

# XP-TEAL: *Gaia* XP tool for emission and absorption lines

## I. Balmer lines in open clusters

J. M. Carrasco<sup>1,2,3,\*</sup>, M. Weiler<sup>1,2,3</sup>, S. Malhotra<sup>1,2,3</sup>, F. Anders<sup>1,2,3</sup>, C. Jordi<sup>3,4</sup>, E. Masana<sup>1,2,3</sup>,  
J. Castañeda<sup>5,2,3</sup>, and F. Boix<sup>1,2,3</sup>

<sup>1</sup> Departament de Física Quàntica i Astrofísica (FQA), Universitat de Barcelona (UB), Martí i Franquès 1, 08028 Barcelona, Spain

<sup>2</sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (UB), Martí i Franquès 1, 08028 Barcelona, Spain

<sup>3</sup> Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels (Barcelona), Spain

<sup>4</sup> Institut d'Estudis Catalans, c. del Carme, 47, 08001 Barcelona, Spain

<sup>5</sup> DAPCOM Data Services, c. dels Vilabella 5–7, 80500 Vic, Barcelona, Spain

Received 18 March 2026 / Accepted 25 May 2026

### ABSTRACT

**Context.** The third *Gaia* data release (*Gaia* DR3) includes blue and red low-resolution spectrophotometry for more than 200 million sources. These *Gaia* XP spectra are represented by sets of coefficients multiplying linear combinations of basis functions related to Hermite functions. In a previous work we introduced a method that maximally exploits the information encoded in the *Gaia* spectrophotometric coefficients in order to identify absorption and emission lines.

**Aims.** Our aim is to detect hydrogen Balmer emission and absorption lines in stars belonging to open clusters included in the released *Gaia* DR3 blue and red spectrophotometry, and to characterise the Balmer line strengths as a function of colour.

**Methods.** We refined our XP tool for emission and absorption lines (XP-TEAL), using derivatives of the internal spectral coefficients to analyse hydrogen Balmer lines in the *Gaia* DR3 XP spectra. We tested our results using stars in open clusters, considering both absorption and emission features.

**Results.** The analysis of the hydrogen lines in the spectra of open cluster stars allows us to assess the performance and limitations of the method. We find an empirical dependence of the absorption-line equivalent widths on the intrinsic colour of the stars, which allows us to derive precise colour excesses caused by interstellar extinction. We also identify 1256 stars with emission lines, 356 of which have not previously been reported in the literature.

**Conclusions.** We successfully tested our XP-TEAL code on stars in clusters and have made both the source code and the online version available to the community. Despite the low resolution of *Gaia* spectrophotometry, our code is able to extract useful line information.

**Key words.** line: profiles – methods: data analysis – techniques: spectroscopic – catalogs – stars: emission-line, Be – open clusters and associations: general

## 1. Introduction

The third *Gaia* data release (*Gaia* DR3) (Gaia Collaboration 2023, 2016) provides, for the first time, low-resolution blue (BP) and red (RP) spectrophotometry for approximately 219 million sources (De Angeli et al. 2023). Most of these sources have apparent *G* magnitudes lower than 17.65 mag, but the catalogue also includes some fainter objects, such as white dwarfs and quasars (De Angeli et al. 2023). The BP spectral range extends from 330 nm to 680 nm, while the RP spectral range extends from 640 nm to 1050 nm (Carrasco et al. 2021). Both spectrophotometers have resolving powers below  $R = \lambda/\Delta\lambda = 100$ , depending on wavelength (Carrasco et al. 2021). In the following, we use the term XP when referring to BP or RP spectra without distinction.

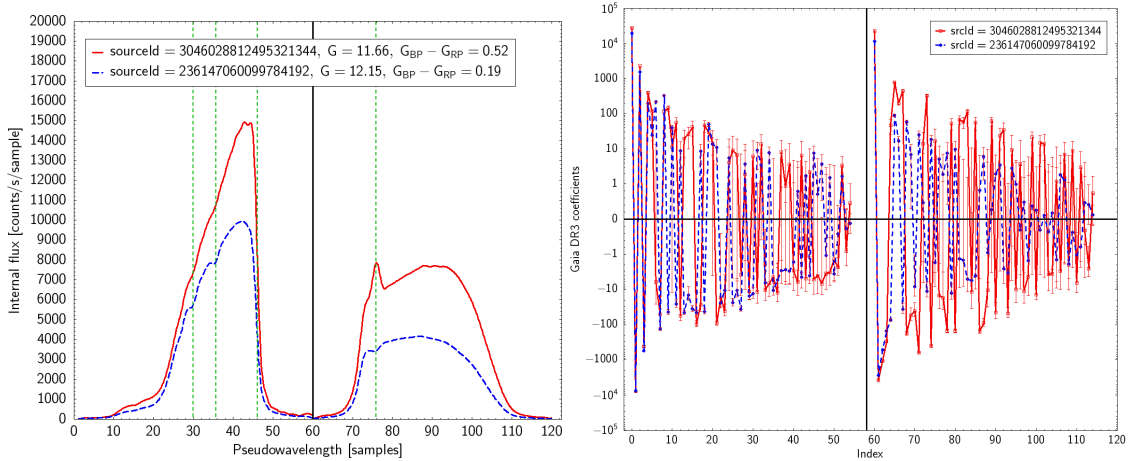
The XP spectra are described in *Gaia* DR3 as linear combinations of Hermite functions (De Angeli et al. 2023). From the same set of coefficients, the spectra can be reconstructed either in an internally calibrated form (still convolved with the total instrument response) or as externally calibrated spectra (in physical flux and wavelength units). The internal spectra express the

flux in photo-electrons per second, per unit pseudo-wavelength (a pixel-based coordinate), within the *Gaia* aperture, whereas the external spectra combine BP and RP into a single spectrum in energy units per unit time, wavelength, and detector area. The construction of the externally calibrated spectra involves deconvolving the low-resolution spectra with the line spread function (LSF) (Montegriffo et al. 2023).

Since the internally calibrated spectra contain complete spectral line information, we used them and avoided the external calibration step, which introduces additional sources of uncertainty. We exploited the specific representation of the internally calibrated XP spectra in *Gaia* DR3 as linear combinations of Hermite functions, following the formalism developed by Weiler et al. (2023). We briefly summarise this formalism in Sect. 2. Using this formalism we developed a web tool that is also available for use by the reader (XP-TEAL, Appendix A).

In this work, we use the *Gaia* low-resolution XP spectra to study the Balmer hydrogen lines of stars belonging to open clusters (OCs) in *Gaia* DR3. We derive equivalent widths (EWs) for three Balmer lines,  $H\alpha$ ,  $H\beta$ , and  $H\gamma$  in OC stars (Sect. 3). In addition, we analyse the sources with detected emission lines in OCs (Sect. 4). We estimate colour excesses for OCs by deriving an empirical dependence of absorption-line EWs on the stars'

\* Corresponding author: carrasco@fqa.ub.edu



**Fig. 1.** Example *Gaia* DR3 internal XP spectra (left) and their coefficients (right) for two sources: one with  $H\alpha$  in emission (solid red line) and one with Balmer absorption lines (dashed blue line). The RP pseudo-wavelengths (left panel) and index coefficients (right panel) are shifted by +60 so that BP and RP appear side by side. The solid vertical black line separates the BP (left) and RP (right) spectra. Left panel: pseudo-wavelengths represented in the abscissa increase from shorter (left) to longer (right) wavelengths; dashed vertical green lines mark the positions of  $H\alpha$  in RP and of  $H\alpha$ ,  $H\beta$  and  $H\gamma$  in BP (from right to left) using the nominal *Gaia* dispersion relation. Right: ordinate uses a symmetric logarithmic scale to better display the small high-order coefficients relative to the large low-order ones.

intrinsic colour (Sect. 5). Finally, in Sect. 6 we summarise our main results and conclusions.

## 2. Methodology

*Gaia* DR3 XP spectra are described by a set of coefficients and a set of Hermite basis functions (Carrasco et al. 2021; De Angeli et al. 2023). The same set of coefficients can be combined with different bases to sample either internal or external spectra; these are related through a transformation (shift and scale) to match the respective XP pseudo-wavelength grids (see Fig. 4 in Carrasco et al. 2021). In *Gaia* DR3, these Hermite bases are further orthogonally transformed to reduce the number of significant coefficients (De Angeli et al. 2023).

Hermite functions have a particularly convenient property: the derivative of any linear combination of Hermite functions is itself a linear combination of Hermite functions. Thus, given the coefficient vector of an internal spectrum, the coefficients of any derivative can be obtained by multiplying the original coefficient vector by an appropriate matrix. Moreover, the roots of a linear combination of Hermite functions are the eigenvalues of a ‘non-standard companion matrix’ associated with that combination. Standard eigenvalue techniques therefore allow us to efficiently compute all roots of a given combination. Together, these properties make it straightforward to locate local minima and maxima in the internally calibrated XP spectra. This framework was developed in detail by Weiler et al. (2023) and implemented in the algorithm XP tool for emission and absorption lines (XP-TEAL<sup>1</sup>; see Appendix A), which we employ here under the narrow-line approximation (see Sect. 8 in Weiler et al. 2023).

Spectral lines appear as local extrema in the internally calibrated spectra, or in their second derivatives when the lines are weak and the gradient of the underlying spectrum (continuum plus response curve) is large. Neglecting intrinsic line widths, the LSF (itself expressed as a linear combination of Hermite

functions) can be scaled and added or subtracted until a smooth interpolation is obtained. Smoothness is defined as the disappearance of either the second or the fourth derivative, depending on whether the line manifests directly in the spectrum or in its second derivative. This procedure yields the line EW from an internally calibrated spectrum with a modest computational cost.

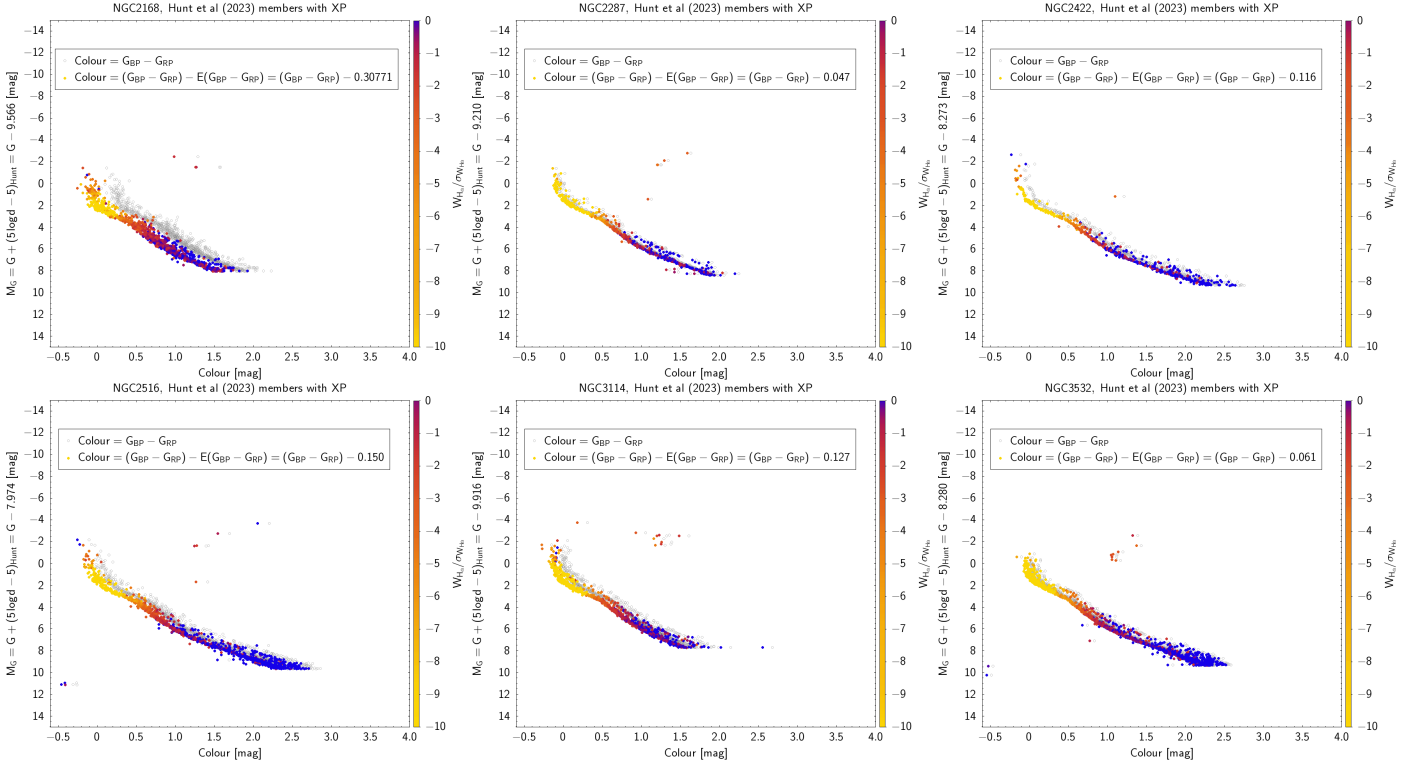
We adopted the formalism of Weiler et al. (2023) to measure Balmer-line strengths using the XP-TEAL narrow-line mode. We analysed three Balmer lines:  $H\alpha$  at 656.470 nm (better observed in RP than in BP, because of the higher resolution at this wavelength), and  $H\beta$  at 486.274 nm and  $H\gamma$  at 434.173 nm (both observed only in BP). We selected these lines because they have the highest signal-to-noise ratio (S/N) in *Gaia* XP spectra and are mostly free from blends (Creevey et al. 2023; Fouesneau et al. 2023). Line strengths are expressed as EWs. Following Weiler et al. (2023), we adopted the convention that negative EWs correspond to absorption and positive EWs to emission.

The Balmer lines of main-sequence stars are usually observed in absorption, with the line strength varying systematically with effective temperature (that is, spectral type). Hydrogen absorption peaks for A-type stars and decreases towards both earlier and later types (Maury & Pickering 1897; Cannon & Pickering 1901).

As illustrated for two example *Gaia* DR3 sources in Fig. 1, the internal XP spectra clearly carry information on the hydrogen Balmer lines. The pseudo-wavelength positions of  $H\alpha$ ,  $H\beta$ , and  $H\gamma$ , computed from the nominal dispersion law for *Gaia* DR3<sup>2</sup>, are indicated by green dashed vertical lines. The  $H\alpha$  line is located in the overlap region between BP and RP and is therefore accessible from both instruments. The  $H\beta$  and  $H\gamma$  lines are only present in BP. The corresponding XP coefficient vectors and their errors are also shown in Fig. 1. The low-order coefficients, which mainly encode the continuum, are larger and have smaller errors, while higher-order coefficients, which carry most of the line information, are smaller and comparable to their errors.

<sup>1</sup> XP-TEAL web tool is available at <https://xpteal.fqa.ub.edu>, and the Python source code at [https://github.com/malhotrasagar15/xpy\\_teal](https://github.com/malhotrasagar15/xpy_teal)

<sup>2</sup> Following Weiler et al. (2023), we also used an extra shift of +0.3 in pseudo-wavelength for  $H\alpha$  in RP in left panel of Fig. 1 in order to better match the spectral line.



**Fig. 2.** *Gaia* DR3 CAMDs for an illustrative set of young OCs from Hunt & Reffert (2023) with prominent blue main sequences. The vertical axis was scaled to absolute magnitudes using distance modulus values from Hunt & Reffert (2023). Grey points represent apparent  $G_{BP} - G_{RP}$  colours, while coloured symbols show intrinsic colours on the  $x$ -axis, corrected using the colour excesses derived in Sect. 5. The colour scale encodes the  $H\alpha$  EW obtained in this work.

In this work, we computed the EWs for the Balmer lines for the sample of *Gaia* objects with XP spectra in OCs (see Sect. 3). For each line, we also derived the ratio between the measured EW and its uncertainty, which we refer to as the S/N of the line.

### 3. Application to OC stars

We investigated the hydrogen Balmer lines in stars that belong to OCs. Open clusters (OCs) are particularly useful for our purposes because their member stars share a common origin (age and metallicity), a coherent space motion (which helps establish membership), and a similar distance, so that apparent magnitudes can be used to construct meaningful colour-magnitude diagrams (CMDs) without using individual parallaxes. Since all members lie approximately along the same line of sight, they are also typically affected by a similar amount of interstellar reddening.

Because of their relatively young ages (the age distribution peaks at  $\log \tau[\text{yr}] \sim 8.2$ ; e.g. Anders et al. 2021), some OCs still host blue main-sequence stars, which are ideal for tracing the evolution of hydrogen Balmer lines with colour or temperature in the *Gaia* XP spectra. Figure 2 presents several colour-absolute magnitude diagrams (CAMDs) for a selection of such young clusters for illustrative purposes.

Several studies (Cantat-Gaudin et al. 2018; Cantat-Gaudin & Anders 2020; Cantat-Gaudin et al. 2020; Castro-Ginard et al. 2020, 2022; Hunt & Reffert 2023, e.g.) provide cluster identifications and memberships from *Gaia*. In this paper we adopted the catalogue of Hunt & Reffert (2023) as our primary reference and used a total of 7135 cluster candidates (6785 of them classified as OCs; see Table B.1). Of the 1 291 929 stars listed by

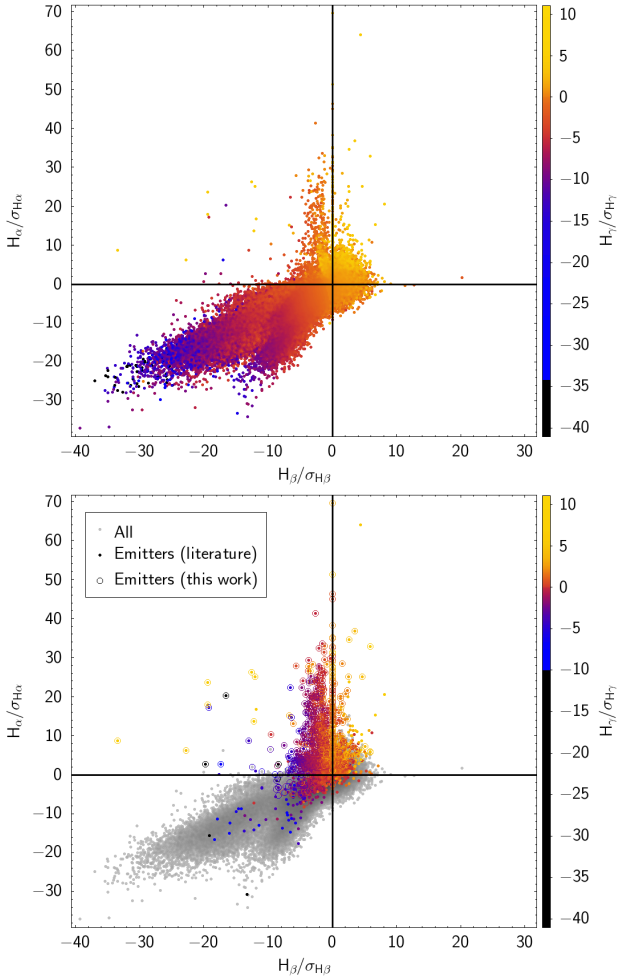
Hunt & Reffert (2023), 608 014 sources have XP spectra in *Gaia* DR3, with 989 of these stars appearing as members of more than one cluster in Hunt & Reffert (2023). The catalogue distinguishes OCs (kind = ‘o’) from globular clusters (kind = ‘g’). As mentioned above, we were interested in OCs and therefore excluded the globular clusters from our analysis. We also excluded 19 additional globular clusters from Vasiliev & Baumgardt (2021).

Figure 3 shows  $(S/N)_{H\alpha}$ ,  $(S/N)_{H\beta}$  and  $(S/N)_{H\gamma}$  for all selected OC stars with XP spectra. Both panels are colour-coded by  $(S/N)_{H\gamma}$ . The top panel highlights the (expected) correlation between the EWs of the three measured Balmer lines in absorption (with two distinct sub-branches in  $H\beta$  versus  $H\alpha$ ). The bottom panel highlights stars that show  $H\alpha$  in emission (see Sect. 4), showing that this emission is often (but not always) accompanied by  $H\beta$  and  $H\gamma$  in absorption. This is also expected, since many different stellar sub-classes can show  $H\alpha$  in emission.

### 4. Emission lines in OC stars

Among the 608 159 stars from Hunt & Reffert (2023) with XP spectra analysed here, *Gaia* DR3 (in particular, the ESP-ELS pipeline) flags 1400 sources as emission-line candidates. In addition, we find 2787 stars with emission-type classifications in SIMBAD astronomical database (Wenger et al. 2000), either in the ‘main\_type’, ‘other\_types’, or ‘sp\_type’ fields. Combining both lists yields 3602 stars previously known or suspected of hosting emission lines (see Figs. 3 lower panel and 4).

In this paper, we classify a star as an emission candidate if its  $H\alpha$  EW is greater than three standard errors above zero. Using this criterion we identify 1 256 emission candidates (1157



**Fig. 3.** Ratio of EW to its uncertainty (S/N) for  $H\alpha$  (ordinate),  $H\beta$  (abscissa), and  $H\gamma$  (colour scale) for all selected OC stars with XP spectra. Bottom panel: same but for known emission-line stars from the literature (solid symbols) and emission candidates identified in Sect. 4 (open symbols).

in OCs), which we list in Table B.2 and plot as coloured points in Fig. 4. Of these, 900 (845 in OCs) stars were already classified as emitters in the literature. The remaining 356 (312 in OCs) candidates are new emission-line candidates that, to our knowledge, had not been reported before. Another paper with an in-depth analysis of the emission-line stars in the full *Gaia* DR3 catalogue in preparation by Malhotra et al. (in prep).

Figure 4 shows how the  $H\alpha$  emission varies in different regions of the CAMD and as a function of the effective temperature and surface gravity, as reported by the *Gaia* DR3 GSP-Phot algorithm (Andrae et al. 2023), which derives the astrophysical parameters from XP spectra. Emission sources populate the upper main sequence, the pre-main sequence, and regions expected for evolved hot stars, in line with their spectroscopic classifications. The emission strength depends on both the temperature and the apparent magnitude in a non-trivial manner. Within a given OC, these quantities are correlated because cooler stars are typically fainter. In a cluster-by-cluster comparison, distance and extinction also play an important role: for more distant and/or reddened clusters, the S/N of the Balmer lines is generally lower, as expected.

The Extended Stellar Parametrizer for Emission-Line Stars (ESP-ELS; Creevey et al. 2023; Foesneau et al. 2023) provides independent  $H\alpha$  EW estimates for emission-line stars in *Gaia* DR3, referred to as ‘pseudo-equivalent widths’. Although we also used XP spectra, we used the spectral coefficients directly, whereas ESP-ELS uses the sampled spectra derived from them to perform its analysis. We compared ESP-ELS results with our XP-TEAL-based measurements in Fig. 5. We restricted the comparison in the left panel to sources with  $|(S/N)_{H\alpha}| > 1$  and a  $p$ -value  $p_{H\alpha} > 0.95^3$ , so that both methods operate in a reasonably high S/N regime. The agreement between the two sets of measurements is good, although our absolute EWs tend to be about a factor of two larger. We also note that the sign convention is the opposite: we adopted positive values for emission, whereas ESP-ELS uses negative values, which leads to a negative correlation in the linear fit in the left panel of Fig. 5. The green line in that panel indicates a slope of  $-2.3$ , which shows that the scale of EWs between this work and *Gaia* DR3 ESP-ELS differs by more than a factor of two. Other datasets also report differences of this order of magnitude with respect to the literature (Shridharan et al. 2022).

Figure 5 summarises the ESP-ELS classification of the emission candidates in OCs. Based on probabilistic classifications, ESP-ELS assigns 637 (587 in OCs) objects as Be stars, 131 (128 in OCs) as Herbig Ae/Be stars, 17 as T Tauri stars, 13 as nitrogen-sequence Wolf–Rayet stars, and four as carbon-sequence Wolf–Rayet stars (all T Tauri and Wolf–Rayet stars lie in OCs).

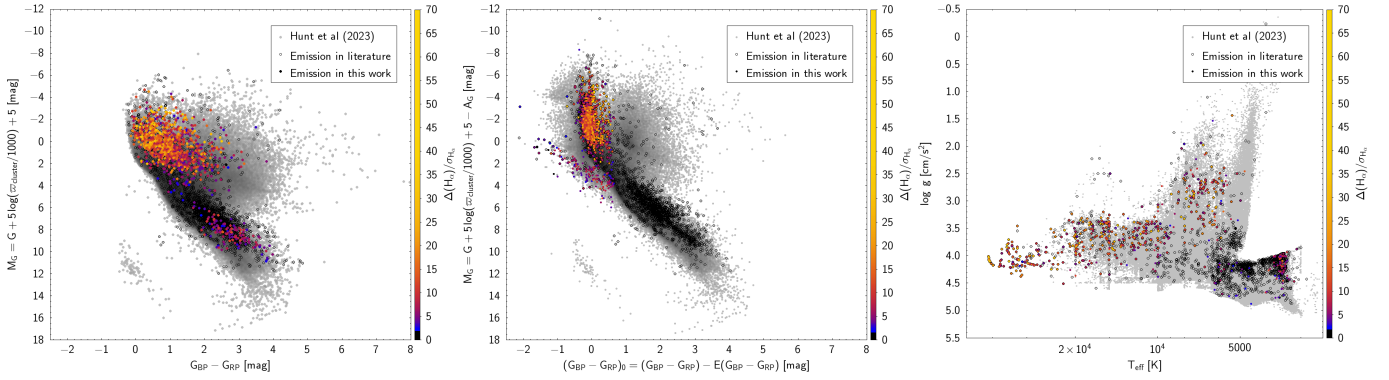
In total we find 3465 previously known emission stars in OCs when combining SIMBAD and ESP-ELS information. Our Balmer-based selection confirms 845 of these among our 1256 candidates. Our method does not flag the remaining 2620 OC stars as emission sources (see Figs. 3 and 4). For 1511 of these sources our algorithm assigns an EW equal to zero for at least one Balmer line, meaning that no significant local extremum is detected in the spectrum or its second derivative at the expected line position. For the others, the  $(S/N)_{H\alpha}$  simply does not exceed our conservative threshold. In practice, therefore, all such stars either have weak or marginal Balmer emission in the XP spectra or are located in regions where the noise and continuum gradients make a robust detection difficult. To illustrate this, in Fig. 6 we plot the internal *Gaia* BP and RP spectra for 45 emitters that our algorithm did not detect (providing EWs equal to zero for all three Balmer lines). At the Balmer-line positions (vertical dashed lines), the emission lines reported in the literature are not visible in these spectra.

## 5. Absorption line strength versus colour

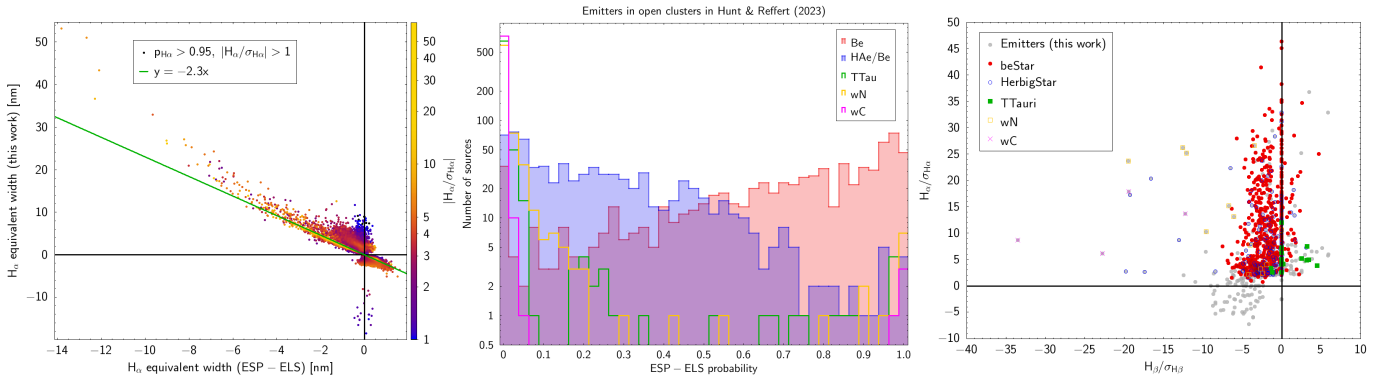
Equivalent widths (EWs) measure line strengths relative to the local continuum. Since hydrogen Balmer lines are narrow and the extinction curve is smooth, the EWs are essentially insensitive to interstellar extinction (see also Domínguez et al. 2013 for EW-based analyses of Balmer lines). As a result, the EW versus colour relations for different clusters share the same intrinsic shape, but shifted in colour by different amounts depending on the cluster reddening (Fig. 7). This behaviour allows us to derive cluster colour excesses and absorption.

Our methodology uses all stars in a cluster to obtain a global estimate of the cluster excess without considering inner-cluster variations in extinction. Although significant differences occur

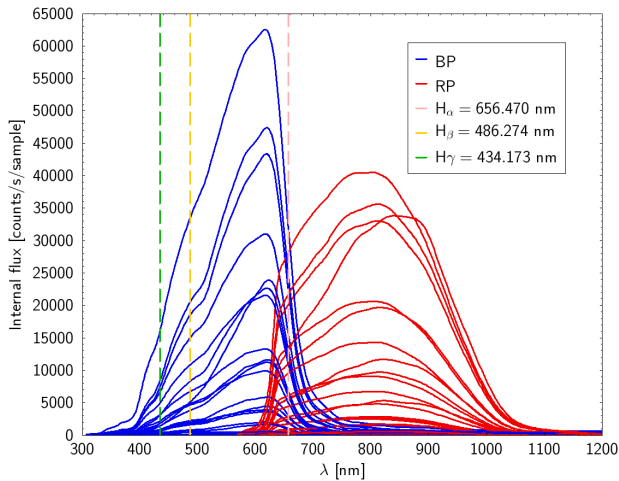
<sup>3</sup> The closer the  $p$ -value is to unity, the higher the significance of the local extremum (see Appendix A).



**Fig. 4.** Location of emission-line cluster members in the Hertzsprung–Russell-like diagrams. Open symbols represent previously known emission stars; filled symbols indicate new emission candidates from Sect. 4. Left: observed CAMD (using the parallax of the cluster from Hunt & Reffert 2023). Centre: de-reddened CAMD using  $A_G$  from BaSeL-3.1-based relations in Eq. (2). Right: GSP-Phot (Creevey et al. 2023; Andrae et al. 2023) parameters derived for the same sources.



**Fig. 5.** Comparison between  $H\alpha$  EWs measured in this work and the *Gaia* DR3 ESP-ELS results (Creevey et al. 2023). Left: direct EW comparison for OC members with  $|(S/N)_{H\alpha}| > 1$  and  $p_{H\alpha} > 0.95$ , colour-coded by  $|(S/N)_{H\alpha}|$ . The difference in sign convention (emission positive here and negative in Creevey et al. 2023) produces the anti-correlation. The green line indicates a slope of  $-2.3$ . Middle: probability distributions for different emission types (Be, Herbig Ae/Be, T Tauri, carbon and nitrogen Wolf-Rayet) from Creevey et al. (2023). Right:  $(S/N)_{H\alpha}$  versus  $(S/N)_{H\beta}$  coloured by ESP-ELS types; grey points without colour coding represent new emission candidates from this work.



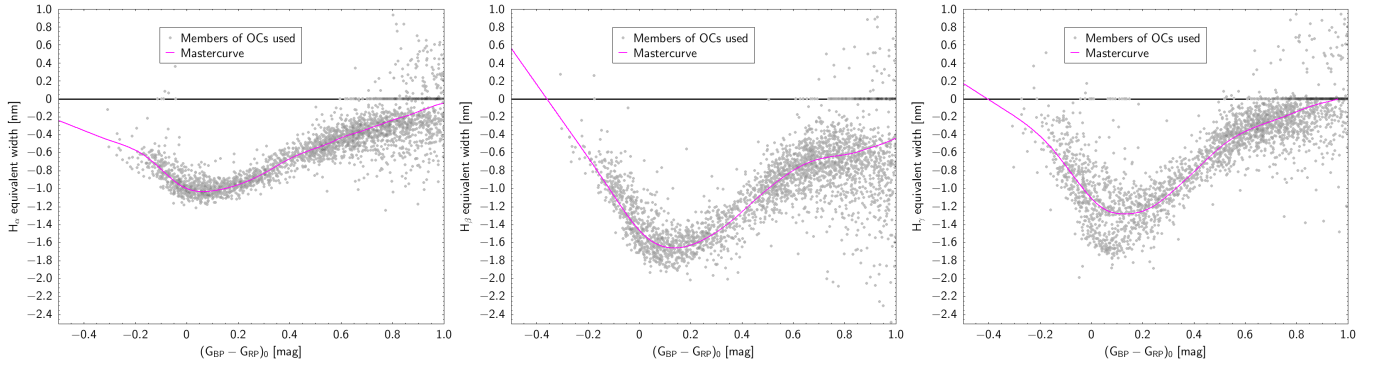
**Fig. 6.** *Gaia* DR3 internally calibrated XP spectra for 45 sources classified as emitters in the literature but not identified by our methodology using XP spectrophotometry (yielding EW equal to zero in all three Balmer lines). Vertical dashed lines indicate the three Balmer lines analysed in this work (from right to left:  $H\alpha$ ,  $H\beta$ , and  $H\gamma$ ).

within a cluster, they can easily be confused with other noise effects, and their detailed analysis is beyond the scope of this work.

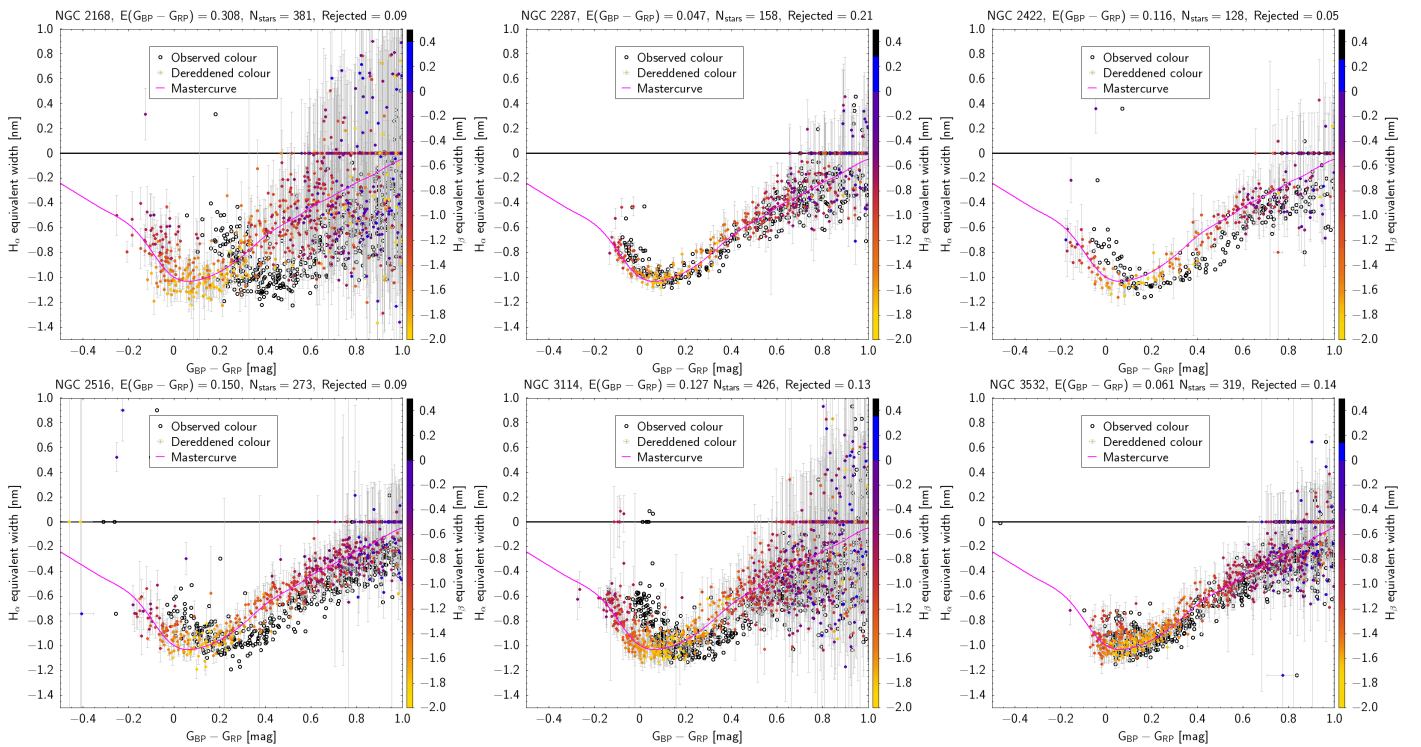
### 5.1. Empirical EW–colour relations (master curves)

We constructed empirical EW–colour relations (using  $G_{BP} - G_{RP}$  colour) for  $H\alpha$ ,  $H\beta$ , and  $H\gamma$  for main-sequence stars selected from the CMDs of a set of reference OCs. The chosen OCs fulfil three criteria: (i) they have a large number of members; (ii) their Balmer lines have good S/N; and (iii) they are young enough to include hot blue stars. The nine clusters that best satisfy these criteria are NGC 3532, NGC 3114, NGC 2287, NGC 6475, NGC 2451A, NGC 2422, BH 99, Trumpler 10, and Pozzo 1 (with 150–1541 members, mean S/N between 2 and 5, and minimum  $G_{BP} - G_{RP}$  between  $-0.5$  and  $-0.1$  mag). For each OC we shifted the  $G_{BP} - G_{RP}$  colours until the EW–colour relation matched that of the bluest reference cluster. We then fitted smoothing splines to the combined sample to obtain master curves for the three Balmer lines (Table B.3 and Fig. 7).

The master curves are reliable over  $-0.2 \lesssim (G_{BP} - G_{RP})_0 \lesssim 0.6$  mag. At bluer colours the number of stars is too small to constrain the relation, and at redder colours molecular bands complicate the interpretation of Balmer EWs. The curves reach minima (i.e. maximum absorption) at intrinsic colours of  $(G_{BP} - G_{RP})_0 = 0.06$  mag for  $H\alpha$ ,  $(G_{BP} - G_{RP})_0 = 0.136$  mag for  $H\beta$ , and  $(G_{BP} - G_{RP})_0 = 0.13$  mag for  $H\gamma$ , which correspond to effective temperatures of roughly 8500–9500 K for main-sequence stars with  $\log g \approx 4.5$ , i.e. spectral types close to A0V (Jordi et al. 2010).



**Fig. 7.** Equivalent widths (EWs) of  $H\alpha$ ,  $H\beta$ , and  $H\gamma$  versus intrinsic  $G_{BP} - G_{RP}$  colour for members of nine OCs with negligible extinction. Grey points represent individual stars, and magenta lines show the fitted EW-colour master curves used to derive colour excesses and identify outliers with anomalous Balmer strengths.



**Fig. 8.** Relation between  $H\alpha$  EW and  $G_{BP} - G_{RP}$  for the same OCs as in Fig. 2. Open black symbols show stars plotted with their observed colours; filled symbols show the same stars after shifting by the colour excess obtained in Sect. 5. Symbol colour encodes the  $H\beta$  EW. The title of each panel includes the values for the colour excess, the number of stars and the fraction (normalised to one) of outliers in the cluster.

Figure 8 shows how Balmer EWs vary with colour for the same sample of OCs, highlighting the maximum absorption at  $(G_{BP} - G_{RP})_0 \approx 0.1$  mag. We obtain the intrinsic colours used in this figure from the colour excesses derived below.

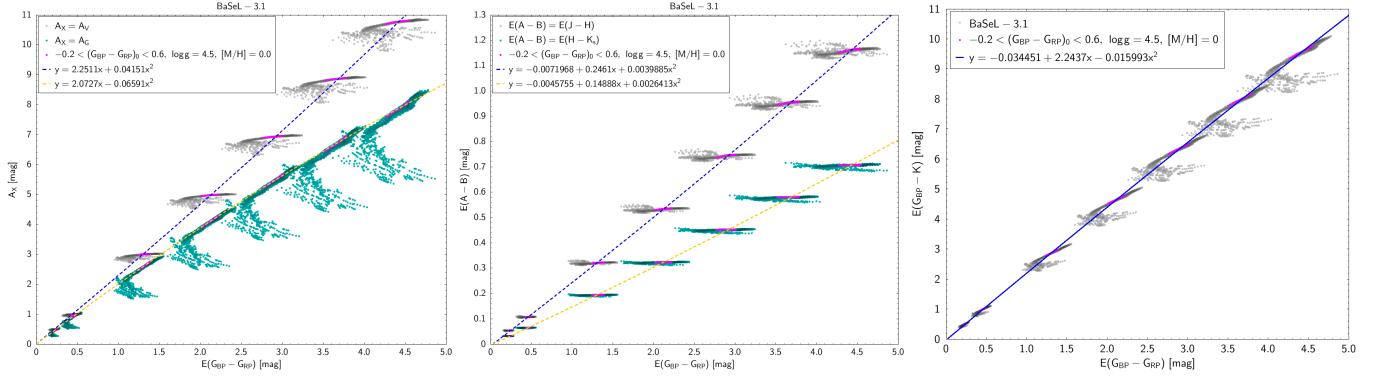
### 5.2. Deriving OC colour excesses

We next used the master curves to derive colour excesses for all OCs. The excess colour derived here is only valid for stars with strong hydrogen absorption, that is, for stars of spectral types B and A. As the *Gaia* passbands are very wide, the excess colour for other spectral types, in particular for much cooler stars, is significantly different from the values derived here, even under the same conditions for extinction (see Jordi et al. 2010).

For each OC we fitted the three master curves simultaneously to its EW-colour data by allowing a shift in  $G_{BP} - G_{RP}$ .

To mitigate the impact of outliers (notably emission-line stars; Sect. 4), we adopted an iterative procedure: we first determined the  $G_{BP} - G_{RP}$  shift that minimises the error-weighted  $\chi^2$ , then rejected data points deviating by more than three standard errors from the curves, and then repeated the procedure. The fits typically converge within a few iterations. We restricted the sample to stars that were bluer than the master-curve minimum minus 0.5 mag in  $G_{BP} - G_{RP}$  to ensure that we worked in the regime where Balmer absorption was strong and the relations were well behaved. We took the final  $G_{BP} - G_{RP}$  shift as the cluster colour excess  $E(G_{BP} - G_{RP})$ . We estimated the uncertainty from the  $\chi^2$  curve as the half-width of the interval in colour shift for which  $\chi^2$  increases by 0.5 from its minimum (Press et al. 2007).

The master-curve for the  $H\alpha$  line is plotted as a magenta line in Fig. 8 (which is the same for all OCs). We included the values of the colour excess determined for the clusters in



**Fig. 9.** Simulations of absorption and colour excesses using the BaSel-3.1 synthetic spectral library (Westera et al. 2002). Left: relationship between absorption in the V (grey) and G (cyan) passbands. For the polynomial fitting in Eqs. (1) (blue line) and (2) (orange line), we used only the magenta points (sources with reliable  $E(G_{BP} - G_{RP})$  estimations). Centre and right: Near-infrared 2MASS colour excesses as functions of  $E(G_{BP} - G_{RP})$ . Centre: 2MASS  $E(J - H)$  (grey) and  $E(H - K_s)$  (cyan) and lines show the polynomial fits in Eqs. (3) and (4). Right: excess in the combined optical-infrared colour  $E(G_{BP} - K_s)$ . The blue line shows the polynomial fit given in Eq. (5).

Hunt & Reffert (2023) in the online version of Table B.1. The grey data points in the CMD shown in Fig. 2 represent the observed data and the coloured symbols show the same data shifted in colour using the colour excess derived here.

The procedure described here provides negative excesses for some clusters, when their curves are located at lower colours than the reference master-curve or when the curve is not easily identified based on the emission lines behaviour. This occurs for 73 of the 7135 clusters in the sample. These clusters also show negative absorption values when derived from the colour excess in Sect. 5.3.

### 5.3. Absorption and excess estimations in other passbands

We verified the  $E(G_{BP} - G_{RP})$  values derived in Sect. 5.2 by comparing them with other estimates of interstellar extinction from other studies (see Sect. 5.4). To enable this comparison, we derived the corresponding absorptions in the *Gaia* G and Johnson V bands,  $A_G$  and  $A_V$ , using BaSel-3.1 synthetic spectra (Westera et al. 2002), following the methodology of Jordi et al. (2010). We used the *Gaia* DR3 *Gaia* passbands (Riello et al. 2021), the Johnson V response from Bessell & Murphy (2012), and the extinction law of Cardelli et al. (1989). We simulated stars with  $0 < A_{550} < 11$  mag, where  $A_{550}$  is the monochromatic extinction magnitude at wavelength  $\lambda = 550$  nm.

We obtained empirical polynomial relations as a function of the colour excess  $E(G_{BP} - G_{RP})$ . We plot the derived polynomials in Fig. 9 (left) and list them in Eqs. (1) and (2) as follows:

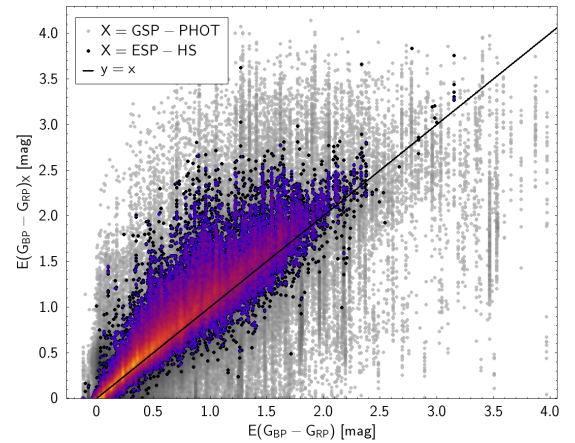
$$A_V = 2.2511E(G_{BP} - G_{RP}) + 0.04151E(G_{BP} - G_{RP})^2 \quad (1)$$

$$A_G = 2.0727E(G_{BP} - G_{RP}) - 0.06591E(G_{BP} - G_{RP})^2. \quad (2)$$

We used the same simulations to relate our *Gaia*-based colour excesses to 2MASS colour excesses (Cutri et al. 2003). We adopted the following polynomial fits (Fig. 9, centre and right panels):

$$E(J - H) = -0.0071968 + 0.2461E(G_{BP} - G_{RP}) + 0.0039885E(G_{BP} - G_{RP})^2 \quad (3)$$

$$E(H - K_s) = -0.0045755 + 0.14888E(G_{BP} - G_{RP}) + 0.0026413E(G_{BP} - G_{RP})^2. \quad (4)$$



**Fig. 10.** Colour excesses from the *Gaia* GSP-Phot (Andrae et al. 2023) and ESP-HS (Creevey et al. 2023) pipelines (ordinate) versus the cluster  $E(G_{BP} - G_{RP})$  derived in this work (abscissa) for OCs. Only the ordinate changes between the two panels; the abscissa is always the Balmer-based  $E(G_{BP} - G_{RP})$  derived in Sect. 5. The black line shows the identity relation.

To quantify the largest optical-infrared colour baseline we also derive a relation for  $E(G_{BP} - K_s)$  (Fig. 9, right panel),

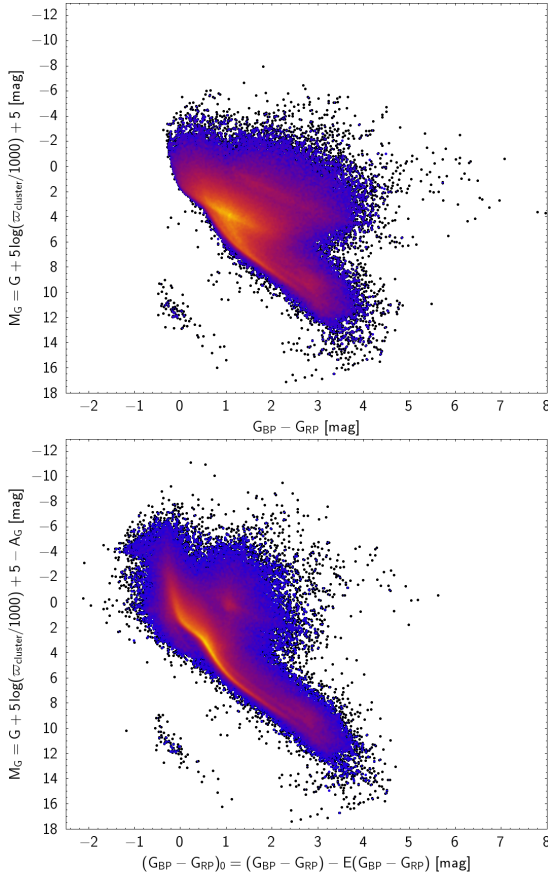
$$E(G_{BP} - K_s) = -0.034451 + 2.2437E(G_{BP} - G_{RP}) - 0.015993E(G_{BP} - G_{RP})^2. \quad (5)$$

Using these equations, we derived the values for  $A_V$ ,  $A_G$ , and the 2MASS excesses from our  $E(G_{BP} - G_{RP})$  values in Sect. 5. We include these estimates in Table B.1.

### 5.4. Comparison with other extinction estimates

This section compares colour excesses and absorption estimates derived in previous sections with estimates provided by other studies.

First, we compared our cluster-averaged  $E(G_{BP} - G_{RP})$  with the star-by-star colour excesses provided by the *Gaia* GSP-Phot and ESP-HS pipelines (Andrae et al. 2023; Creevey et al. 2023). Figure 10 shows that the correlation is tighter with ESP-HS (which also uses the *Gaia* radial velocity spectrometer data, RVS) than with GSP-Phot (using only XP spectra), suggesting that our Balmer-based colour excesses are of comparable quality



**Fig. 11.** Colour-absolute magnitude diagrams (CAMDs) for all selected OC stars using observed  $G_{BP} - G_{RP}$  colours (left) and intrinsic colours (right), obtained by subtracting the cluster  $E(G_{BP} - G_{RP})$  and  $A_G$  from BaSeL-3.1-based relations derived in Sect. 5.

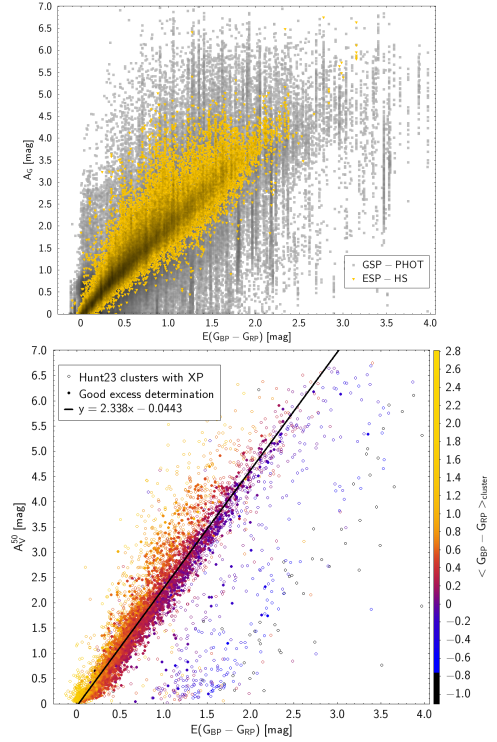
to those obtained from more complex models that also include RVS data.

Using *Gaia* colour excesses (Sect. 5) and absorption in the *Gaia*  $G$  passband (Eq. (2) in Sect. 5.3), we de-reddened the CMD. Figure 11 illustrates the impact of our de-reddening procedure on the global CMD built from all OC members. After correcting the colours using our  $E(G_{BP} - G_{RP})$  estimates (Table B.2) and applying the corresponding  $A_G$ , the main sequence becomes narrower, the giant and binary sequences are better defined, and the morphology of the diagram is significantly cleaner.

We expect that the colour excess is linearly related to the interstellar absorption (see Fig. 9 for the behaviour of the spectral library), with the slope given by the interstellar reddening<sup>4</sup>. This almost linear relationship is expected even when mixing passbands from different photometric systems<sup>5</sup>, but with a different slope (see left panel in Fig. 9). For example, in Fig. 12 (bottom) we compare our  $E(G_{BP} - G_{RP})$  values with the interstellar absorption in the Johnson-Cousins  $V$  band ( $A_V^{50}$ , the 50th percentile of  $V$ -band cluster extinction obtained using a Bayesian neural network) from Hunt & Reffert (2023) for the same cluster, showing that our estimated colour excesses are clearly correlated with  $A_V^{50}$ .

<sup>4</sup> Nevertheless, some variation with source colour is also present (see Fig. 9).

<sup>5</sup> The advantage of comparing colour excesses and absorptions from different photometric systems is to minimise the chain of transformations required to compare them, thereby reducing systematic effects.



**Fig. 12.** Interstellar absorption from different works plotted against the  $E(G_{BP} - G_{RP})$  colour excess derived in this work for OCs (see Sect. 5). Top: correlation between our colour excesses compared with the interstellar absorption derived from *Gaia* GSP-Phot Andrae et al. (2023) (grey squares) and ESP-HS (Creevey et al. 2023) (orange triangles). Bottom:  $A_V^{50}$  values in the ordinate axis from Hunt & Reffert (2023), with solid symbol marking sources with good excess determination. The colour index indicates the median intrinsic colour of each OC. The black line shows a linear fit. Both panels use the same axis scales.

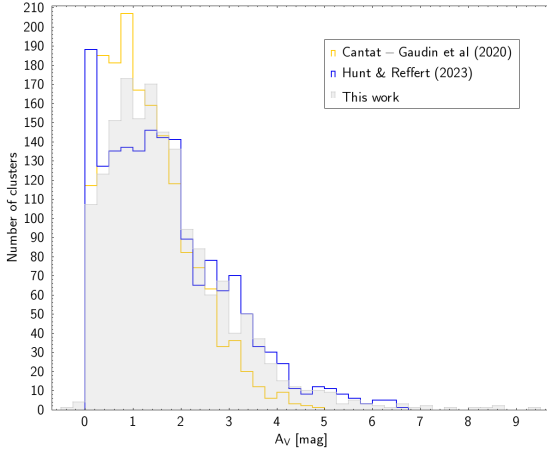
We find a weaker correlation (see Fig. 9, top) when using the individual  $A_G$  values derived by the *Gaia* collaboration (Andrae et al. 2023) for each member of the OC<sup>6</sup>.

The better agreement of our colour excesses with Hunt & Reffert (2023) than with Andrae et al. (2023) (see Fig. 12) is probably due to the fact that both Hunt & Reffert (2023) and our work use combined information from all sources in the OC, whereas Andrae et al. (2023) performs a star-by-star analysis without using additional information from other stars in the cluster. We obtain a better correlation (although still worse than for the Hunt & Reffert 2023 values) using the values from the *Gaia* Extended Stellar Parametrizer for Hot Stars (ESP-HS, Creevey et al. 2023, orange points in Fig. 12, top) which, in addition to XP information, uses data from the RVS.

Most of the stars in the top panel of Fig. 9 are located close to the linear fit, although there are some outliers with larger  $E(G_{BP} - G_{RP})$  values. This secondary sequence corresponds to inaccurate excess flux estimates caused by poorly shaped for the observed curves of hydrogen EWs as a function of colour or the low number of sources used for the fitting. To clean this diagram, we applied the following filters:

- To ensure agreement between the original fit of the master curve and the minimum chi-squared as a function of the shift in colour, we only used clusters with disagreement smaller

<sup>6</sup> We also used  $A_0$  GSP-SPEC and MSC values to check whether the correlation improved, but they showed similar results to those obtained with  $A_G$  shown in Fig. 12 (top).



**Fig. 13.** Histogram of  $A_V$  values included in [Cantat-Gaudin et al. \(2020\)](#) (orange), [Hunt & Reffert \(2023\)](#) (blue), and derived using Eq. (2) for clusters in common with  $A_G > 0$  mag from Eq. (2).

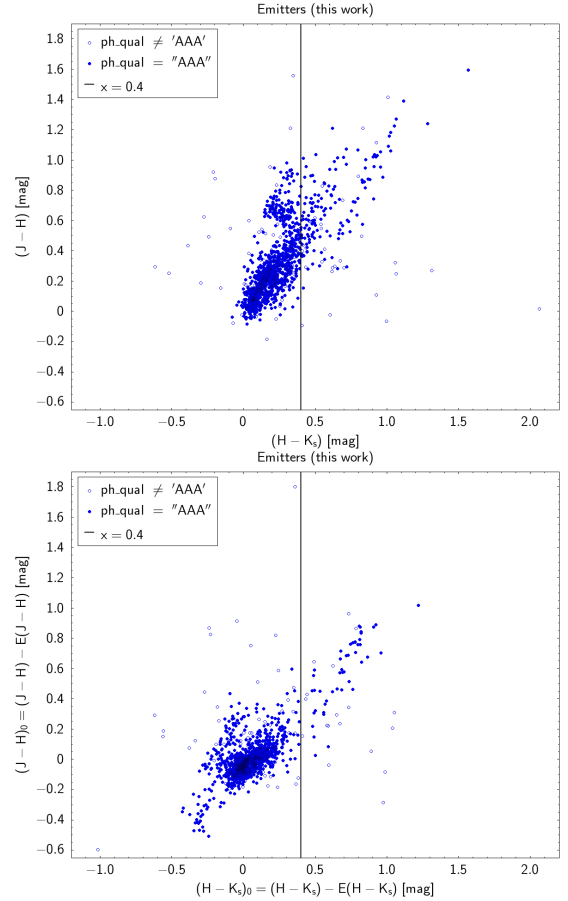
than  $|E(G_{BP} - G_{RP})_{\text{corr}} - E(G_{BP} - G_{RP})| < 2 \cdot 10^{-5}$  mag. The small threshold value reflects the fact that the two fitting methods used to derive the colour excess provide essentially the same result.

- We also ensured that a significant number of stars in each cluster was available for the fit by imposing  $N_{\text{stars}} > 10$ .
- We excluded clusters with too many outliers so that the fit remained meaningful. We enforced this by selecting only clusters with fewer than 15% outliers (rejectedFraction < 0.15).
- The last criterion ensured that the confidence interval was narrow ( $\Delta(E(G_{BP} - G_{RP})) < 0.1$  mag), i.e. the value of the colour excess was well constrained. We estimated the uncertainty of the colour excess by shifting the master curves over the observed  $EW-G_{BP} - G_{RP}$  relation on a fixed grid of shifts and computing the chi-squared value for each shift. We then used an interpolating spline to estimate the continuous chi-squared function with shift. We defined the upper and lower limit of the confidence interval as the shifts for which the chi-squared value increased with respect to the minimum value by a fixed amount. We adopted a value of 0.5, which corresponds to the  $1\sigma$  uncertainty interval for normally distributed errors.

This procedure yields  $E(G_{BP} - G_{RP})$  values for 6813 clusters of the 7135 with XP spectra analysed in Sect. 3. After applying additional quality cuts (e.g. on the fraction of rejected points and on the width of the confidence interval), we flag 2723 clusters as having reliable colour excesses (‘GoodExcess’ in Table B.1). Nevertheless, despite the filtering, a few OCs still lie in the secondary sequence. To provide the fitted line shown in Fig. 12 (bottom), we added an extra filter, using only sources with  $A_V > 2.4E(G_{BP} - G_{RP}) - 1$ . This mild filter leaves 2671 OCs, which we used to derive a global relation between  $E(G_{BP} - G_{RP})$  and the median  $V$ -band absorption from [Hunt & Reffert \(2023\)](#):

$$A_V^{50} = 2.338E(G_{BP} - G_{RP}) - 0.0443. \quad (6)$$

This relation captures the mean trend for all clusters, but the slope depends to some extent on the median intrinsic cluster colour (redder clusters exhibit larger slopes, bluer clusters smaller ones; see Fig. 12, bottom).



**Fig. 14.** Observed (top) and intrinsic (bottom) 2MASS colour-colour diagrams for the emission-line stars identified in this work. Filled symbols mark sources with 2MASS photometric quality ‘AAA’ in all bands. Intrinsic colours are derived using Eqs. (3) and (4).

Using the OCs in common between [Hunt & Reffert \(2023\)](#) and [Cantat-Gaudin et al. \(2020\)](#), we also compared their  $A_V$  values as functions of our  $E(G_{BP} - G_{RP})$ . When we adopt  $A_V$  from [Cantat-Gaudin et al. \(2020\)](#), we obtain a shallower relation:

$$A_V^{50} = 1.8684E(G_{BP} - G_{RP}) + 0.0409. \quad (7)$$

The differences between the absorption scales of [Hunt & Reffert \(2023\)](#) and [Cantat-Gaudin et al. \(2020\)](#) are discussed in detail by [Hunt & Reffert \(2023\)](#), who attribute them mainly to differential extinction of stars inside the clusters.

We find that the slope recovered from our simulations (2.2511, Eq. (1)) for the  $A_V = f(E(G_{BP} - G_{RP}))$  relation is closer to that obtained using the [Hunt & Reffert \(2023\)](#) data (2.338, Eq. (6)) than that derived from [Cantat-Gaudin et al. \(2020\)](#) (1.8684, Eq. (7)).

Figure 13 compares the distribution for the clusters in common between [Hunt & Reffert \(2023\)](#) and [Cantat-Gaudin et al. \(2020\)](#), considering only positive values for  $A_G$  as derived by Eq. (2). Our  $A_V$  values are more similar to those of [Hunt & Reffert \(2023\)](#), although we obtain more sources with  $A_V < 2$  mag.

We also examined the near-infrared colours of our emission candidates using 2MASS photometry ([Cutri et al. 2003](#)) and the *Gaia* DR3–2MASS crossmatch ([Marrese et al. 2019](#)). [Nidhi et al. \(2023\)](#), using 2MASS colours ([Skrutskie et al. 2006](#)), adopts the criterion  $(H - K_s)_0 > 0.4$  mag to select Herbig Ae/Be stars from classical Be and Ae stars. Of the 1256 emission candidates

studied in Sect. 4, 1228 have 2MASS counterparts, and 1108 of these have photometric quality flag ‘AAA’ (with  $S/N > 10$  in all three 2MASS passbands). Using Eqs. (3) and (4) to remove reddening effects, we constructed intrinsic 2MASS colour–colour diagrams (Fig. 14). We identify 70 stars (51 with ‘AAA’ quality) with intrinsic  $(H - K_s)_0 > 0.4$  mag, making them good candidates for Herbig Ae/Be stars.

## 6. Summary and conclusions

We analysed the *Gaia* DR3 internal XP spectra of stars in stellar clusters from Hunt & Reffert (2023) using the Hermite-function formalism of Weiler et al. (2023) to measure EWs for the Balmer lines  $H\alpha$ ,  $H\beta$ , and  $H\gamma$ . Open clusters (OCs), whose members share common ages, metallicities, distances, and reddening, provide an ideal test-bed for assessing the sensitivity and robustness of our method. The empirical Balmer EW–colour relations derived from a set of reference OCs define master curves that are nearly insensitive to extinction. We used these to obtain cluster colour excesses and to identify stars with anomalous Balmer strengths.

From these measurements we obtain  $E(G_{BP} - G_{RP})$  colour excesses for more than 6800 clusters, derive empirical relations linking  $E(G_{BP} - G_{RP})$  to  $A_G$ ,  $A_V$ , and 2MASS colour excesses, and demonstrate that the resulting de-reddened CMDs show significantly tightened main sequences and clearer cluster features. We also identify 1256 emission-line candidates (1157 in OCs), of which 356 (312 in OCs) appear to be new compared to ESP-ELS and SIMBAD. Finally, the XP-based colour excesses show good agreement with independent estimates from ESP-HS, confirming that Balmer-line strengths in *Gaia* XP spectra can provide reliable reddening indicators when applied to suitable samples of stars.

We have demonstrated that despite the very low resolution of the XP spectra, useful line information can be extracted, and the large number of sources can yield high quality scientific results.

Our next steps are to extend this approach to the full *Gaia* DR3 catalogue and to additional spectral features, such as the Ca II triplet in white dwarfs, helium lines in hot stars, and molecular bands in cool dwarfs and carbon stars.

## Data availability

Tables B.1–B.3 are available at the CDS via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/711/A215>.

**Acknowledgements.** This work was (partially) supported by the Spanish MICIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”

by the “European Union” through grant PID2021-122842OB-C21 and PID2024-157964OB-C21, and the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia ‘María de Maeztu’) through grant CEX2024-001451-M and the project 2021-SGR-00679 GRC de l’Agència de Gestió d’Ajuts Universitaris i de Recerca (Generalitat de Catalunya). This work was partially supported by the OCRE awarded project Galactic Research in Cloud Services (Galactic RainCloudS). OCRE receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 824079. This work has made use of data from the European Space Agency (ESA) mission *Gaia*<sup>7</sup>, processed by the *Gaia* Data Processing and Analysis Consortium (DPAC)<sup>8</sup>. Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This research has made use of the SIMBAD<sup>9</sup> database, operated at CDS, Strasbourg, France. This research has made use of the TOPCAT<sup>10</sup> and STILTS tools (Taylor 2005). We also thank R. Sordo for her insightful comments on the manuscript, which helped improve the paper.

## References

- Anders, F., Cantat-Gaudin, T., Quadrino-Lodoso, I., et al. 2021, *A&A*, 645, L2
- Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, *A&A*, 674, A27
- Bessell, M., & Murphy, S. 2012, *PASP*, 124, 140
- Cannon, A. J., & Pickering, E. C. 1901, *Ann. Harvard Coll. Observ.*, 28, 129
- Cantat-Gaudin, T., & Anders, F. 2020, *A&A*, 633, A99
- Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *A&A*, 618, A93
- Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, *A&A*, 640, A1
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, *A&A*, 652, A86
- Castro-Ginard, A., Jordi, C., Luri, X., et al. 2020, *A&A*, 635, A45
- Castro-Ginard, A., Jordi, C., Luri, X., et al. 2022, *A&A*, 661, A118
- Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, *A&A*, 674, A26
- Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, *VizieR Online Data Catalog*, II/246
- De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, *A&A*, 674, A2
- Domínguez, A., Siana, B., Henry, A. L., et al. 2013, *ApJ*, 763, 145
- Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, *A&A*, 674, A28
- Gaia* Collaboration (Prusti, T., et al.) 2016, *A&A*, 595, A1
- Gaia* Collaboration (Vallenari, A., et al.) 2023, *A&A*, 674, A1
- Hunt, E. L., & Reffert, S. 2023, *A&A*, 673, A114
- Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, *A&A*, 523, A48
- Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, *A&A*, 621, A144
- Maury, A. C., & Pickering, E. C. 1897, *Ann. Harvard Coll. Observ.*, 28, 1
- Montegriffo, P., De Angeli, F., Andrae, R., et al. 2023, *A&A*, 674, A3
- Nidhi, S., Mathew, B., Shridharan, B., et al. 2023, *MNRAS*, 524, 5166
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, *Numerical Recipes*, 3rd edn. (Cambridge University Press)
- Riello, M., De Angeli, F., Evans, D. W., et al. 2021, *A&A*, 649, A3
- Shridharan, B., Mathew, B., Bhattacharyya, S., et al. 2022, *A&A*, 668, A156
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Taylor, M. B. 2005, in *Astronomical Society of the Pacific Conference Series*, 347, *Astronomical Data Analysis Software and Systems XIV*, eds. P. Shopbell, M. Britton, & R. Ebert, 29
- Vasiliev, E., & Baumgardt, H. 2021, *MNRAS*, 505, 5978
- Weiler, M., Carrasco, J. M., Fabricius, C., & Jordi, C. 2023, *A&A*, 671, A52
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, 143, 9
- Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, *A&A*, 381, 524

<sup>7</sup> <https://www.cosmos.esa.int/gaia>

<sup>8</sup> <https://www.cosmos.esa.int/web/gaia/dpac/consortium>

<sup>9</sup> <http://simbad.cds.unistra.fr/simbad/0>

<sup>10</sup> <http://www.starlink.ac.uk/topcat/>

## Appendix A: XP-TEAL

XP-TEAL can be used through a web interface<sup>11</sup> (see screenshot in Fig. A.1). The underlying Python source code is also publicly available<sup>12</sup>. It implements the formalism of Weiler et al. (2023) for the analysis of emission and absorption lines in internally calibrated *Gaia* DR3 XP spectra, exploiting the Hermite-function representation to locate spectral features efficiently and to avoid deconvolution artefacts.

As explained in Sect. 2, the method searches for local extrema in the internally calibrated spectra or in their second derivatives, and provides several options for computing EWs. When no significant extremum compatible with a spectral line is found, XP-TEAL returns upper or lower limits for the EW. Users should, however, carefully inspect the results: not every extremum is a genuine spectral line, even if its position in wavelength matches with the expected position of a line. Their interpretation must always take into account the expected spectral properties of the targets and the noise characteristics of the data. The web interface of the tool is directly using *Gaia* DR3 XP spectra from the *Gaia* archive<sup>13</sup>. Four different options for providing input XP spectra are offered by the "Input" field (Fig. A.1):

1. Choosing the option "Use source\_ID dataset" allows the user to upload an ascii file containing a list of *Gaia* DR3 source identification number (source\_id), in a single column. The XP spectra for these sources, if available, are downloaded from the *Gaia* archive and processed. This option is preferable if the user already has a list of favourite sources available.
2. Choosing the "Use XP coefficients" option allows the user to upload a file containing the XP continuous spectra as they can be downloaded via the datalink from the *Gaia* Archive, in csv format. This method is preferable if the user already has downloaded continuous XP spectra of the sources of interest from the *Gaia* Archive.
3. Choosing the "ADQL Query" option allows the user to enter an ADQL query that is passed on to the *Gaia* Archive, and the available XP spectra for the sources matching the query are downloaded and processed.
4. Choosing the "Basic Query" option opens a field in which simple conditions based on the fields in the DR3 *Gaia* source table can be combined to select sources. The available XP spectra for all matching sources are downloaded from the *Gaia* archive and processed.

If the user wants the computation of EWs, the corresponding spectral lines need to be specified in the "Options" field of the web interface, under "List of lines".

There are currently two ways available to specify for which spectral lines to compute EWs. If "Balmer lines" is selected, the user can specify the hydrogen Balmer lines of interest. If "File with list of wavelengths" is chosen, the user can upload an ascii file containing the vacuum wavelengths, expressed in nm, of the spectral lines he is interested in.

Please note that in spectra of such low resolution as the *Gaia* XP spectra, spectral lines can easily become blended. The consistent treatment of blended lines is not yet available. All lines are treated as being unblended, and the results for lines which are blended might be inaccurate.

For the computation of the EWs of lines different assumptions can be made. These assumptions can be chosen under "Fitting of the lines". If the option "Narrow line" is chosen, the line is treated as if it had a negligible intrinsic width as compared to the width of the XP LSF. This option is fastest in computation, the most robust approach and with the lowest random noise on the resulting EWs. Using this option for broad lines or absorption bands can however result in systematic errors, which depend on the kind of sources.

The computational stability can be improved by imposing a certain line profile. The user can choose a Gaussian profile, a Lorentzian profile, or a Voigt profile. For each of these line profiles, the user can optionally provide the Full Width Half Maximum (FWHM), in nm, of the profile. If the FWHM is not provided, the code determines this value internally, again based on the position of the inflexion points around a local extremum. As a rule, the more assumptions are made on the line profile or the intrinsic width of the line, the more stable the continuum interpolation becomes.

Finally, the user can choose to use truncation in the number of coefficients used. If "Use truncated coefficients" is chosen, only the first XP\_N\_RELEVANT\_BASES coefficients, as specified in the DR3 spectroscopic table, are used for the computation. This approach reduces noise on the spectra, but it might affect the representation of narrow spectral features such as spectral lines. It is therefore not recommended to use this option.

The web interface can provide different kinds of output. If lines for which to compute the EWs are specified, an output file with the results is provided. This file is a csv file that contains in its first column the *Gaia* DR3 source\_id. Then, for each line five columns are provided. The line is either named by its Balmer designation, or by this wavelength in nm. The output data provided by the code are:

W: EW of the line (nm), negative for absorption and positive for emission.

Werror: standard error of the EW (nm).

p: p-value quantifying the significance of the local extremum (between 0 and 1); larger values indicate a lower probability that the extremum is due to noise.

D: Line width relative to the LSF width. Values close to 1 are consistent with a narrow line. This quantity is affected by some systematic errors, mainly resulting from an imperfect LSF model, and some slag should be permitted. However, values clearly smaller than 1 (say, around 0.5) might be an indicator that the local extremum is actually not a line, but a noise pattern. Values clearly larger than 1 (say, close to 2) might indicate that either the narrow line approximation is poor, or that the pattern is not a line at all, independent of the p-value.

order: Indicator of the detection order: "0" for extrema found in the spectrum, "2" for extrema found in the second derivative, and "3" when no significant extremum is detected and only an upper limit is provided. In this last case, W is set to 0, there are no entries for p and D, and Werror contains the upper limit for W corresponding to a  $1\sigma$  uncertainty.

If the output option "All extrema in the spectra" is chosen, then two further output files are provided, containing all local extrema present in the spectrum and its second derivative, respectively. These are csv files, with a semicolon, ";", as the secondary separator. They contain, for BP and for RP:

source\_id: The *Gaia* DR3 source\_id

extrema\_position\_xp: The position of the local extrema, in pseudo-wavelength units.

<sup>11</sup> <https://xpteal.fqa.ub.edu>

<sup>12</sup> [https://github.com/malhotrasagar15/xpy\\_teal](https://github.com/malhotrasagar15/xpy_teal)

<sup>13</sup> <https://gea.esac.esa.int/archive/>

Fig. A.1: Screenshot of XP-TEAL web tool that could be used by the reader to run the same code used in this work.

`extrema_positionError_xp`: The errors of the position of the local extrema, in pseudo-wavelength units.

`extrema_kind_xp`: The kind of extrema, either "minimum" or "maximum".

`extrema_estimSignif_xp`: The  $p$ -values of the extrema.

`extrema_estimWidth_xp`: The estimated width of the extrema, measured as the distance between the inflexion points surrounding the extremum, in pseudo-wavelength units.

`extrema_estimWidthError_xp`: the standard error on `extrema_estimWidth_xp`, in pseudo-wavelength units.

These two files are rather large, and the interpretation may not be straightforward, so requesting them is recommended only if they are really needed by the user. If the output option "sampled spectra" is chosen, then furthermore a file containing plots of all spectra are provided.

## Appendix B: Online tables

This paper provides several online tables via CDS. Here we show brief extracts of their content.

Table B.1: Sample of the information in an online table with information of the clusters analysed in this work.

| Cluster         | RA      | DEC     | $E(G_{BP} - G_{RP})$ | $A_G$ | $A_V$ | $E(H - K)$ | $E(J - K)$ | $E(G_{BP} - K)$ | GoodExcess | OC    |
|-----------------|---------|---------|----------------------|-------|-------|------------|------------|-----------------|------------|-------|
| 1636-283        | 249.856 | -28.399 | 2.473                | 5.820 | 4.722 | 0.380      | 0.626      | 5.416           | false      | false |
| ADS_16795       | 352.592 | 58.553  | 0.081                | 0.183 | 0.168 | 0.008      | 0.013      | 0.147           | false      | true  |
| AH03_J0748+26.9 | 117.156 | -26.973 | 0.554                | 1.260 | 1.128 | 0.079      | 0.130      | 1.204           | true       | true  |
| .               |         |         |                      |       |       |            |            |                 |            |       |
| .               |         |         |                      |       |       |            |            |                 |            |       |
| .               |         |         |                      |       |       |            |            |                 |            |       |

**Notes.** The table contains some information directly extracted from [Hunt & Reffert \(2023\)](#) (right ascension, RA, declination, DEC and interstellar absorption in  $V$  band,  $A_V^{\text{Hunt}}$ ) and others derived in this paper (the median colour of the cluster,  $\langle G_{BP} - G_{RP} \rangle$ , the interstellar absorption in  $G$  (using Eq. 2) and  $V$  (using Eq. 1) bands,  $A_G^{\text{pred}}$ , using Eq. 2 and  $\langle G_{BP} - G_{RP} \rangle$  in this table, the colour excesses of the clusters derived using Eqs. 3–5, and two flags indicating if we considered the cluster as an OC and if the colour excess derived can be considered of good quality, according to the filters mentioned in Sect. 5).

Table B.2: Sample of the information contained in an online table with the information of the members of the clusters analysed in this work.

| srcId               | ... | $H\alpha$ | $\sigma_{H\alpha}$ | $p_{H\alpha}$ | $D_{H\alpha}$ | order $_{H\alpha}$ | RelEmis $_{H\alpha}$ | ... | OC | Emitter |
|---------------------|-----|-----------|--------------------|---------------|---------------|--------------------|----------------------|-----|----|---------|
| 5866370214722778880 | ... | -0.237    | 0.094              | 0.931         | 0.898         | 2                  | 0.140                | ... | T  | F       |
| 5866193811838054016 | ... | -0.2811   | 0.056              | 0.999         | 0.870         | 2                  | 0.246                | ... | T  | F       |
| 276804702298080768  | ... | -0.137    | 0.077              | 0.531         | 2.033         | 2                  | 0.278                | ... | T  | F       |
| .                   |     |           |                    |               |               |                    |                      |     |    |         |
| .                   |     |           |                    |               |               |                    |                      |     |    |         |
| .                   |     |           |                    |               |               |                    |                      |     |    |         |

**Notes.** The members included here were extracted from [Hunt & Reffert 2023](#). The table contains some information extracted from *Gaia* DR3 (source identification, position, magnitude and colour). The rest of the columns are derived in this work. For each source the width of the  $H\alpha$ ,  $H\beta$  and  $H\gamma$  is provided (together with some relevant parameters of this determination). Finally two boolean flags indicating if a given star was considered to be part of an OC and if it is considered as an emitter in the previous literature are provided.

Table B.3: Sample of the mastercurves of the de-reddened width of the Balmer EWs as a function of the  $G_{BP} - G_{RP}$  colour.

| $G_{BP} - G_{RP}$ | $H\alpha$ | $H\beta$ | $H\gamma$ |
|-------------------|-----------|----------|-----------|
| -0.5              | -0.240    | 0.562    | 0.171     |
| -0.499            | -0.241    | 0.558    | 0.170     |
| -0.498            | -0.242    | 0.554    | 0.168     |
| .                 |           |          |           |
| .                 |           |          |           |
| .                 |           |          |           |