrected color, $C$ is the Gaia DR2 color, and $\Delta C$ is the color correction term.

In Figure 8, there are three marked variations at around $G \sim 11.2$, 13.0, and 15.7 mag that are probably related to the changes of instrument configurations. Gates are implemented on the Gaia CCDs to ease the saturation issues in the case of bright stars, e.g., $G < 12$ mag. Sizes of Windows in the along-scan and across-scan directions are modified at $G = 13.0$ and 16 mag for the $G$ band, and at $G = 11.5$ for the $G_{BP}$ and $G_{RP}$ bands (Evans et al. 2018). However, the choice of the
instrument configurations depends on the observed signals in the $G$ band, so the corresponding calibrated $G$ magnitudes are not exactly identical for every single observation. For example, the transition of the $G$ band from Window Class 0 to 1 may move around $G = 13$ mag, and the transition of the $G_{BP}$ and $G_{RP}$ may move around $G = 11.5$ mag, due to issues including flat-fielding and stray light contamination. Since the homogeneous calibration adopted by Gaia is a function of the $G$ magnitude and may differ significantly before and after the transition magnitude, some corrections could be incorrect. Consequently, the random uncertainties around the transitions are larger than those in adjacent regions, as demonstrated by Figure 9 in Evans et al. (2018). The features at $G \sim 11.2, 13.0,$ and $15.7$ mag in Figure 8 are consistent with Figure 9 in Evans et al. (2018), suggesting that these fine structures are real and accurate.

The final results and a comparison of calibrations of different subsamples are plotted in Figure 10. Positions of changes of

Figure 5. Same as Figure 4, but for the RGB stars.

Figure 6. Distributions of the color residuals (fitted colors – observed colors) in the $T_{\text{eff}}$ and $(G_{BP} - G_{RP})_0$ plane for the dwarf stars. The black solid lines mark the boundaries of different subsamples. To avoid the selection effect, only colored points are used. Only 1 in 100 stars are plotted.

Figure 7. Same as Figure 6, but for the RGB stars. Only 1 in 10 stars are plotted.
4. Discussion

4.1. Color Dependence

As shown in Figure 10, despite places that have few sources and relatively larger errors at the very bright ends, the calibration curves of $G - G_{RP}$ yielded by the six different subsamples agree well with each other for $G < 14$ mag. However, for $G > 14$ mag, discrepancies exist between the curves of the two RGB subsamples (RGB 1.13 and RGB 1.01) and the red dwarf subsample (MS 1.06), while those of the other three blue dwarf subsamples (MS 0.90, MS 0.77, and MS 0.66) still agree well. As for the calibration curves for $G_{BP} - G_{RP}$, the same trend exists for $G > 14$ mag. For $G < 11.5$ mag, the curves are divided into two groups (RGB 1.13 and RGB 1.01 versus MS 0.77 and MS 0.66) according to their colors.

The discrepancies of the three red lines (MS 1.06, RGB 1.13, and RGB 1.01) at $G > 14$ mag in Figure 10 are not real, but are caused by the combination of two reasons: the selection function of the LAMOST data and the spatially dependent systematics of the SFD reddening map. The LAMOST surveys have four different types of plates: very bright (VB), bright (B), median bright (M), and faint (F) plates, in which VB plates target stars with $r < 14$ mag and the others target fainter stars (Chen et al. 2018a). Compared to blue stars, bright red stars are more likely to be giants while faint red stars are dwarfs. The
comparisons of the spatial distributions of dwarfs and RGB stars are shown in Figure 11. The whole sky is divided into different bins of 10 degree by 10 degree in size. The percentages \( f_i = \frac{N_i}{N_{total}} \) of RGB stars and dwarfs in each bin are calculated, and the color bars represent the ratios of dwarf percentage to RGB percentage. The closer the ratios get to 1, the smaller the differences between the spatial distributions of RGB stars and dwarfs are. Clearly, the differences are larger for \( G > 14 \) mag. Note that the Gaia \( G \) magnitudes are very close to the SDSS \( r \) magnitudes for typical stars.

On the other hand, the SFD reddening map is found to show systematic errors that depend on spatial position and dust temperature (e.g., Peek & Graves 2010; Y. Sun et al. 2021, in preparation). The top and middle panels of Figure 12 show sky distributions of \( G - G_{RP} \) and \( G_{BP} - G_{RP} \) residuals (fitted colors—observed colors) after correcting for the trends with \( G \) magnitude, which are almost identical and therefore caused by the same problem. The SFD map overestimates reddening in the pink regions (positive residuals) and underestimates reddening in the blue regions (negative residuals), which is consistent with the result of Y. Sun et al. (2021, in preparation). A strong correlation is seen in the bottom panel of Figure 12 that compares the corrected \( G - G_{RP} \) and \( G_{BP} - G_{RP} \) residuals of the whole sample. The red crosses are the median values of two corrected residuals. They agree very well with the cyan line, whose slope is determined by the median value (\( \sim 1.99 \)) of the \( R(G_{BP} - G_{RP})/R(G - G_{RP}) \) ratios of the sample, as expected.

For the \( G_{RP} \) and \( G_{BP} \) bands, even though the Gaia photometric pipeline treats both 2D and 1D Windows as aperture photometry in the same way (Evans et al. 2018), different instrument
configurations may still cause problems. WEI18 points out that convergence during the calibration process may be dominated by stars of the largest number, which can cause discrepancies among colors. They find that $G_{BP}$ magnitude has a systematic tendency among $G_{BP} - G_{RP}$ and the offset happens around 10.99 mag.

MAW18 further confirms that the color-dependent break happens in the $G_{BP}$ band at $G \sim 10.87$ mag by analyzing the uncertainties of the $G_{BP}$ fluxes. Likewise, we believe that the color-dependent feature at $G < 11.5$ mag in our result for the $G_{BP} - G_{RP}$ calibration is due to the same reason.

We combine the three unbiased dwarf subsamples (MS 0.90, MS 0.77, and MS 0.66) together to determine the recommended calibration curves with the same procedure in Section 3. The recommended calibration curves are overplotted in black lines in Figure 10. The differences between the recommended curve and the three colored ones (MS 0.90, MS 0.77, and MS 0.66) are computed, with standard deviations of 0.3 and 0.6 mmag for $G - G_{RP}$ and $G_{BP} - G_{RP}$, respectively. This suggests that the calibration curve of each subsample has a typical random error smaller than 1.0 mmag. Note that the recommended curves work well in most cases except for $G_{BP} - G_{RP}$ when $G < 11.5$ mag, which is plotted as the dashed–dotted line. The recommended color calibration curves as well as the calibration curves of different subsamples are listed in Table 1.

4.2. Comparisons With Previous Works

Evans et al. (2018) have published a set of revised sensitivity curves called REV, which are preferred to, but have no
significant differences from, the DR2 ones. In CV18, with the well-calibrated CALSPEC spectral library, they compute the synthetic $G$, $G_{BP}$, and $G_{RP}$ magnitudes using the REV curves and compare with those in the Gaia DR2 catalog, yielding a linear correction of $3.5 \text{ mmag/mag}$ in the $G$ magnitude. WEI18 reconstructs the Gaia passbands in a basic functional analytic framework with four spectral libraries including the CALSPEC and finds a close magnitude drift with CV18. Later, MAW18 updates the spectral libraries and applies the same method as in WEI18, suggesting a $3.2 \text{ mmag/mag}$ correction in the $G$ magnitude.

Figure 13 compares our results with those of CV18 and MAW18. For the $G - G_{RP}$ color, systematic offsets can be seen between any two of the three references because of the different processing methods: (1) we take $G = 13.5 \pm 0.2 \text{ mag}$ as the control sample; (2) CV18 applies the REV transmission curves and the official zero-points; and (3) MAW18 uses their own transmission curves called MAW and a zero-point of $0$. Both CV18 and MAW18 provide a linear fitting of the systematic trend in the $G$ band. Our result shows a very similar trend, consistent with the results of CV18 and MAW18. However, thanks to the much larger sample of stars used (half a million compared to about one hundred), finer details are revealed in this work, particularly those due to the changes of the instrument configurations as discussed in Section 3. For the $G_{BP} - G_{RP}$ color, our result is also consistent with those of CV18 and MAW18 within their errors. Our result shows a small but clear trend with $G$ magnitude at $G > 11.5 \text{ mag}$. At the bright end, our result is slightly higher, probably due to the fact that most stars there in CV18 and MAW18 are bluer.

### 4.3. Revised Color–Color Diagram

Here we carry out a test of the validity of our calibration curves by comparing the observed and the revised color–color curves.

#### Table 1

| $G$   | recom | MS 1.06 | MS 0.89 | MS 0.77 | MS 0.67 | RGB 1.13 | RGB 1.02 |
|-------|-------|---------|---------|---------|---------|----------|----------|
| 13.01 | 0.57  | 0.34    | 1.37    | 0.53    | 1.46    | 1.07     | 1.84     | 1.13     | 1.27     | 1.52     | 2.05     |
| 13.02 | 0.46  | 0.28    | 1.34    | 0.46    | 1.42    | 0.94     | 1.78     | 1.06     | 1.24     | 1.44     | 2.00     |
| 13.03 | 0.37  | 0.22    | 1.31    | 0.39    | 1.38    | 0.81     | 1.72     | 0.99     | 1.21     | 1.36     | 1.95     |
| 13.04 | 0.28  | 0.16    | 1.28    | 0.32    | 1.34    | 0.68     | 1.66     | 0.92     | 1.18     | 1.28     | 1.90     |
| 13.05 | 0.21  | 0.10    | 1.25    | 0.25    | 1.30    | 0.55     | 1.60     | 0.85     | 1.15     | 1.20     | 1.85     |
| 13.06 | 0.14  | 0.04    | 1.22    | 0.18    | 1.26    | 0.42     | 1.54     | 0.78     | 1.12     | 1.12     | 1.80     |
| 13.07 | 0.09  | 0.11    | 1.22    | 0.29    | 1.48    | 0.71     | 1.09     | 0.14     | 1.75     |           |         |
| 13.08 | 0.05  | 0.04    | 1.16    | 0.04    | 1.18    | 0.16     | 1.42     | 0.64     | 1.06     | 0.96     | 1.70     |
| 13.09 | 0.01  | 0.08    | 1.13    | 0.03    | 1.14    | 0.03     | 1.36     | 0.57     | 1.03     | 0.88     | 1.65     |
| 13.10 | 0.01  | 0.02    | 1.10    | 0.10    | 1.10    | 0.10     | 1.30     | 0.50     | 1.00     | 0.80     | 1.60     |
| 13.11 | 0.02  | 0.01    | 1.07    | 0.09    | 1.07    | 0.08     | 1.26     | 0.47     | 0.98     | 0.76     | 1.57     |
| 13.12 | 0.02  | 0.01    | 1.04    | 0.08    | 1.04    | 0.06     | 1.22     | 0.44     | 0.96     | 0.72     | 1.54     |
| 13.13 | 0.02  | 0.01    | 1.01    | 0.07    | 1.01    | 0.04     | 1.18     | 0.41     | 0.94     | 0.68     | 1.51     |
| 13.14 | 0.01  | 0.01    | 0.98    | 0.06    | 0.98    | 0.02     | 1.14     | 0.38     | 0.92     | 0.64     | 1.48     |
| 13.15 | 0.00  | 0.01    | 0.95    | 0.05    | 0.95    | 0.00     | 1.10     | 0.35     | 0.90     | 0.60     | 1.45     |
| 13.16 | 0.02  | 0.01    | 0.92    | 0.04    | 0.92    | 0.02     | 1.06     | 0.32     | 0.88     | 0.56     | 1.42     |
| 13.17 | 0.03  | 0.01    | 0.98    | 0.03    | 0.98    | 0.04     | 1.02     | 0.29     | 0.86     | 0.52     | 1.39     |
| 13.18 | 0.04  | 0.01    | 1.02    | 0.12    | 0.86    | 0.04     | 1.02     | 0.29     | 0.86     | 0.52     | 1.39     |
| 13.19 | 0.05  | 0.01    | 1.06    | 0.11    | 0.83    | 0.03     | 1.02     | 0.29     | 0.86     | 0.52     | 1.39     |
| 13.20 | 0.06  | 0.01    | 1.10    | 0.10    | 0.80    | 0.05     | 1.02     | 0.29     | 0.86     | 0.52     | 1.39     |

Note. The first column is $G$ magnitude. The second and third columns are, respectively, recommended $G - G_{RP}$ and $G_{BP} - G_{RP}$ calibration curves, followed by those yielded by different subsamples. All calibration curves are in units of mmag.

This table is available in its entirety in machine-readable form.
The recommended blank solid curves are adopted to correct $G - G_{\text{RP}}$ and $G_{\text{BP}} - G_{\text{RP}}$ for stars with $G > 11.5$ mag. For stars with $G < 11.5$ mag, $G_{\text{BP}} - G_{\text{RP}}$ colors are corrected by colored curves. For stars of $G_{\text{BP}} - G_{\text{RP}} > 1.13$ mag, the calibration curve of $G_{\text{BP}} - G_{\text{RP}} = 1.13$ mag is used. We correct the whole sample using Equation (3) and compare color–color diagrams before and after correction in the left and middle panels in Figure 14. To exhibit the metallicity-dependent stellar locus clearly, only 300 stars in each metallicity bin are plotted and colored by their metallicities. After correction, the distribution of stars with similar $[\text{Fe/H}]$ gets tighter, as expected. Following the work of Yuan et al. (2015a), a fourth-order polynomial is adopted to fit the metallicity-dependent stellar color loci of the Gaia colors. The histograms of the fitting residuals are plotted in the right panels of Figure 14. Both dwarfs and giants show a much smaller $\sigma$ after correction, from 2.51 mmag to 1.76 mmag for dwarfs and from 2.59 mmag to 1.55 mmag for giants. The distribution for giants also becomes more symmetric and Gaussian. Like in Figure 31 in Arenou et al. (2018), we also plot a 2D histogram of the $G - G_{\text{BP}}$ residuals in Figure 15 after subtracting the metallicity-dependent stellar color locus. The subplot labeled DR2 is identical to Arenou et al. (2018), showing a clear trend with $G$ magnitude. After the linear correction in the $G$ magnitude proposed by CV18, the overall trend disappears, but still shows small-scale features at the level of a few mmag. Likewise, the liner corrections suggested by WEI18 and MAW18 suffer the same problem. After the corrections of this work, the trend is flat everywhere.

The revised color–color diagram is essential in a number of studies, for instance, detecting peculiar objects or determining reliable metallicities for an enormous and magnitude-limited sample of stars from the Gaia photometry (S. Xu et al. 2021, in preparation). We notice that the $G_{\text{BP}} - G$ residual distribution of dwarfs after correction is asymmetric at its wings. There is an excess of stars whose colors are redder compared to those predicted by the metallicity-dependent color locus. This asymmetric features can be well explained by binary stars (Yuan et al. 2015b) and used to estimate binary fractions for a huge number of field stars (Z. X. Niu et al. 2021, in preparation).

5. Summary

In this work, using about a half million stars with high-quality Gaia DR2 photometry and LAMOST DR5 stellar parameters, we apply the SCR method to calibrate the Gaia $G - G_{\text{RP}}$ and $G_{\text{BP}} - G_{\text{RP}}$ colors. An unprecedented precision of about 1 mmag is achieved, suggesting the great power of the SCR method in calibrating photometric surveys. Magnitude-dependent trends are revealed for both the $G - G_{\text{RP}}$ and $G_{\text{BP}} - G_{\text{RP}}$ colors in great detail, reflecting changes in instrument configurations. Color-dependent trends are found for the $G_{\text{BP}} - G_{\text{RP}}$ color and for stars brighter than $G \sim 11.5$ mag. The maximum correction term of the calibration is about 20 mmag in general and varies by a few mmag/mag. Our results are consistent with previous results but with a much improved precision. Implementations of the color and magnitude corrections are provided. A revised color–color diagram of Gaia DR2 is given to demonstrate some potential applications of the corrected mmag photometry. The $G$, $G_{\text{BP}}$, and $G_{\text{RP}}$ passbands are different in Gaia EDR3 in the same way that the $G$ passbands were different between DR1 and DR2. We will extend the work to Gaia EDR3 data in a separate paper.
We acknowledge the anonymous referee for valuable comments that improved the quality of this paper significantly. We acknowledge Profs. L. Casagrande and J. Maíz Apellániz for providing data of their work. We greatly thank Huiqin Yang for a careful reading of the manuscript. This work is supported by National Science Foundation of China (NSFC) under grant Nos. 11988101, 11603002, and 113300034, National Key Research and Development Program of China (NKRDP) under grant Nos. 2016YFA0400804, 2019YFA0405503, and 2019YFA0405504.

This work has made use of data products from the Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST). LAMOST is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

ORCID iDs
Zexi Niu @ https://orcid.org/0000-0002-3651-0681
Haibo Yuan @ https://orcid.org/0000-0003-2471-2363

References
Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17
Casagrande, L., & VandenBerg, D. A. 2018, MNRAS, 479, L102
Chen, B. Q., Liu, X. W., Yuan, H. B., et al. 2018a, MNRAS, 476, 3278
Chen, X., Wang, S., Deng, L., de Grijs, R., & Yang, M. 2018b, ApJS, 237, 28
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 735
Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, MNRAS, 469, 3688
Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, AJ, 156, 241
Huang, Y., Yuan, H., Li, C., et al. 2020, arXiv, arXiv:2011.0712
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2020, MNRAS, 491, 13
Liu, X. W., Yuan, H. B., Huo, Z. Y., et al. 2014, in IAU Symp. 298, Setting the Scene for Gaia and LAMOST, ed. S. Feltzing et al. (Cambridge: Cambridge Univ. Press), 310
Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095
Maíz Apellániz, J., & Weiler, M. 2018, A&A, 619, A180
Onken, C. A., Wolf, C., Bessell, M. S., et al. 2019, PASA, 36, e033
Peek, J. E. G., & Graves, G. J. 2010, ApJ, 719, 415
Riello, M., De Angeli, F., Evans, D. W., et al. 2018, A&A, 616, A3
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Weiler, M. 2018, A&A, 617, A138
Wu, Y., Luo, A. L., Li, H.-N., et al. 2011, RAA, 11, 924
Wu, Y., Xiang, M., Zhao, G., et al. 2019, MNRAS, 484, 5315
York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579
Yuan, H., Liu, X., Xiang, M., et al. 2015b, ApJ, 799, 135
Yuan, H., Liu, X., Xiang, M., et al. 2015c, ApJ, 799, 133
Yuan, H., Liu, X., Xiang, M., Huang, Y., & Chen, B. 2015a, ApJ, 799, 134
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723

Figure 15. A 2D histogram of the $G - G_{BP}$ residual after subtracting the metallicity-dependent color locus. Top: published DR2 data. Middle: the same after applying CV18 corrections. Bottom: applying corrections from this work.