Emission line star catalogues post-Gaia DR3

A validation of Gaia DR3 data using the LAMOST OBA emission catalogue

B. Shridharan¹, B. Mathew¹, S. Bhattacharyya¹, T. Robin¹, R. Arun², S. S. Kartha¹, P. Manoj³, S. Nidhi¹, G. Maheshwar², K. T. Paul¹, M. Narang³, and T. Himanshu³

¹ Department of Physics and Electronics, CHRIST (Deemed to be University), Hosur Main Road, Bangalore, India
e-mail: shridharan.b@res.christuniversity.in
² Indian Institute of Astrophysics, Koramangala, Bangalore, India
³ Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai, India

Received 26 June 2022 / Accepted 22 September 2022

ABSTRACT

Aims. Gaia Data Release 3 (DR3) and further releases have the potential to identify and categorise new emission-line stars in the Galaxy. We perform a comprehensive validation of astrophysical parameters from Gaia DR3 with the spectroscopically estimated emission-line star parameters from the LAMOST OBA emission catalogue.

Method. We compare different astrophysical parameters provided by Gaia DR3 with those estimated using LAMOST spectra. By using a larger sample of emission-line stars, we performed a global polynomial and piece-wise linear fit to update the empirical relation to convert the Gaia DR3 pseudo-equivalent width to the observed equivalent width, after removing the weak emitters from the analysis.

Results. We find that the emission-line source classifications given by DR3 is in reasonable agreement with the classification from the LAMOST OBA emission catalogue. The astrophysical parameters estimated by the esphs module from Gaia DR3 provides a better estimate when compared to gspphot and gspspec. A second degree polynomial relation is provided along with piece-wise linear fit parameters for the equivalent width conversion. We notice that the LAMOST stars with weak Hα emission are not identified to be in emission from BP/RP spectra. This suggests that emission-line sources identified by Gaia DR3 are incomplete. In addition, Gaia DR3 provides valuable information about the binary and variable nature of a sample of emission-line stars.

Key words. stars: emission-line, Be – catalogs; stars: variables: T Tauri, Herbig Ae/Be – methods: data analysis – techniques: spectroscopic

1. Introduction

Emission-line stars (ELS) are a class of objects with emission lines, particularly Hα, at 6563 Å in the spectrum. They also exhibit physical processes such as stellar winds, jets or outflows, and/or mass accretion through the circumstellar disc. The hot ELS are mainly classified into main-sequence classical Be (CAe/Be; Rivinius et al. 2013) and pre-main-sequence (PMS) Herbig Ae/Be (HAE/Be; Waters & Waelkens 1998) stars based on their evolutionary stage. Many large sky surveys, such as 2MASS (Cutri et al. 2003), WISE (Cutri et al. 2012), and IPHAS (Drew et al. 2005), have improved the ELS research by providing precise photometric measurements which are used to classify the ELS into various categories (Koenig & Leisawitz 2014; Witham et al. 2008).

The Gaia Data Release 3 (Gaia DR3) catalogue represents a substantial advance in Galactic stellar astronomy. Gaia DR3 (Gaia Collaboration 2021) builds on previous releases by improving the quality of previously released data and introducing entirely new data products, such as mean dispersed BP/RP spectra from spectro-photometry and radial velocity spectra (RVS), in addition to their integrated photometry in G_bp, G_rp, and the white light G band published in Gaia EDR3 (De Angeli et al. 2022). Gaia BP/RP spectra and/or RVS are now available for sources with G < 19 mag, and astrophysical parameters for sources with G < 17.6 mag.

The previous Gaia releases played a pivotal role in identifying and studying new populations of ELS in the Galaxy. Some notable examples are the selection of 11 000 high confidence PMS stars from the Sco OB2 association (Damiani et al. 2019), and understanding the dynamics of young stellar objects (YSOs) in the Vela OB association (Cantat-Gaudin et al. 2019). The Spitzer/IRAC Candidate YSO (SPICY) catalogue was compiled from the YSO candidates identified using the high-quality astrometric data from Gaia EDR3 along the Galactic midplane (Kuhn et al. 2022). More homogeneous studies on the stellar parameters of YSOs were carried out by Arun et al. (2019) and Wichittanakom et al. (2020) using Gaia DR2, and Guzmán-Díaz et al. (2021) and Vioque et al. (2022) using Gaia EDR3. Even though Gaia has extensively improved stellar parameters of the previously known ELS in the Milky Way, the unavailability of Hα emission measurements for the Gaia sources hindered the classification of more ELS.

The Large sky Area Multi-Object fibre Spectroscopic Telescope (LAMOST) has observed and catalogued 10 431 197 spectra of astronomical sources in their latest DR7 data release. Due to the availability of such a large database of spectra, the number of newly identified ELS has improved. Hou et al. (2016) identified 10 436 early-type ELS using LAMOST DR2 and studied various Hα profiles. Shridharan et al. (2021, hereafter, called as LEMC) compiled a catalogue of 3339 hot ELS from 451 695 O-, B-, and A-type spectra from the LAMOST DR5.
release. After careful spectral type re-estimation, they reported 1088 CBe, 233 CAe, and 56 HAeBe stars based on the analysis of optical/IR magnitudes and colours. This makes it one of the largest homogeneous ELS catalogues with a thorough classification using spectroscopy and available photometry. More recently, Zhang et al. (2022) identified 25,886 early-type ELS from LAMOST DR7. Even though the number of ELS objects increased with such large spectroscopic surveys, they cannot be classified accurately unless astrometric and photometric data are available. Hence, the field of ELS improves when the large spectroscopic surveys and all-sky astrometric surveys progress in tandem. This is achieved by the recently released Gaia DR3 data which provides astrometric, photometric, and spectroscopic parameters for more than 200 million objects. There is no doubt that the DR4 and further releases will greatly improve the ELS research.

As the first step in this direction, we compare the new dataset released by Gaia DR3 with a previously existing, well-characterised spectroscopic catalogue. In this work, we aim to provide an external validation for the astrophysical parameters and to improve our ELS catalogue with newly available data from Gaia DR3.

2. Data analysis and results

We used the 3339 ELS from LEMC and queried various DR3 tables using the source identifier from EDR3. The query was made using the ADQL facility in the Gaia archive\(^1\). We have explored the different datasets that Gaia provides with its new release.

2.1. Classification and astrophysical parameters from Gaia DR3

The gaiadr3.astrophysical_parameters table provides plenty of information using the BP/RP spectra; the details of which can be found in Fouesneau et al. (2022, hereafter APSIS-II). The comparison between the sub-classification of the ELS reported in LEMC with the classification done using the Extended Stellar Parametrizer for Emission-Line Stars (ESP-ELS) module of Gaia DR3 (mentioned as classlabel_espels), for a sample of 506 stars, is shown as a heatmap in the top panel of Fig. 1. The bottom panel of Fig. 1 shows the heatmap of the spectral type comparison between 3109 ELS from LEMC with those estimated from the Extended Stellar Parametrizer for Hot Stars (ESP-HS) module in Gaia DR3 (denoted as spectraltype_esphs).

From the figure (Fig. 1: top panel), we see that the classification provided by LEMC and classlabel_espels DR3 matches well. Of 315 CBe stars with Gaia DR3 estimates, 303 (96%) stars are classified as ‘BeStar’, 11 (4%) stars as ‘HerbigStar’, and one star as ‘WN’ by Gaia DR3. The quality of the classlabel_espels classification is given by classlabel_espels_flag, where classlabel_espels_flag <= 2 denotes a probability larger than 50%. Interestingly, the 11 stars which are classified as ‘HerbigStar’ have a quality flag classlabel_espels_flag >= 4. For the 303 stars classified as ‘BeStar’, 155 stars have classlabel_espels_flag <= 2 and 148 have classlabel_espels_flag > 2. The sub-sample of 89 stars with unclear classifications in LEMC (Be**, Em*)\(^2\) can now be classified as ‘BeStar’ (83) and ‘HerbigStar’ (6).

The bottom panel of Fig. 1 shows the comparison between the spectral type given in LEMC and those estimated by Gaia DR3, spectraltype_esphs. It can be seen that stars with spectraltype_esphs = ‘B’, the spectral type estimated, reasonably match the LEMC spectral types ranging from O (<1%), B0–B5 (25%), B5–B9 (57%), and to A0–A5 (17%). However, the problem with spectraltype_esphs can be seen clearly when we consider the stars with LEMC spectral type B8 (767 stars). Of the 767 stars, 255 (33%) stars are classified by Gaia DR3 to be spectraltype_esphs = F/G/K. This is a very significant deviation from the accurate spectral type

\(^1\) https://gea.esac.esa.int/archive/

\(^2\) Be** = LEMC B-type star, but no detection in Gaia EDR3.
Be* = LEMC B-type star with a Gaia EDR3 detection, but not in 2MASS.
Em* = Hα emission object for which the spectral type could not be calculated.
given in LEMC, which was performed through a semi-automated template matching technique. The deviation of 33% towards later spectral types should be kept in mind before using the spectraltype_esphs in future studies. A possible explanation for the observed deviation can be the line-of-sight extinction. Thirty-three percent of the B8 stars misclassified by Gaia DR3 as F/G/K have higher extinction values in both Green’s 3D dustmap (Green et al. 2019) and Gaia DR3 (Ag–DR3), whereas the extinction value for 59% of the B8 stars classified to have a B spectral type is within 0–1 mag. Thus, the higher the observed extinction value is, the higher the chances of Gaia B spectral type is within 0–1 mag. Thus, the higher the observed extinction value is, the higher the chances of Gaia DR3 spectral type estimation being different from the spectral type in LEMC.

Gaia DR3 provides several astrophysical parameters such as $T_{\text{eff}}$, logg, $V_{\text{sin}}$, mass, radius, and luminosity based on the BP/RP spectrum. For hot stars and ELS, they have used special modules to estimate these parameters. We compared all the different $T_{\text{eff}}$ estimates with our spectral type to identify the best value for hot ELS. It should be noted that spectral type estimates from LEMC, although performed meticulously, have errors of about ±2 subtypes. Figure 2 shows the distribution of various $T_{\text{eff}}$ and logg estimates of ELS available from Gaia DR3 with the spectral type estimated in LEMC. It is very evident from Fig. 2 (top) that for B-type stars, $T_{\text{eff}}$ is significantly underestimated using RVS (teff_gspspec). Two different modules were used to estimate $T_{\text{eff}}$ using BP/RP spectra, that is, teff_gsphopt and teff_esphs. Figure 2 (top) reveals that the $T_{\text{eff}}$ esphs value matches better when compared to $T_{\text{eff}}$ from the Pecaut & Mamajek (2013) calibration table and, also, it has a significantly lower inter-quartile range (IQR) when compared to teff_gsphopt. We notice a large number of outliers in the teff_gsphopt boxplot for each spectral type, which puts its validity into question. Hence it is clear from our analysis that teff_esphs provides a better $T_{\text{eff}}$ estimate for B-type stars. In addition, there are other $T_{\text{eff}}$ estimates available from modules such as teff_gspphot_marcs, teff_gspphot_ob, and teff_gspphot_a in the gaiadr3.astrophysical_parameters_supp table. An appropriate model selection can be done based on the object of interest.

Similarly, Fig. 2 (bottom) shows the distribution of log g values for a sub-sample of LEMC stars. The log g estimate from RVS (logg_gspspec) shows a large scatter when compared to the log g estimates from BP/RP spectra, that is, logg_gspphot and logg_esphs, which are distributed in the range 3–4 dex. Since the LEMC sample contains mainly CBe and HAEbe stars, it is fair to expect log g to be within 3–5. Hence we conclude that, when compared to other modules used in Gaia DR3, the ESPHS module provides accurate astrophysical parameters and can be used for the analysis of OBA stars. According to Frémat et al. (2022), the Vsin i estimations from the vbroad module degrades noticeably at $T_{\text{eff}} > 7500$ K and $G_{RVS} > 10$. Therefore, Vsin i would be highly inaccurate for our sample of hot ELS. Consequently, we have not included a Vsin i analysis in the present study.

2.2. Comparison of Gaia DR3 pEW with the EW from LAMOST spectra

Gaia DR3 made available pseudo-equivalent width (pEW) measurements of Hα for about 235 million sources, which are given in the Gaia DR3 astrophysical_parameters table as the ew_espels_halpha parameter. The classification and the ELS catalogue provided by Gaia DR3 are dependent on this pEW calculation. However, due to the low resolution of BP/RP spectra, using pEW solely may not provide a complete list of ELS which can be identified from Gaia DR3. Hence it is important to calibrate pEW values with actual EW measurements carefully. APSIS-II provides an empirical relation between pEW and the
EW values available from various ELS catalogues in the literature (Fig. 21 and Table 3 of APSIS-II). They estimated the slope of the linear fit to be in the range of 2.26 and 2.83, which can be used to convert the pEW to actual Hα EW.

We improved upon this analysis by performing a second degree polynomial fit to a large sample of 1088 CBe stars from LEMC. Even though we have a bigger sample of 3339 ELS, we did not attempt to perform a fit with other classes to avoid problems such as emission inside the absorption core (Caε stars; Anusha et al. 2021), low number statistics (HaBe), and contamination from [NII] forbidden lines. We used the sample of 1088 CBe stars from LEMC for which the EW were measured homogeneously using IRAF (Anusha et al., in prep.). Stars showing a Hα emission peak inside the absorption core are shown (light blue diamonds) in Fig. 3 and were not used in the analysis. We emphasise here that Gaia DR3 identifies the Hα to be in emission only if the emission peak is above the local continuum. Thus, for B-type stars, Gaia DR3 can identify sources as ELS only if the observed EW is greater than 0.5 nm. For A-type stars, the threshold value will only increase, since the Hα absorption peaks at A0 spectral type (Gray & Corbally 2009). Hence the catalogue of ELS provided by Gaia DR3 may not be complete with weak emitters, specifically those with an emission peak inside the absorption core. This is a known caveat owing to the very low resolution of BP/RP spectra (Martayan et al. 2008). The second degree polynomial relation is shown in Eq. 1:

\[
EW_{\text{LEMC}}(\text{nm}) = -0.54 + 1.60 \times pEW_{\text{Gaia}} - 0.49 \times (pEW_{\text{Gaia}})^2 \text{ (nm)}. \tag{1}
\]

Since we have larger sample of CBe stars when compared to Silaj et al. (2010) and Raddi et al. (2015), we also performed a piece-wise linear fit in intervals of 0.5 nm. The slope and intercept of the linear fit along with the IQR (EW_{\text{LEMC}}) and median absolute deviation (MAD) along the EW_{\text{LEMC}} axis as a representative of the scatter are given for each interval range. A global polynomial fit and a piece-wise fit for different intervals of pEW values are shown in Fig. 3. As seen from piece-wise linear fit values, the slope gets steeper as we move towards intense emitters. We suggest using the respective slopes and intercepts for calculating observed EWs from each pEW range for hot ELS (Fig. 3). However, the users should be aware that LAMOST and Gaia have obtained the spectra at different epochs; the scatter and deviation of some points can be attributed to the intrinsic variability of some CBe stars that range in the orders of days to weeks (Mathew & Subramaniam 2011; Cochetti et al. 2021). The addition of a pEW measurement in DR3 will improve the sample of ELS and can serve as a target list for future Hα ELS surveys.

### 2.3. Synthetic photometry from BP/RP spectra

For 686 stars in the LEMC, we could not estimate the spectral type due to the low signal-to-noise ratio (S/N) in the bluer region of the LAMOST spectra. Due to the observation strategy of LAMOST DR5, the majority of our sample is towards the...
galactic anti-centre direction (Fig. 1 of Shridharan et al. 2021). Limited photometric survey footprints towards this region inhibited us from studying these stars photometrically or using SEDs to estimate their stellar parameters. For our sample of 3339 ELS, 2872 stars have continuous BP/RP spectra. The gaiaxpy package enables the user to calculate the synthetic magnitudes based on the continuous BP/RP spectra of variable stars with a good quality classification (> 0.6) are shown in Fig. 5. The cause of variability can also be related to the evolving nature of the Hα emission region and hence, a detailed analysis of these stars will be carried out in a future work.

3. Summary

The newly released Gaia DR3 data will accelerate the field of astronomy as they provide astrophysical parameters for 470 759 263 sources using the mean BP/RP spectra. Of which, 2 382 015 sources are classified as hot stars which can increase the number of known CBε and HAeBe stars. As a first step towards achieving this, we compared the astrophysical parameters provided by DR3 with carefully classified OBA-type ELS identified from LAMOST DR5.

We see that the ELS classification provided by Gaia DR3 as classlabel_elseps matches with our LEMC catalogue reasonably well. Gaia DR3 also provides a new classification and spectral type estimate for stars classified as ‘Em’ and ‘Em[e]’ in the LEMC catalogue. The mismatch between the spectral types provided by Gaia DR3 (spectraltype_elsps) and LEMC was evident upon comparison. The spectraltype_elsps estimates should be used with caution along with the quality flag provided. Gaia DR3 also provides T_{eff} from three different modules using both BP/RP spectra and RVS. Based on our comparison of T_{eff} values with spectral types from the LEMC catalogue, we see that teff_elsps values match well with the theoretical values. The teff_gspspec values are severely underestimated for early B-type stars. Similarly, the teff_gspphot estimate may not be reliable because of the scatter and high number of outliers. We conclude that teff_elsps should be used as the T_{eff} estimate for early-type ELS.

We used the sample of 1088 CBε stars from LEMC to perform a global polynomial fit and piece-wise fit analysis to obtain a relation to convert the pEW to the actual Hα EW. In cases where one needs a more accurate estimate of actual Hα EW for a specific range of pEW, the piece-wise slope and intercept values can be used. It should be noted that the weak emitters (with an emission peak inside the absorption core) in LEMC have positive pEW values in Gaia DR3. This directly implies the incompleteness of the ELS catalogue provided by Gaia DR3.

We also checked for non-single stars and variable stars present in the LEMC catalogue. Among our sample, there are ten non-single stars with seven of them being classified as spectroscopic binaries for which various parameters are provided. From LEMC, 363 stars are classified as variables. These Hα emitting binaries and variable ELS will be studied in a future work.

To summarise, this work provides an account of how the data provided by the recent Gaia DR3 can improve the study of ELS. Along with photometry and astrometric measurements, the availability of BP/RP spectra for a large number of sources will increase the number of already known ELS. The astrophysical parameters estimated from the BP/RP and RVS will help to study a large number of ELS with ease.
Fig. 5. G-band multi-epoch photometry of stars classified as variables. The different subplots show the different variable classes as provided by Gaia DR3 with the class specification shown in red letters. The y axis is limited to a range of two mag to visualise the variability in each class. The description on each variable class is provided in Table 2 of Eyer et al. (2022).

Acknowledgements. We would like to thank the Science & Engineering Research Board (SERB), a statutory body of the Department of Science & Technology (DST), Government of India, for funding our research under Grant Number CRG/2019/005380. We thank the Center for Research, CHRIST (Deemed to be University), Bangalore, India, for funding our research under the Grant Number MRP DSC-1932. RA acknowledges the financial support from SERB POWER fellowship Grant SPF/2020/00009. 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. Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. We thank the SIMBAD database and the online VizieR library service for helping us with the literature survey and obtaining relevant data.

References
Anusha, R., Mathew, B., Shridharan, B., et al. 2021, MNRAS, 501, 5927
Arun, R., Mathew, B., Manoj, P., et al. 2019, AJ, 157, 159
Arun, R., Mathew, B., Maheswar, G., et al. 2021, MNRAS, 507, 267
Bhattacharyya, S., Mathew, B., Ezhikode, S. H., et al. 2022, ApJ, 933, L34
Cantat-Gaudin, T., Jordi, C., Wright, N. J., et al. 2019, A&A, 626, A17
Chini, R., Barr, A., Buda, L. S., et al. 2013, Cent. Eur. Astron. Phys. Bull., 37, 295
Cochetti, Y. R., Arias, M. L., Kraus, M., et al. 2021, A&A, 647, A164
Cutri, R., Skrutskie, M., Van Dyk, S., et al. 2003, The IRSA 2MASS All-Sky Point Source Catalog VizieR On-line Data Catalog: II/246
Cutri, R., et al. 2012, VizieR Online Data Catalog: II/311
Dami, F., Prisinzano, L., Pallottini, L, Micela, G., & Sciortino, S. 2019, A&A, 623, A112
De Angeli, F., Weiler, M., Montegriffo, P., et al. 2022, A&A, in press, http://doi.org/10.1051/0004-6361/202243659
Drew, J. E., Greimel, R., Irwin, M. J., et al. 2005, MNRAS, 362, 753
Eyer, L., Audard, M., Holl, B., et al. 2022, A&A, submitted, [arXiv:2206.06416]
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2022, A&A, in press, http://doi.org/10.1051/0004-6361/202243919
Frémat, Y., Royer, F., Marchal, O., et al. 2022, A&A, in press, http://doi.org/10.1051/0004-6361/202243899
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
Gray, R. O., & Corbally, C. J. 2009, Stellar Spectral Classification (Princeton University Press)
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Guzmán-Díaz, J., Mendigutía, I., Montesinos, B., et al. 2021, A&A, 650, A182
Hou, W., Luo, A. L., Hu, J.-Y., et al. 2016, Res. Astron. Astrophys., 16, 138
Koenig, X., & Leisawitz, D. 2014, ApJ, 791, 131
Kuhn, M. A., Saber, R., Povich, M. S., et al. 2022, AJ, accepted [arXiv:2206.04999]
Martayan, C., Frémat, Y., Blomme, R., et al. 2008, in SF2A-2008, eds. C. Charbonnel, F. Combes, & R. Samadi, 499
Mathew, B., & Subramaniam, A. 2011, Bull. Astron. Soc. India, 39, 517
Pecaut, M. J., & Mamajek, E. E. 2013, ApJS, 208, 9
Raddi, R., Dreiz, J., Steeghs, D., et al. 2015, MNRAS, 446, 274
Rivinius, T., Carciofi, A. C., & Martayan, C. 2013, A&ARv, 21, 1
Shridharan, B., Mathew, B., Nidhi, S., et al. 2021, Res. Astron. Astrophys., 21, 28
Silaj, J., Jones, C., Tycner, C., Sigut, T., & Smith, A. 2010, ApJS, 187, 228
Vioque, M., Oudmaijer, R. D., Wichittanakom, C., et al. 2022, ApJ, 930, 39
Waters, L. B. F. M., & Waelkens, C. 1998, ARA&A, 36, 233
Wichittanakom, C., Oudmaijer, R. D., Fairlamb, J. R., et al. 2020, MNRAS, 493, 234
Witham, A., Knigge, C., Drew, J., et al. 2008, MNRAS, 384, 1277
Zhang, Y.-J., Hou, W., Luo, A. L., et al. 2022, ApJS, 259, 38