A&A 674, A28 (2023)
https://doi.org/10.1051/0004-6361/202243919

Gaia Data Release 3

Astronomy & Astrophysics
Special issue

Gaia Data Release 3

Apsis. II. Stellar parameters[⋆]

M. Fouesneau[1,⋆⋆], Y. Frémat[2], R. Andrae[1], A. J. Korn[3], C. Soubiran[4], G. Kordopatis[5], A. Vallenari[6], U. Heiter[3], O. L. Creevey[5], L. M. Sarro[7], P. de Laverny[5], A. C. Lanzafame[8,9], A. Lobel[2], R. Sordo[6], J. Rybizki[1], I. Slezak[5], M. A. Álvarez[10], R. Drimmel[11], D. Garabato[10], L. Delchambre[12], C. A. L. Bailer-Jones[1], D. Hatzidimitriou[13,14], A. Lorca[15], Y. Le Fustec[16], F. Pailler[17], N. Mary[18], C. Robin[18], E. Utrilla[15], A. Abreu Aramburu[20], J. Bakker[19], I. Bellas-Velidis[14], A. Bijaoui[5], R. Blomme[2], J.-C. Bouret[21], N. Brouillet[4], E. Brugaletta[8], A. Burlacu[16], R. Carballo[22], L. Casamiquela[4,23], L. Chaoul[17], A. Chiavassa[5], G. Contursi[5], W. J. Cooper[24,11], C. Dafonte[10], C. Demouchy[25], T. E. Dharmawardena[1], P. García-Lario[19], M. García-Torres[26], A. Gomez[10], I. González-Santamaría[10], A. Jean-Antoine Piccolo[17], M. Kontizas[13], Y. Lebreton[27,28], E. L. Licata[11], H. E. P. Lindstrøm[11,29,30], E. Livanou[13], A. Magdaleno Romeo[16], M. Manteiga[31], F. Marocco[32], C. Martayan[33], D. J. Marshall[34], C. Nicolas[17], C. Ordenovic[5], P. A. Palicio[5], L. Pallas-Quintela[10], B. Pichon[5], E. Poggio[5,11], A. Recio-Blanco[5], F. Riclet[17], R. Santoveña[10], M. S. Schultheis[5], M. Segol[25], A. Silvelo[10], R. L. Smart[11], M. Süveges[35], F. Thévenin[5], G. Torralba Elipe[10], A. Ulla[36], E. van Dillen[25], H. Zhao[5], and J. Zorec[37]

(Affiliations can be found after the references)

Received 2 May 2022 / Accepted 9 June 2022

ABSTRACT

Context. The third *Gaia* data release (*Gaia* DR3) contains, beyond the astrometry and photometry, dispersed light for hundreds of millions of sources from the *Gaia* prism spectra (BP and RP) and the spectrograph (RVS). This data release opens a new window on the chemo-dynamical properties of stars in our Galaxy, essential knowledge for understanding the structure, formation, and evolution of the Milky Way.

Aims. To provide insight into the physical properties of Milky Way stars, we used these data to produce a uniformly derived all-sky catalogue of stellar astrophysical parameters: atmospheric properties ($T_{\rm eff}$, $\log g$, [M/H], [α/Fe], activity index, emission lines, and rotation), 13 chemical abundance estimates, evolution characteristics (radius, age, mass, and bolometric luminosity), distance, and dust extinction.

Methods. We developed the astrophysical parameter inference system (Apsis) pipeline to infer astrophysical parameters of *Gaia* objects by analysing their astrometry, photometry, BP/RP, and RVS spectra. We validate our results against those from other works in the literature, including benchmark stars, interferometry, and asteroseismology. Here we assess the stellar analysis performance from Apsis statistically.

Results. We describe the quantities we obtained, including the underlying assumptions and the limitations of our results. We provide guidance and identify regimes in which our parameters should and should not be used.

Conclusions. Despite some limitations, this is the most extensive catalogue of uniformly inferred stellar parameters to date. They comprise $T_{\rm eff}$, $\log g$, and [M/H] (470 million using BP/RP, 6 million using RVS), radius (470 million), mass (140 million), age (120 million), chemical abundances (5 million), diffuse interstellar band analysis (half a million), activity indices (2 million), Hα equivalent widths (200 million), and further classifications of spectral types (220 million) and emission-line stars (50 thousand). More precise and detailed astrophysical parameters based on epoch BP, RP, and RVS spectrophotometry are planned for the next *Gaia* data release.

Key words. stars: distances – stars: fundamental parameters – methods: statistical – Galaxy: stellar content – dust, extinction – catalogs

1. Introduction

Studying the present-day structure and substructures of the Milky Way is one of the most direct ways for understanding the true nature of the Galaxy formation mechanism and evolutionary history. *Gaia* is an ambitious space mission of the European Space Agency (ESA) to primarily provide a three-dimensional map of the Milky Way with an unprecedented volume and precision (Gaia Collaboration 2016). It represents a revolution in galactic archaeology and is a leap forward in revealing how galaxies take shape and investigating the exciting complexities of our own Galaxy. Although it observes only one percent of the stellar population of our own Galaxy, *Gaia* still characterises $\sim$1.8 billion stars across the Milky Way, measuring their positions, parallaxes, and proper motions. It provides us not only with their three-dimensional positions, but also with their two- or three-dimensional velocities through the proper motion for $\sim$1.4 billion stars and radial velocity measurements for $\sim$33 million bright stars (Gaia Collaboration 2021, *Gaia* eDR3).

Gaia DR3 (Gaia Collaboration 2023a) improves upon the previous releases by improving the quality of the previously released data and by providing entirely new data products: (i) dispersed light spectra from spectrophotometry (BP blue photometer [330–680] nm; RP: red photometer [640–1050] nm) for $\sim$100 million stars, in addition to their integrated photometry ($G_{\rm BP}$, $G_{\rm RP}$) and the white-light G band published in *Gaia* eDR3

[⋆] Our catalogue is available from the *Gaia* Archive and partner data centres `https://gea.esac.esa.int/archive/documentation/GDR3/`.

[⋆⋆] Corresponding author: M. Fouesneau, e-mail: `fouesneau@mpia.de`

Table 1. Apsis module acronyms.

Acronym	Description
Apsis	*Astrophysical Parameters Inference System*
DSC	Discrete Source Classifier
GSP	*Generalized Stellar Parametrizer*
GSP-Phot	– from Photometry (photometry & BP, RP spectra)
GSP-Spec	– from Spectroscopy (RVS spectra)
ESP	*Extended Stellar Parametrizer*
ESP-CS	– for Cool Stars
ESP-ELS	– for Emission Line Stars
ESP-HS	– for Hot Stars
ESP-UCD	– for Ultra Cool Dwarfs
FLAME	Final Luminosity Age Mass Estimator
OA	Outlier Analysis
MSC	Multiple Star Classifier
TGE	Total Galactic Extinction

Notes. Refer to Sect. 3 for module descriptions.

(De Angeli et al. 2023) and (ii) medium-resolution spectroscopy (RVS, radial velocity spectrometer [845–872] nm, $\lambda/\Delta\lambda \sim$ 11 500) for ~1 million stars (Seabroke et al., in prep.).

In the previously released data, Andrae et al. (2018) published the first set of stellar parameters from the analysis of the integrated photometry and parallaxes available in *Gaia* DR2 (Gaia Collaboration 2018a). In contrast, *Gaia* DR3 provides a complex set of astrophysical parameters (APs) obtained from the analysis of the *Gaia* astrometry measurements and the BP, RP, and RVS spectra. This wide variety of information enables us to conduct a hyper-dimensional analyses of the Milky Way populations that have never been possible before the *Gaia* era.

The present work is one of a series of three papers on the *Gaia* DR3 astrophysical parameters. Creevey et al. (2023) presented an overview of the astrophysical parameter inference system (Apsis) and its overall contributions to *Gaia* DR3. This paper focuses on the stellar content description and quality assessments. The non-stellar content is presented in Delchambre et al. (2023). For more technical details of the Apsis modules, we refer to the online documentation[1] (Gaia Collaboration 2022) and specific publications describing some of the modules (GSP-Phot in Andrae et al. 2023, GSP-Spec in Recio-Blanco et al. 2023, and ESP-CS in Lanzafame et al. 2023). We list the relevant module acronyms in Table 1.

We only processed stellar sources down to $G = 19$ mag for which *Gaia* provides a BP/RP or RVS spectrum, except for ultra-cool dwarfs (UCDs). We selectively processed 78 739 UCDs fainter than this limit (see Fig. 1). This limiting magnitude choice was driven primarily by the limited processing time of the BP/RP spectra. The astrophysical parameter dataset contains stellar spectroscopic and evolutionary parameters for 470 million sources. These comprise T_{eff}, $\log g$, and [M/H] (470 million using BP/RP, and 6 million using RVS), radius (470 million), mass (140 million), age (120 million), chemical abundances (up to 5 million), diffuse interstellar band analysis (0.5 million), activity indices (2 million), Hα equivalent widths (200 million), and further classifications of spectral types (220 million) and emission-line stars (50 thousand).

The work described here was carried out within the *Gaia* Data Processing and Analysis Consortium (DPAC) within Coordination Unit 8 (CU8; see Gaia Collaboration 2016 for an overview of the DPAC). We realise that one can create more precise and possibly more accurate estimates of the stellar parameters by cross-matching *Gaia* with other survey data, such as GALEX (Morrissey et al. 2007), Pan-STARRS (Chambers et al. 2016), or catWISE (Eisenhardt et al. 2020), and spectroscopic surveys such as LAMOST (Luo et al. 2019), GALAH (Buder et al. 2021), or APOGEE (Jönsson et al. 2020). For example, Fouesneau et al. (2022), Anders et al. (2023), and Huang et al. (2022) combined *Gaia* data with other photometry and spectroscopic surveys to derive APs for millions of stars[2]. However, the remit of the *Gaia*-DPAC is to process the *Gaia* data. Further exploitation, for instance, including data from other catalogues, is left to the community at large. These *Gaia*-only stellar parameters will assist the exploitation of *Gaia* DR3 and the validation of such extended analyses, however.

We continue this article in Sect. 2 with a brief overview of our assumptions and key processing aspects. In Sect. 3, we describe the *Gaia* DR3 AP content, the validation of our results, and their internal consistenc. We compare them against other published results (e.g. benchmark stars, interferometry, and asteroseismology). Finally, we highlight a few applications of our catalogue in Sect. 4 and its limitations in Sect. 5 before we summarise in Sect. 6.

2. Overview of stellar APs in GDR3

The goal of Apsis is to classify and estimate astrophysical parameters for the *Gaia* sources using (only) the *Gaia* data (Bailer-Jones et al. 2013; Creevey et al. 2023). In addition to assisting the exploitation of *Gaia* DR3, the DPAC data processing itself uses these APs internally, for example, to help extract template-based radial velocities from the RVS spectra, identify quasars that were used to fix the astrometric reference frame, or the optimisation of the BP/RP calibration.

We designed the Apsis software to provide estimates for a broad class of objects covering a significant fraction of the *Gaia* catalogue, rather than treating specific types of objects. Apsis consists of several modules with different functions and source selections. Creevey et al. (2023) presented the architecture and the modules of Apsis separately. We provide in Fig. 2 a schematic overview of the source selection per Apsis module in the Kiel diagram. Some modules do not appear in this diagram as they have a more complex role (e.g., emission lines or classification).

2.1. Source processing selection function

This section details the source selection and assumptions we applied during the processing of stellar objects. First, we processed only sources for which one of the BP, RP, or RVS spectra was available with at least ten focal plane transits (repeated observations). The sources are processed by the specific modules depending on (1) the availability of the necessary data (2) the signal-to-noise ratio (S/N) of the data, the brightness to first order; and (3) potentially the outputs from other modules.

GSP-Phot (Andrae et al. 2023) operates on all sources with BP/RP spectra down to $G = 19$ mag. As we expect that more

---

[1] *Gaia* DR3 online documentation: `https://gea.esac.esa.int/archive/documentation/GDR3/`

[2] Survey acronyms: GALEX: the Galaxy Evolution Explorer; Pan-STARRS: the Panoramic Survey Telescope and Rapid Response System; APOGEE: the Apache Point Observatory Galactic Evolution Experiment; catWISE: the catalogue from the Wide-field Infrared Survey Explorer; LAMOST: the Large Sky Area Multi-Object Fibre Spectroscopic Telescope; and GALAH: the Galactic archaeology with HERMES.

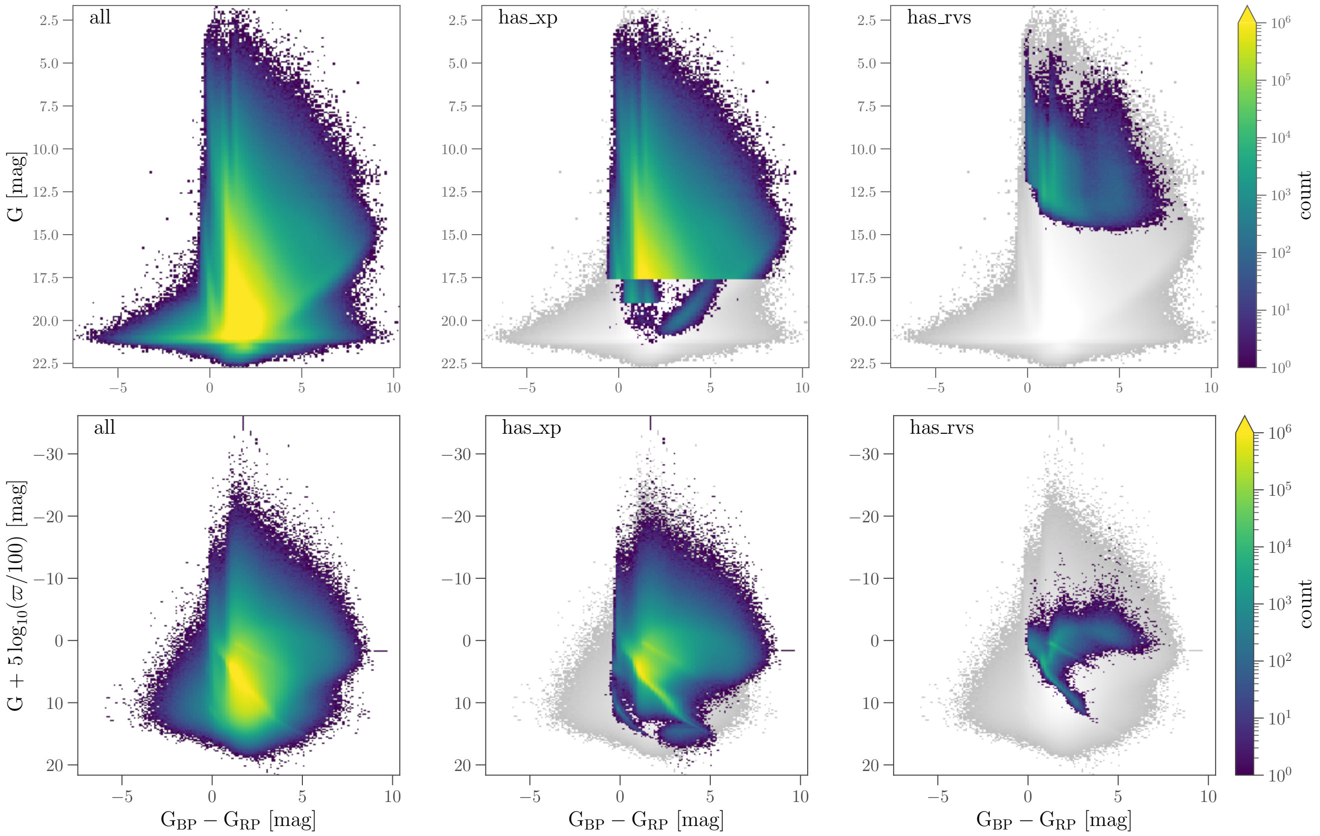

Fig. 1. Distribution of the sources in colour–magnitude space processed by Apsis according to the available measurements. The top panels show the observed colour-magnitude diagram. In contrast, the bottom panels show their absolute magnitude computed using the inverse parallax as the distance and assuming zero extinction for sources with positive parallax measurements. From left to right, the sources with G, BP, and RP photometry ("all"), those with published BP/RP spectra (`gaia_source.has_xp_continuous`), and those with RVS spectra (`gaia_source.has_rvs`). The gray density in the middle and right panels indicates the entire sample for reference. The peculiar distribution of BP/RP fainter than G = 17.65 mag in the top middle panel corresponds to selected UCDs (red sources) and extragalactic sources (blue sources). The inverse parallax used in the bottom panels includes low-quality parallaxes that cause the unphysically high brightness of many sources.

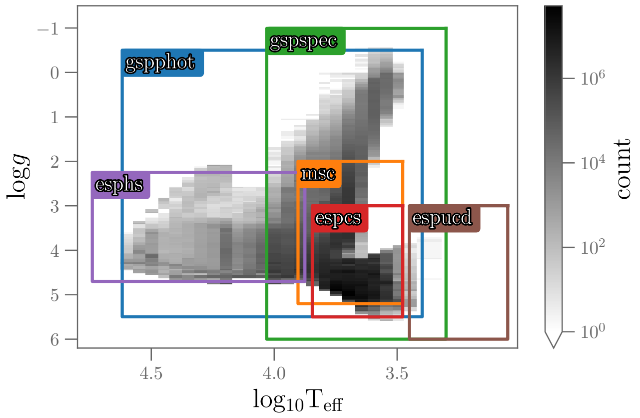

Fig. 2. Stellar parameter space spanned by Apsis modules in the Kiel diagram. Boxes indicate the modules producing estimates for either T_{eff}, $\log g$, or both. The density distribution represents the content from `gaiadr3.gaia_source`

than 99% of sources down to this brightness are stars, there is a minor overhead of computation time in applying GSP-Phot to every source and GSP-Spec (Recio-Blanco et al. 2023) on all sources with with a $S/N > 20$ in their RVS spectra, that is, $G \lesssim 13{-}14$ mag.

Following these two independent general analyses, Apsis refines the characterisation of *Gaia* sources with specific modules. FLAME operates on a subset of sources with APs of sufficient precision from GSP-Phot ($G < 18.25$ mag) and GSP-Spec ($G < 14$ mag), based on their reported uncertainties. MSC analyses all sources with $G < 18.25$ mag and treats every source as though it were a system of two unresolved stars. The remaining modules, specifically ESP-CS (Lanzafame et al. 2023), ESP-HS, ESP-ELS, and ESP-UCD only analyse objects of their class, that is, active cool stars, hot stars, emission-line stars, and ultra-cool dwarfs. Apart from ESP-UCD, which analyses UCDs fainter than $G = 19$ mag, the other specific modules only produce results for sources with $G < 17.65$ mag. Finally, GSP-Phot also provides the A_0 estimates used by TGE to produce an all-sky (two-dimensional) map of the total Galactic extinction, meaning the cumulative amount of extinction in front of objects beyond the edge of our Galaxy (see Sect. 3.4 and Delchambre et al. 2023). The various quoted magnitude limits are independent of the physical properties of the star and of the quality of the spectra. Instead, these limits came from the Apsis processing scheme and from processing time limitations[3]. In addition to and in contrast with the classifications from some of these analysis

---

[3] We used G to divide the entire *Gaia* data set into chunks of approximately 150 million sources. Some modules (e.g. GSP-Phot) ran faster than others (e.g. the ESP modules) and processed fainter chunks of data.

modules, Apsis comprises two modules dedicated to empirical classifications of sources. DSC classifies sources probabilistically into five classes: quasar, galaxy, star, white dwarf, and physical binary star. It is primarily intended to identify extragalactic sources, however, and OA complements this classification by clustering the sources with the lowest classification probabilities from DSC. See Sect. 3.6 and details in Creevey et al. (2023) and Delchambre et al. (2023).

We summarise the target selection of the Apsis modules in Fig. 3. We used the inverse parallax as a proxy to emphasise the stellar loci of the targets. Even though we did not explicitly select for $G_{\rm BP}$–$G_{\rm RP}$ colours, we note that most of the sources with $G_{\rm BP}$–$G_{\rm RP} < -0.8$ mag in *Gaia* DR3 are not stellar objects according to the Apsis processing definitions. This selection means that stellar evolution models (e.g. PARSEC[4]) do not predict bluer stars than $G_{\rm BP}$–$G_{\rm RP} < -0.6$ mag in the absence of noise in the measurements and within the chemical abundance regime of our analysis.

2.2. Stellar processing modules and stellar definition(s)

A principle of Apsis in *Gaia* DR3 is to use only *Gaia* data on individual sources when inferring the APs. We only used non-*Gaia* observations for validation and calibration. We defined stellar objects as those that remain after removing other types of objects: for instance, extragalactic sources (i.e. galaxies and quasars; Gaia Collaboration 2023b) through dedicated modules such as DSC and with proper motion, *Gaia* brightness, and colour selections. Apsis currently ignores morphological information (Ducourant et al. 2023) and does not take stellar variability (Rimoldini et al. 2023) into account. As it works with combined epoch spectra (BP, RP, and RVS), some time-variable sources (e.g. Cepheids) received spurious APs from Apsis. Eyer et al. (2023) summarises the characterisation of variable sources with dedicated pipelines. In the future, we plan to investigate using epoch data and determine whether variability information could improve the quality of our results.

A consequence of our analysis design is that Apsis can assign multiple sets of APs to any given source. Figure 2 illustrates the overlap between modules, which for example, leads to four temperature estimates for some main-sequence stars. The values we derive not only depend on the data we measure, but also on the stellar models we adopt (as embodied in the training data) and other assumptions made, see Creevey et al. (2023) for a brief overview and the online documentation for details. We can never know the true APs of a star with 100% confidence. Which estimate to use inevitably remains a decision for the user. For those users who do not wish to make this choice, GSP-Phot estimates APs for all the stars, so that a homogeneous set of stellar APs is always available.

The situation is even more complex in the details because a few of the modules themselves comprise multiple algorithms or multiple sets of assumptions, each providing separate estimates. One reason for this choice is that we cross-validate our results: if two or more algorithms give similar results for the same source (and training data), our confidence in the results may increase. For example, GSP-Spec provides estimates from Matisse-Gauguin (Recio-Blanco et al. 2016) and a neural-network approach (Manteiga et al. 2010) using the same RVS data. Another reason is that we do not use a common set of stellar models: GSP-Phot operates with four different atmospheric libraries with overlapping parameter spaces but significant differences (see Sect. 3.2.1).

Finally, while *Gaia* DR3 reports APs for a wide range of stellar types, we did not optimise Apsis to derive parameters for white dwarfs (WDs), horizontal branch (HB), and asymptotic giant branch (AGBs) stars. We did not attempt to model their specific physical conditions (e.g. compositional changes due to dredge-up, atomic diffusion, enriched atmosphere, and circumstellar dust).

2.3. Input data of Apsis processing

As Creevey et al. (2023) described the Apsis input data and their pre-processing exhaustively, here we briefly summarise the most relevant aspects of stellar APs. In the context of determining the stellar APs, we used sky positions, parallaxes, integrated photometry measurements, and BP/RP and RVS spectra. However, we note that the classifications by DSC also used proper motions.

Although Apsis mainly processed the sources independently (apart from TGE and OA), their positions on the sky were informative to determine their APs. For instance, we may see a source located near the Galactic center behind a significant amount of extinction, while it would be less likely towards high Galactic latitudes. Therefore, we defined sky-position-dependent priors using Rybizki et al. (2020) as a representative view of the *Gaia* sky, for instance. The details varied from one module to the next.

We implemented the parallax zero-points from Lindegren et al. (2021), which vary with magnitude, colour, ecliptic latitude, and astrometric solution type (`gaia_source.astrometric_params_solved`). A code is provided with *Gaia* DR3 to compute the parallax zero-points[5].

We used the integrated photometry in the G, $G_{\rm BP}$, and $G_{\rm RP}$ bands, in association with the zero-points provided by Riello et al. (2021). In addition, we also implemented the correction to the G-band photometry from Montegriffo et al. (2023), which depends on G, $G_{\rm BP}$–$G_{\rm RP}$ colour, and the astrometric solution type. We emphasise that the parallax zero-point remains calibrated on the original G-band photometry. However, *Gaia* DR3 publishes these corrected values in `gaia_source.phot_g_mean_mag`.

Apsis derived some of the APs from the analysis of the RVS spectra. The RVS processing pipeline provided us with time- or epoch-averaged spectra, also called mean spectra, after removing potential cosmic rays and the deblending of overlapping sources. The pipeline delivers the spectra in their stellar rest-frame, that is corrected for the radial velocity of the star (`gaia_source.radial_velocity`), and normalised at the local (pseudo-)continuum ($T_{\rm eff} \geq 3500$ K). Our analysis used these final spectra resampled from 846 to 870 nm, with a constant spacing of 0.01 nm. Seabroke et al. (in prep.) described the processing of RVS spectra in detail. However, Apsis modules rebin the spectra to their optimal use-cases to increase the signal-to-noise ratio of their relevant spectral features (Creevey et al. 2023, for details).

Most of the Apsis modules produced APs from the analysis of the BP and RP spectra (see examples in Fig. 4). *Gaia*

---

[4] PARSEC isochrones available from `http://stev.oapd.inaf.it/cgi-bin/cmd`

[5] *Gaia* eDR3 provides the parallax zero-point code `https://www.cosmos.esa.int/web/gaia/edr3-code`

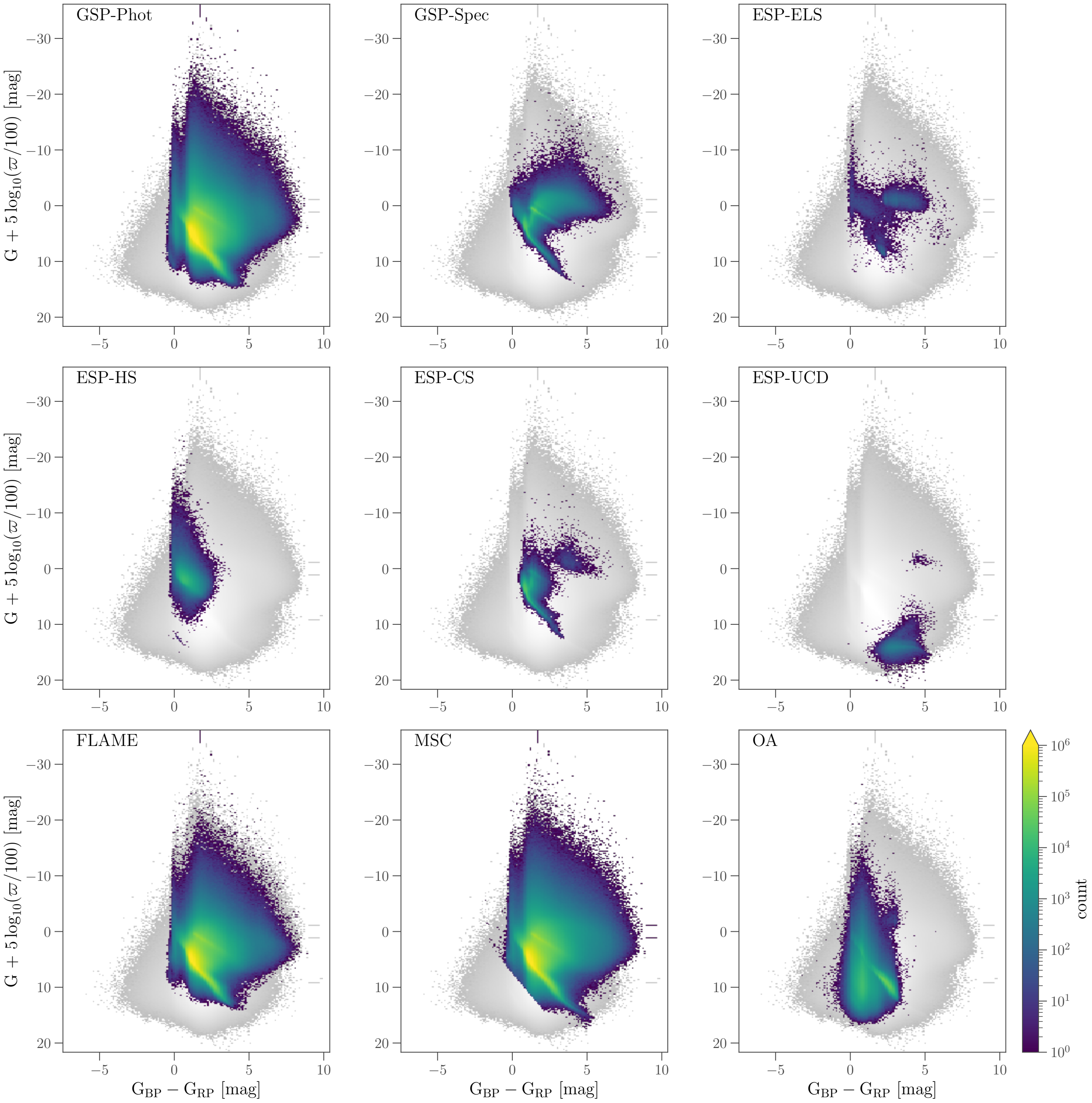

Fig. 3. Distributions of the sources processed by Apsis in *Gaia* DR3 in CMD space. Each panel highlights the sources a module processed, in addition to the sources with G, BP, and RP photometry (grey density).

DR3 provides the (epoch) mean BP and RP spectra in a series of coefficients associated with Gauss-Hermite polynomials. This format results from the complexity of the prism observations. Carrasco et al. (2021) described the processing of the spectra. These coefficients contain a flux-calibrated (mathematical) continuous representation of the spectra that the Apsis pipeline internally samples[6] approximately uniformly in pseudo-instrumental pixel space, but non-uniform in wavelengths (see Fig. 4 from Creevey et al. 2023).

2.4. *Typical examples and challenges of stellar BP/RP spectra*

The BP and RP spectra reside at the boundary between photometry and spectroscopy. Due to the low effective spectral resolution from the prisms, these data present only a few noticeable features, as opposed to individual spectral lines in spectroscopy. On the other hand, where spectroscopy often provides an uncertain determination of a stellar continuum, the BP and RP data provide robust determinations with high signal-to-noise ratios similar to photometric measurements. To illustrate this further, Fig. 4 shows the variation in spectra of dwarf stars with effective temperature. In this figure, we divided the spectrum fluxes by the instrument filter responses as provided by the simulation tool

---

[6] *Gaia* DR3 provides GaiaXPy, a Python package to sample the BP and RP spectra; `https://www.cosmos.esa.int/web/gaia/gaiaxpy`

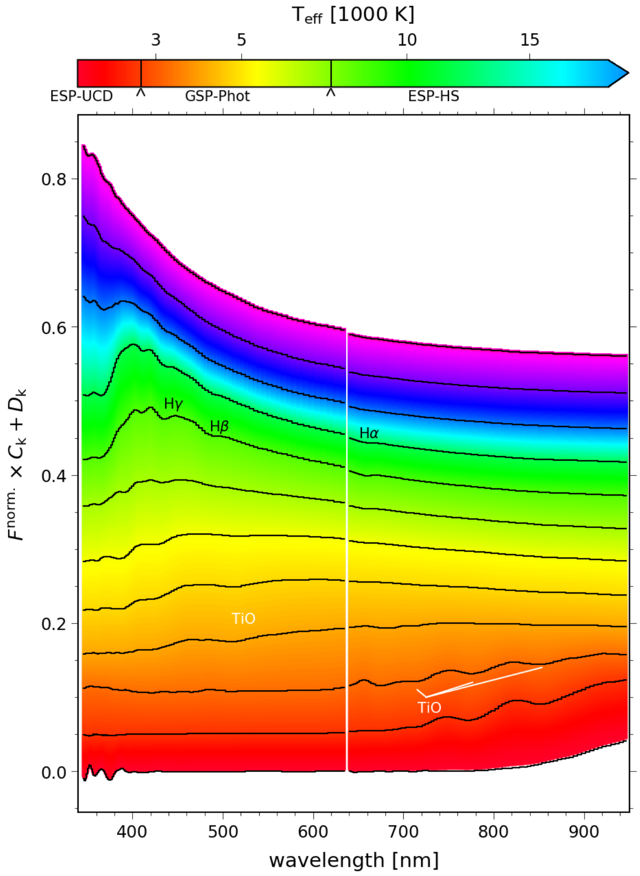

Fig. 4. Variations in BP and RP spectra of main-sequence stars with effective temperature. The background colour-coding follows the effective temperature scale provided by ESP-UCD, GSP-Phot, and ESP-HS (also indicating their optimal $T_{\rm eff}$ performance regimes from our validation). We highlighted some spectra for reference and labelled some spectral features. We normalised the spectra to their integrated flux after correcting the BP/RP for the instrument response (see Montegriffo et al. 2023). We further stretched and vertically shifted the resulting normalised flux ($F^{\rm norm.}$). We restricted our selection to comparable dwarfs: GSP-Phot stars with $4 \leq \log g, < 4.5$ dex, $A_0 < 0.2$ mag and $-0.1 \leq$ [Fe/H] $\leq +0.1$ dex, and ESP-HS stars with $4 \leq \log g, < 4.5$ dex, $A_0 < 0.2$ mag (see the discussion in Sect. 2.4).

internally available to DPAC (Montegriffo et al. 2023). GaiaXPy provides the community with a similar tool[7]. Ultra-cool stars mainly emit photons in the RP passband, and their spectra depict strong molecular features. The almost featureless A-, B-, and O-type stars exhibit the Balmer hydrogen lines and the Balmer jump. In between, we have the F-, G-, K-, M-type stars characterised by the appearance of TiO bands and metal line blends. Figure 4 from Creevey et al. (2023) compares the variation in BP and RP spectra with effective temperature and extinction using simulations and observational examples. Based on these data, we also classify emission-line stars (ELS) by their stellar class by measuring the Hα line strength and identifying significant emissions in other wavelength domains. In Fig. 5 we plot the spectral energy distribution (SED) of some of the stellar classes that the ESP-ELS module estimated. While one can usually find the strongest features in some planetary nebula and Wolf-Rayet stars, weaker Hα emission is more challenging to measure due to the low resolving power of BP and RP spectra[8]. The difficulty increases further for the cool ELS stars ($T_{\rm eff} \leq 5000$ K), whose spectra mainly show a weak Hα emission blended into the local pseudo-continuum shaped by the TiO molecular bands. Combining the BP and RP data with higher-resolution spectra (e.g., RVS, LAMOST, or APOGEE) will become an obvious path of choice for the next decades.

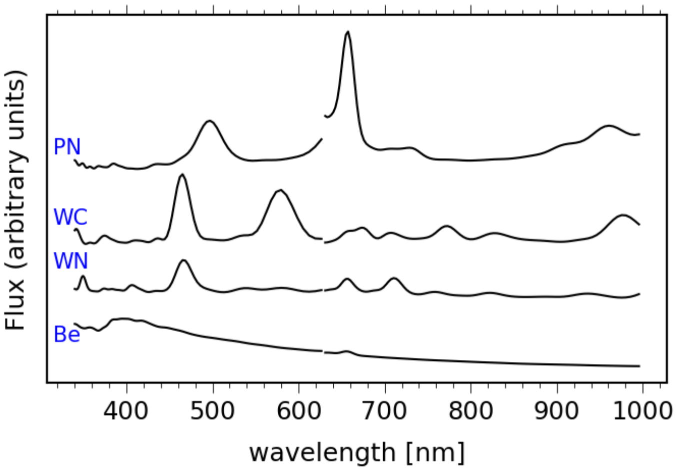

Fig. 5. Emission-line features in the BP/RP spectra of various types of objects: planetary nebula (PN), Wolf-Rayet Carbon-rich (WC) and nitrogen-rich (WN) stars, and Be stars. We divided the spectrum flux by the instrument filter responses. We added offsets between the SEDs to place them in the same figure (see discussion in Sect. 2.4).

2.5. Typical RVS spectra

The RVS spectra share a lot of similarities with those from RAVE. The RVS spectra have a slightly shorter wavelength window, but a higher resolution (~11 500): from 846 to 870 nm, with a resolution element of 0.001 nm.

Figure 6 presents a selection from *Gaia* DR3 of typical RVS spectra in the OBAFGKM sequence, a sequence from the hottest (O-type) to the coolest (M-type). Each letter class subdivides itself using numbers where 0 is the hottest and 9 is the coolest (e.g., A0, A4, A9, and F0 from hotter to cooler). We selected these spectra based on their spectroscopic temperatures and surface gravities.

The variations in RVS spectra with the effective temperature are strong, and the spectra of F-, G-, and K-type stars present many atomic lines, but a reliable measurements depends strongly on the temperature and gravity of the star. The *Gaia* Image of the Week 2021-07-09 presents an animation of several *Gaia* RVS stellar spectra and their element abundances. This figure also illustrates the challenge of characterising O-type stars that present nearly featureless RVS observations.

3. AP content description and performance

This section describes the AP content of *Gaia* DR3, their performance, and limitations. We first discuss the object APs individually: their distances in Sect. 3.1, their stellar atmospheric parameters in Sect. 3.2 (i.e., $T_{\rm eff}$, $\log g$, metallicity, individual abundances, rotation, and activity), and their evolution

---

[7] GaiaXPy: `https://www.cosmos.esa.int/web/gaia/gaiaxpy`

[8] The effective resolution of BP and RP spectra decreases towards the red wavelengths, and the RP response steeply drops at the blue edge at 640 nm.

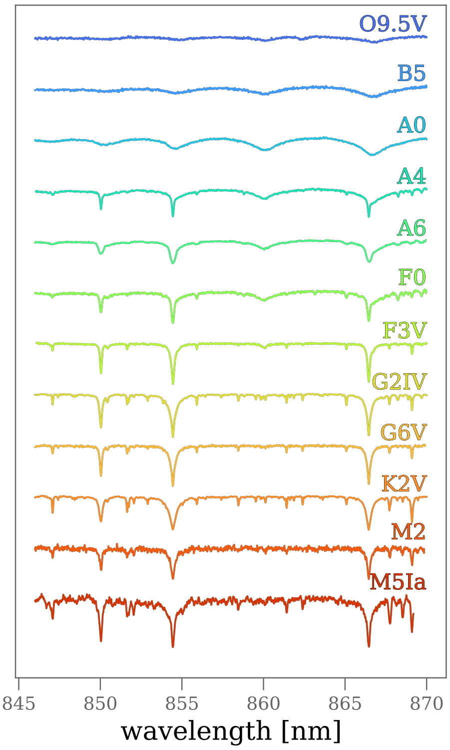

Fig. 6. Typical RVS spectra in the OBAFGKM sequence published with the *Gaia* DR3 release. The vertical axis is in arbitrary units for the comparison. The luminosity class V indicates main-sequence stars, and the M5Ia source is a supergiant.

parameters in Sect. 3.3 (i.e., absolute and bolometric luminosities, radius, gravitational redshift, mass, age and evolution stage). These require us to account for dust effects along the line of sight summarised in Sect. 3.4 and analysed in depth in Delchambre et al. (2023) and Gaia Collaboration (2023c). In Sect. 3.5, we further assess the quality of our APs by focusing on objects in groups (i.e., clusters and binaries). Finally, we discuss the detection of peculiar cases and outliers in Sect. 3.6.

We emphasise that to avoid repetitions, we summarise only the complete description of some internal precisions of the APs as a function of magnitude, colour, sky position, and other parameters that appear in other publications (e.g. Andrae et al. 2023; Creevey et al. 2023; Delchambre et al. 2023; Recio-Blanco et al. 2023; Lanzafame et al. 2023).

For guidance, Appendix D compiles the various estimates of stellar parameters from *Gaia* DR3 cast into the categories described above (corresponding to the following subsections). The compilation indicates which Apsis module produces them, and which table and fields store the values in the *Gaia* catalogue. We emphasise that the field names correspond to the catalogue in the *Gaia* Archive but names may differ when using partner data centers.

3.1. Distances

Two Apsis modules provide distance estimates: GSP-Phot for single stars and MSC for unresolved binary stars. Both modules analyse the BP and RP spectra with the *Gaia* parallaxes to derive distance estimates simultaneously with other astrophysical parameters. We list the catalogue fields related to the distance estimates of the two modules in Table D.1.

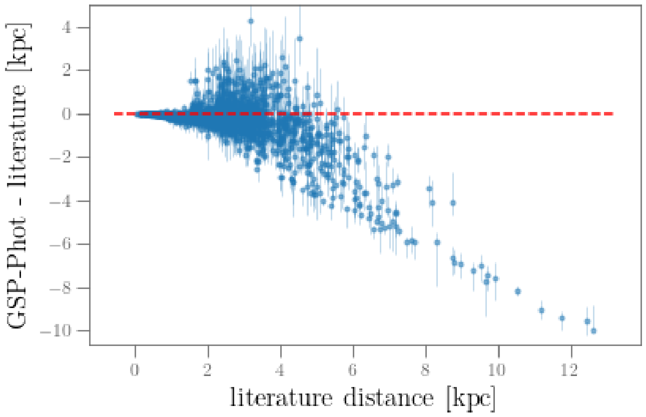

Fig. 7. Comparison of GSP-Phot distances in star clusters against literature values. For 2019 star clusters (nearly 200 000 stars) spanning distances up to 13 kpc, we indicate on the *y*-axis the median offset (dot), and the 50% quantile (line) of all member star distances with respect to cluster distances taken from Cantat-Gaudin et al. (2020). Above ~2.5 kpc (~15% of the cluster members), GSP-Phot systematically underestimates distances (see the discussion Sect. 3.1).

For GSP-Phot, the distances are reliable out to ~2 kpc. Beyond 2 kpc, GSP-Phot systematically underestimates distance, as is evident from star clusters, for example. Figure 7 compares the median GSP-Phot distances of stellar members for each cluster with their literature values by Cantat-Gaudin et al. (2020) derived using *Gaia* DR2 data through maximum likelihood. We included the *Gaia* DR3 variable zero-point on parallaxes mentioned in Sect. 2.3. We obtain similar results when comparing this to the photometric distances by Kharchenko et al. (2013) and in BOCCE (Bragaglia & Tosi 2006; Cantat-Gaudin et al. 2018) catalogues based on fitting a colour-magnitude diagram. However, when the parallax measurement is good (about $\varpi/\sigma_\varpi > 10$), the GSP-Phot distances remain reliable even out to 10 kpc, as we show in Fig. 8a. The reason for this systematic underestimation of distances by GSP-Phot is an overly harsh distance prior. Andrae et al. (2023) discussed the prior and showed that we could resolve this issue by updating its definition. A prior optimisation remains necessary and will be part of further releases. Figure 8 also compares the distances from Bailer-Jones et al. (2021) and Anders et al. (2023) to the *Gaia* DR3 parallaxes. We note that they perform better than GSP-Phot distances[9]. For this reason, various DR3 publications chose to not use the GSP-Phot distances but rather EDR3 distances from Bailer-Jones et al. (2021) (e.g., Gaia Collaboration 2023c,d,e). A further comparison of GSP-Phot distances with those from asteroseismic analyses confirmed a good agreement to 2 kpc, and some outliers beyond (see Fig. 9).

MSC provides distance estimates assuming sources are unresolved binaries with luminosity ratios ranging from 5 to 1. The MSC distance estimates would differ from GSP-Phot estimates (equivalent to an infinite luminosity ratio) by a factor 10 to 50% at best, respectively. We highlight that distances with luminosity ratios of 5 significantly differ from single-star assumptions. Figure 10 compares MSC distance estimates and those from GSP-Phot to the *Gaia* parallaxes for the spectroscopic binary samples from Pourbaix et al. (2004) (mostly $G < 10$ mag) and Traven et al. (2020) (mostly between $G = 10$ and 15 mag).

---

[9] GSP-Phot derived distances from parallaxes with the zero-point correction from Sect. 2.3, but Fig. 8 uses uncorrected values on the *x*-axis. The zero-point alone is unlikely to generate these differences, however.

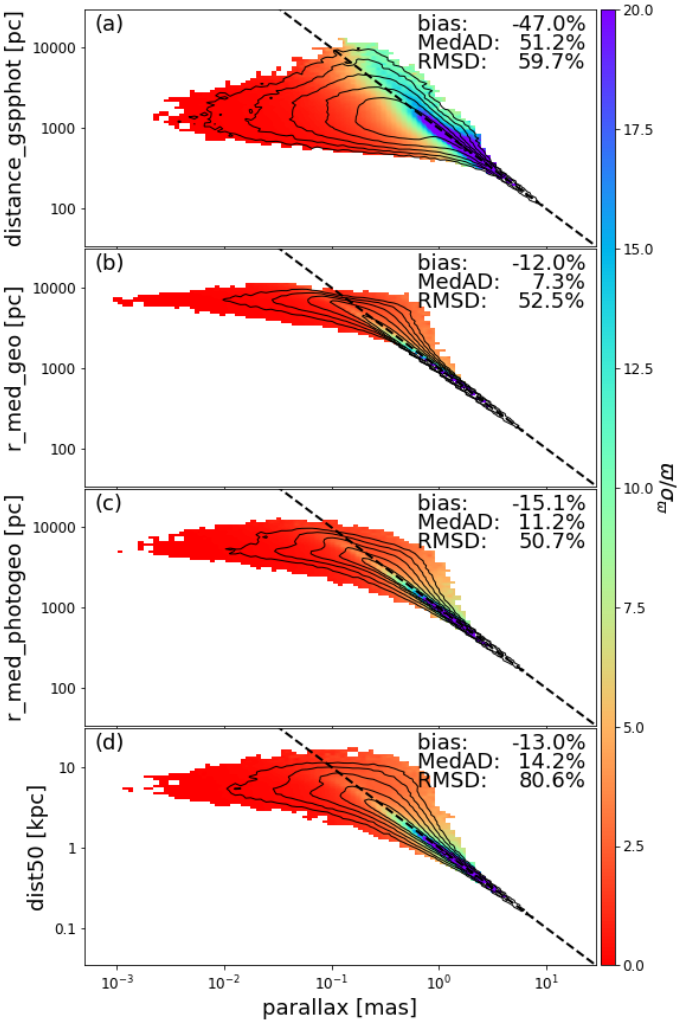

Fig. 8. Comparison of distances with parallaxes for a random subset of one million stars, colour-coded by parallax signal-to-noise ratio. The dashed black line indicates the expected inverse parallax-distance relation, $\varpi = 1/d$. Black contours indicate densities dropping by factors of 3. From top to bottom, the panels show the GSP-Phot distances from `gaia_source.distance_gspphot`, the geometric distances from Bailer-Jones et al. (2021), the photogeometric distances from Bailer-Jones et al. (2021), and the distances from Anders et al. (2023), respectively.

The distances from the two modules and the measured parallaxes agree qualitatively overall. However, GSP-Phot distances exhibit a significantly tighter agreement with the parallaxes than those from MSC, despite the single-star assumption: their mean absolute differences are only half of those for MSC and the rms differences are more than ten times smaller. However, the rms differences are dominated by a handful of outliers, whereas the absolute difference at 90% confidence is more robust, but still much higher for MSC than for GSP-Phot. One source of this mismatch likely comes from the differences in exploiting the information from BP and RP spectra: while MSC and GSP-Phot make use of the parallax and the apparent G magnitude, MSC normalises the spectra, whereas GSP-Phot keeps their calibrated amplitudes in their spectra likelihoods (see Andrae et al. 2023, for further details).

Furthermore, interpreting the difference between the two sets of estimates is more complex in practice. Modules adjust their AP sets altogether to fit the observed BP and RP spectra. We emphasise that the double-star assumption of MSC allows more free fit parameters than the single-star assumption of GSP-Phot (8 and 5, respectively). The increased number of fit parameters is likely a source of the more significant dispersion in the MSC estimates. We discuss the other APs from MSC in Sect. 3.5.2.

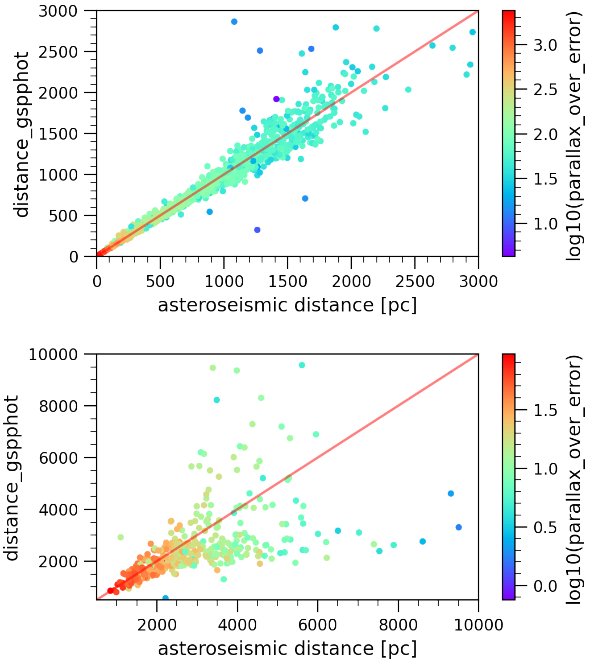

Fig. 9. Comparison of GSP-Phot distances (`gaia_source.distance_gspphot`) with asteroseismic ones for 2236 and 606 stars from Huber et al. (2017) (top) and Anders et al. (2017) (bottom). The continuous red line indicates the bisector for reference in both plots. We obtain distances in good agreement up to 2 kpc, and some outliers beyond.

3.2. Atmospheric APs

The atmospheres of stars produce the photons that *Gaia* collects. Through these photons, we can infer the physical conditions of these layers, which relate to the fundamental stellar parameters. In this section, we characterise the *Gaia* DR3 APs that describe the atmospheric state of the observed stars. We loosely split the APs into three groups: first, the basic static (equilibrium) state of an atmosphere defined by $T_{\rm eff}$, $\log g$, metallicity, [M/H], and α-abundance, $[\alpha/{\rm Fe}]$[10]; then the dynamic (departure from equilibrium) state given by the stellar classes, rotation, line emissions, magnetic activity, and mass loss or accretion; and finally, the chemical abundances.

The *Gaia* data set is primarily magnitude limited and does not select objects on any specific colour or class of stars. Consequently, the atmospheric parameters span a great variety of spectral types, from O to M, and even some L-type stars, some of which require target-specific treatment (partly handled by the ESP-modules in Apsis). Depending on the star (spectral and luminosity) class, we used either empirical or theoretical atmospheric models to estimate the atmospheric parameters of the stars, and sometimes both. The theoretical models try to model the relevant physical processes of the matter-light interaction in stellar atmospheres, while the empirical models capture some hard-to-model observational effects. The overlap between

---

[10] α-elements with respect to iron ($[\alpha/{\rm Fe}]$) refer to O, Ne, Mg, Si, S, Ar, Ca, and Ti and are considered to vary in lockstep.

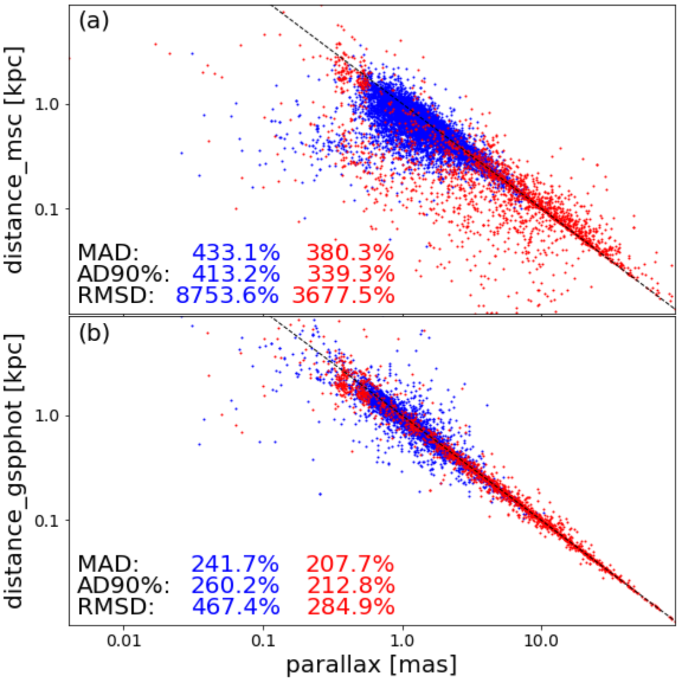

Fig. 10. Comparison of *Gaia* parallaxes with distance estimates from MSC (panel a; `astrophysical_parameters.distance_msc`) and GSP-Phot (panel b; `gaia_source.distance_gspphot`) for 2253 known spectroscopic binaries from Pourbaix et al. (2004) (red points) and 10 407 from Traven et al. (2020) (blue points). We quote the mean absolute differences (MAD), the absolute difference at 90% confidence (AD90%) and the rms differences (RMSD) to the measured parallax for the two samples in both panels. The panels show the same set of stars with estimates from both modules. The anti-diagonal dashed line highlights the parallax as inverse distance.

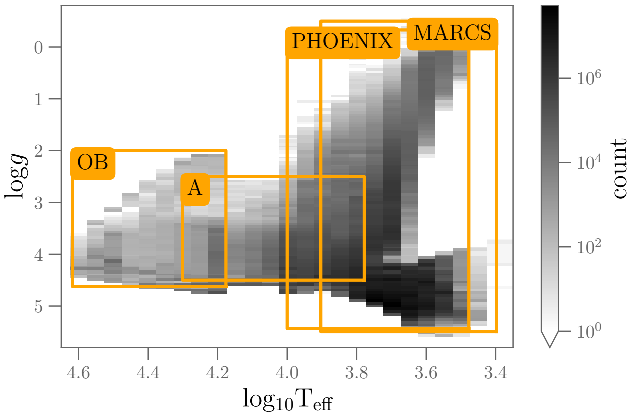

Fig. 11. Parameter space in the Kiel diagram spanned by the stellar atmosphere libraries used by GSP-Phot. Boxes indicate the spans of the libraries producing independent estimates. The density distribution represents the content from `gaiadr3.gaia_source`, which contains only one set of APs per source using the (statistically) best library (`libname_gspphot` field) for that one source.

models and application ranges of Apsis modules allows us to confirm consistency or the lack thereof (see overlaps in Figs. 2 and 11).

3.2.1. Primary atmospheric parameters: $T_{\rm eff}$, log g, [M/H] and [α/Fe]

Below we summarised our validation results for the parameters $T_{\rm eff}$, log g, [M/H], and [α/Fe] that various modules of Apsis estimate (see Table D.2). We first focus on the FGK-type stars as these constitute the majority of stars in the *Gaia* data set. Mainly GSP-Phot and GSP-Spec overlap on this stellar type interval. We emphasise that the application range of the Apsis modules varies significantly. To facilitate, we thus organise the description per module. One way to validate the *Gaia*-based APs and simultaneously quantify their precision is to compare them with large stellar surveys in the literature. The numbers below serve as a guideline for the global precision of the *Gaia* DR3 results relative to literature works. Accuracy is harder to quantify globally, but we can assess it in some specific cases, for instance, relative to *Gaia* benchmark stars (e.g. Heiter et al. 2015) and spectroscopic solar analogs (e.g. Tucci Maia et al. 2016).

GSP-Phot. Analysing BP/RP spectra, GSP-Phot provides multiple sets of APs, one for each of the four supporting theoretical atmospheric libraries: MARCS (Gustafsson et al. 2008), PHOENIX (Brott & Hauschildt 2005), A (Shulyak et al. 2004), and OB (Lanz & Hubeny 2003, 2007). Figure 11 shows their parameter space. GSP-Phot analyzes the BP/RP spectra with a Markov chain Monte Carlo approach (MCMC), which also characterises the uncertainties (method in Andrae et al. 2023). The reported estimates and uncertainties correspond to the 50th (median) and 16th and 84th percentiles of the (marginalised) MCMC samples. We also publish the MCMC chains with the catalogue through the DataLink protocol (Dowler et al. 2015) implemented by the *Gaia* Archive. We compared our APs to those reported in the APOGEE (Abdurro'uf et al. 2022), *Gaia*-ESO (Gilmore et al. 2012; Blomme et al. 2022), GALAH (Buder et al. 2021), LAMOST (Wu et al. 2011, 2014), and RAVE (Steinmetz et al. 2020) catalogues. We characterised the GSP-Phot results by a median absolute error in $T_{\rm eff}$ of 119 K, and a mean absolute error of 180 K across the various mentioned datasets (details in Andrae et al. 2023). The difference between the two statistics translates into the complexity of the distributions. The variations in log g and [M/H] affect the BP and *RP* spectra only weakly in contrast with the temperature. Although GSP-Phot analyses the BP/RP spectra with the *Gaia* parallax information and isochrone models, this combination allows GSP-Phot to determine log g and [M/H] estimates. In Andrae et al. (2023), we compared our values to seismic log g values of solar-like oscillators from Serenelli et al. (2017) and Yu et al. (2018) and we found a median absolute error of 0.2 dex for log g. For [M/H], we find that GSP-Phot estimates are typically too low by 0.2 dex and exhibit additional systematics. We thus caution against using [M/H] estimates without further investigation. However, we find that [M/H] still encodes some useful information about metallicity, for instance, with empirical calibrations (see Andrae et al. 2023, Sect. 3.5.3 for details). The comparison of [M/H] to [Fe/H] from APOGEE DR17 (Abdurro'uf et al. 2022) gives a median absolute deviation of 0.2 dex, with globally no offset for stars with log g $>$ 2.5, based on more than 400 000 FGK stars in common (see Andrae et al. 2023, their Fig. 10). We assessed the typical precision of [M/H] by measuring the dispersion among FGK members in 187 open clusters with known metallicities, from −0.50 dex to +0.43 dex (see Andrae et al. 2023, their Fig. 11). The residuals show an explicit dependency on the parallax S/N (see Fig. 9 of Andrae et al. 2023). The median absolute deviation is 0.2 dex for G $<$ 16 mag, while for fainter stars, the GSP-Phot metallicities appear to be underestimated by 0.6 dex (median offset) with a dispersion reaching

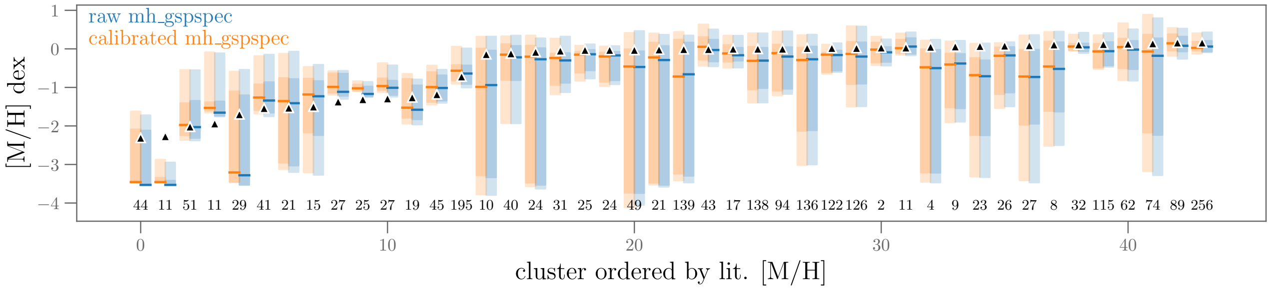

Fig. 12. [M/H] abundance distributions of member stars per cluster for the GSP-Spec Matisse-Gauguin algorithm (`astrophysical_parameters.mh_gspspec`) before and after recommended adjustements (see Sect. 3.2.1) The clusters are ordered by ascending [M/H] literature values (the lowest values to the left). These literature values are given by the black triangles. Cantat-Gaudin et al. (2020) provided the cluster members. Marks indicate the median of the distributions, and shaded regions indicate the 68% and 98% intervals. Numbers at the bottom indicate how many estimates were available to represent the distribution. Ideally, all predictions are within a small interval, which agrees with the triangles. We did not filter the estimates using the flags to keep enough stars per cluster, but nevertheless, the agreement is remarkable.

0.5 dex. As one can expect, the GSP-Phot performance varies significantly from star to star and depends on the stellar physical conditions encoded in the spectral type, luminosity class, and [M/H], for instance (see Fig. 11 from Andrae et al. 2023). For metal-poor stars with $[\mathrm{M/H}] \leq -1.5$ dex, the metallicity sensitivity of the BP and RP spectra diminishes drastically, which causes GSP-Phot to overestimate [M/H]. This loss of sensitivity is typical of optical photometric metallicity indicators, which is one of the reasons for dedicated passband designs (e.g. Jordi et al. 2010; Starkenburg et al. 2017; López-Sanjuan et al. 2021) and spectral indices (e.g. Johansson et al. 2010). Andrae et al. (2023) interpret this as a consequence of [M/H] having the weakest impact on BP and RP spectra and thus being the parameter that is easiest to compromise.

GSP-Spec. Analysing RVS spectra with primarily $S/N > 20$ (i.e, $G \lesssim 16$ mag), GSP-Spec estimates the stellar APs using synthetic spectra based on MARCS models and with two different algorithms (Matisse-Gauguin and ANN; see Manteiga et al. 2010; Recio-Blanco et al. 2016, 2023, for details). Unlike GSP-Phot, GSP-Spec does not exploit additional information such as parallax or photometric measurements. GSP-Spec estimates uncertainties per star from the ensemble of APs from 50 Monte Carlo realisations of the spectra: for each, GSP-Spec draws a spectrum from the noise (i.e. spectral flux covariances estimated by Seabroke et al., in prep.) and derives a set of atmospheric parameters and chemical abundances (see Sect. 3.2.3). The reported lower and upper confidence values correspond to the 16th and 84th percentiles of the MC results per star, respectively. In addition, we provide quality flags to identify estimates potentially suffering from bad pixels, a low S/N, significant line broadening due for instance to stellar rotation ($v \sin i$), poor radial velocity (RV) correction, and grid border effects. We discuss the results from the Matisse-Gauguin and ANN algorithms below. They are available in the `astrophysical_parameters` table and `astrophysical_parameters_supp`, respectively.

We validated and quantified the accuracy of the Matisse-Gauguin parameters for FGK stars against literature data. We selected results with corresponding AP flags equal to zero and compared our estimates with APOGEE DR17 (Abdurro'uf et al. 2022), GALAH-DR3 (Buder et al. 2021), and RAVE-DR6 (Steinmetz et al. 2020). We find with a comparison with APOGEE-DR17 a median offset and MAD of $(-32; 58)$ K, $(-0.32, 0.12)$ dex, and $(+0.04, 0.08)$ dex for T_{eff}, $\log g$, and [M/H]. The spectra from RAVE and RVS share very similar wavelength coverage, which led Recio-Blanco et al. (2023) to extensively compare the GSP-Spec performance against these stellar parameters. We find similar statistics when comparing with the other catalogues (see details in Recio-Blanco et al. 2023, especially their Fig. 11).

However, we found a bias in the $\log g$ and smaller biases in the [M/H] and [α/Fe] values from Matisse-Gauguin for giant stars. Recio-Blanco et al. (2023) provided corrective prescriptions in the form of a polynomial function of $\log g$ and suggested an analogous ($\log g$-dependent) correction for [M/H] and [α/Fe] to reduce this issue between dwarfs and giants. We calibrated the $\log g$ and [M/H] corrections on the AP values from APOGEE-DR17, GALAH-DR3, and RAVE-DR6 simultaneously. These corrections lead to $\log g$ and [M/H] median offsets and MAD of APOGEE-DR17 to $(-0.005; 0.15)$ dex and $(0.06; 0.12)$ dex, respectively. However, we calibrated the [α/Fe] correction on a sample of solar-like stars (in terms of metallicity, galactocentric position, and velocity; see Sect. 3.2.3). This correction reconciles dwarf and giant on the same [α/Fe] scale. We found $[\alpha/\mathrm{Fe}] = 0$ for all stellar types after calibration in the solar-like sample on average. We further assessed the typical precision of the uncorrected [M/H] by measuring the dispersion in stellar clusters of known metallicity, similarly to what we did for GSP-Phot. Figure 12 compares the dispersion of the [M/H] abundance distributions of member stars per clusters for GSP-Spec Matisse-Gauguin algorithm before and after the recommended adjustments. Even though the corrections did not affect the overall agreement, we note that we did not apply filters based on the associated flags. We further restricted ourselves to the FGK members in 162 open clusters of Cantat-Gaudin et al. (2020), and we found an average MAD of 0.11 dex per cluster. We noted a larger dispersion and a negative offset (-0.12 dex) for dwarfs. For 64 globular clusters ($[\mathrm{M/H}] \leq -0.50$ dex), the typical dispersion per cluster is 0.20 dex with an median offset of +0.12 dex. However, these statistics describe the data regardless of the quality flags. If we require unset [M/H] flag bit zero (see details in Recio-Blanco et al. 2023), the metallicities agree better with the literature, with absolute offsets values lower than 0.10 dex, and with typical dispersions of 0.075 dex for open clusters and 0.05 dex for globular clusters. However, the filtering also reduces the number of stars significantly, leaving us with 40% of the 2271 members of open clusters and only 4% of 1224 members in globular clusters. These sources are primarily removed for low-S/N spectra because GCs lie far away. These settings also remove fast rotators, hot stars, and some K- and M-giants in

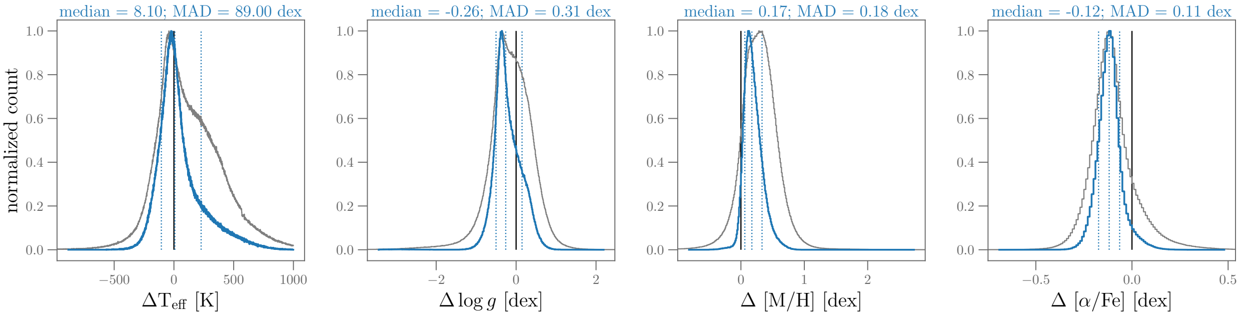

Fig. 13. Comparison of the APs from GSP-Spec-Matisse Gauguin and ANN for the *Gaia* DR3 sample with the first 13 and 8 values in `astrophysical_parameters.flags_gspspec` and `astrophysical_parameters_supp.flags_gspspec_ann` equal to zero. This represents 1 084 427 stars in the *Gaia* DR3 catalogue. For reference, we also indicate the distribution without the flag filtering in grey.

OCs. Finally, they also filter out stars nearby the model grid borders, predominantly hot dwarfs, and cool giants in the case of the OC and GC, respectively. However, we cannot conclude a metallicity-dependent performance from this test as metal-poor stars are rare and predominantly known in GCs.

The artificial neural networks algorithm (ANN) in GSP-Spec ANN provides a different parametrisation of the RVS spectra, independent of the Matisse-Gauguin approach. In contrast with the forward-modelling of Matisse-Gauguin, ANN projects the RVS spectra onto the AP label space. We trained the network on the same grid of synthetic spectra as the Matisse-Gauguin algorithm, in this case, adding noise according to different S/N scales in the observed spectra (Manteiga et al. 2010). The internal errors of ANN are about a fraction of the model-grid resolution and show no significant bias, confirming the ANN projection consistency of the synthetic spectra grid. In Recio-Blanco et al. (2023), we compared the ANN results with the literature values and found similar biases to those of Matisse. Equivalently, we also provide calibration relations for $T_{\rm eff}$, $\log g$, [M/H], and [α/Fe] to correct for these biases.

Figure 13 compares the APs from the two algorithms of GSP-Spec in a sample of 1 084 427 in *Gaia* DR3 with respective estimates. We also restricted this comparison to the good flag status: the first 13 and 8 values in `astrophysical_parameters.flags_gspspec` and `astrophysical_parameters_supp.flags_gspspec_ann` equal to zero. Overall, the algorithms agree with each other. Once we apply the calibration relations to both algorithm estimates, we found for spectra with $S/N \geq 150$ deviations with median values of −94 K, −0.05 dex, 0.1 dex, and 0.04 dex for $T_{\rm eff}$, $\log g$, [M/H], and [α/Fe], respectively. For the sample, we found MAD values of 93 K, 0.11 dex, 0.10 dex, and 0.05 dex, respectively.

GSP-Phot and GSP-Spec overlaps. Figure 14 compares the temperatures and gravity estimates from GSP-Phot and GSP-Spec. The $T_{\rm eff}$ estimates strongly agree overall, but some outliers remain visible in the plot. They most likely originate from GSP-Phot sensitivity to low-quality parallaxes. In particular, we traced back the plume at $\log_{10}$`teff_gspphot` ~ 3.8 to variable stars (see Andrae et al. 2023 for details). In this sample, we found a median offset of 98 K, an MAD of 246 K. It is very apparent that the $\log g$ estimates systematically differ strongly between the modules. The recalibration prescription from Recio-Blanco et al. (2023) mitigates the differences, but does not remove them completely. We found a median offset of 0.35 dex, and an MAD of 0.34 dex. Recio-Blanco et al. (2023) identified a similar trend in the GSP-Spec $\log g$ values when they compared their values to those of the literature (see their Fig. 10).

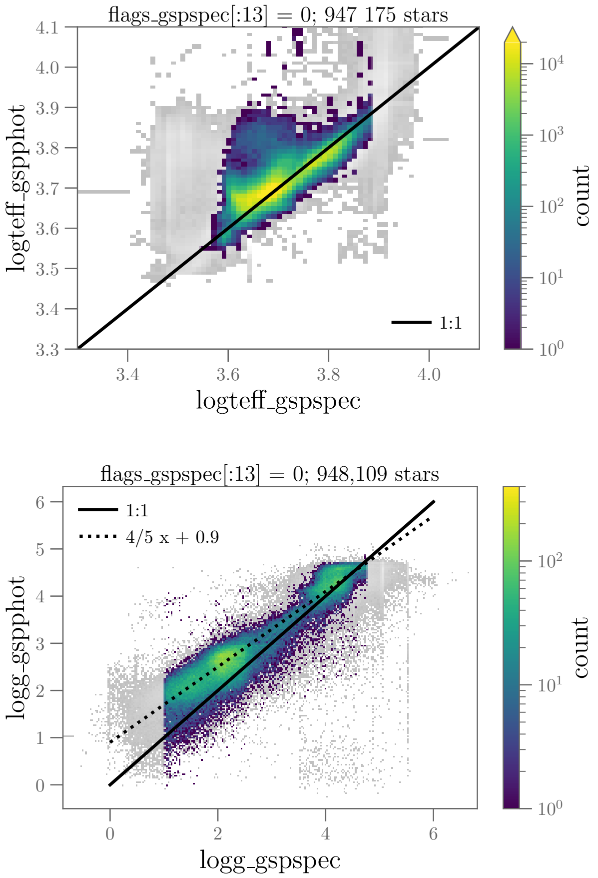

Fig. 14. Comparison of the temperatures (top) and surface gravity (bottom) estimates from GSP-Phot and GSP-Spec. The GSP-Phot and GSP-Spec algorithms are calibrated to the literature values. We plot in grey the ~3.2 million sources in *Gaia* DR3 with both `astrophysical_parameters.teff_gspphot`, `teff_gspspec` and `astrophysical_parameters.logg_gspphot`, `logg_gspspec`. The highlighted distribution corresponds to those with the first 13 values in `flags_gspspec` equal to zero (~1 million sources). We indicate the identity lines and the identified divergence in $\log g$ between the modules. We note that the GSP-Spec recommended calibration of $\log g$ does not affect this comparison significantly.

Solar analogues are stars that are closest to the Sun in temperature, gravity, and metallicity. We selected more than 200 spectroscopic solar analogues from the literature (mostly from Datson et al. 2015; Tucci Maia et al. 2016) with relative $T_{\rm eff}$ within ±100 K, $\log g$, and [Fe/H] within ±0.1 dex to those of the solar values. We compared the biases and dispersion of the GSP-Phot and GSP-Spec Matisse-Gauguin APs in this sample of stars. We note that solar analogues are dwarf stars, which are little to not at all affected by the Matisse-Gauguin corrections mentioned above. We find that GSP-Phot underestimates $T_{\rm eff}$ by between 30 K (PHOENIX) and 90 K (MARCS), with a standard deviation $\sigma \sim 100$ K in both cases. In contrast, GSP-Spec estimates have essentially no $T_{\rm eff}$ bias (+10 K), but a slightly higher dispersion ($\sigma \sim 130$ K). Irrespective of the atmosphere library (`libname_gspphot`), GSP-Phot underestimated the $\log g$ values by 0.12 dex, but with a standard deviation of $\sigma \sim 0.14$ dex, they remain statistically compatible with the solar value. GSP-Spec results are as accurate as those from GSP-Phot around the solar locus, but they present a higher dispersion of 0.42 dex (calibration of $\log g$ does not change this value). We recall that GSP-Spec uses only the RVS spectra as input, while GSP-Phot also uses parallaxes and constraints from isochrones. [M/H] values are nearly solar for GSP-Spec with an offset of 0.1 dex and $\sigma \sim 0.05$ (again, without significant impact of the recommended corrections), but we found larger offsets when GSP-Phot used PHOENIX (−0.4±0.2 dex) and MARCS models (−0.2±0.2 dex). Andrae et al. (2023) discussed the systematic and significant discrepancies between APs based on the PHOENIX and MARCS libraries. For solar-like stars, they found substantial differences in the original atmosphere models that are still under investigation at the time of writing this manuscript.

Ideally, GSP-Phot and GSP-Spec would return results in perfect agreement with each other. In practice, they do not, but instead complement each other. The two modules analyse data with different spectroscopic resolutions and wavelength ranges. To first order, GSP-Phot relies on the stellar continuum over the whole optical range from the BP/RP low-resolution spectra (from 330 to 680 nm). In contrast, GSP-Spec investigates atomic and molecular lines in the continuum-normalised medium-resolution spectra in the narrow infrared window of RVS (from 846 to 870 nm). Hence the modules analyse different aspects of the light emitted from stars. Additionally, interstellar extinction significantly affects the BP and RP spectra, but RVS data only in the region of the diffuse interstellar band around 860 nm (e.g. Gaia Collaboration 2023c). Therefore, AP determination from GSP-Phot significantly depends on a correct determination of the amount of extinction, while it has little impact on the AP inference from GSP-Spec (see Sect. 3.4).

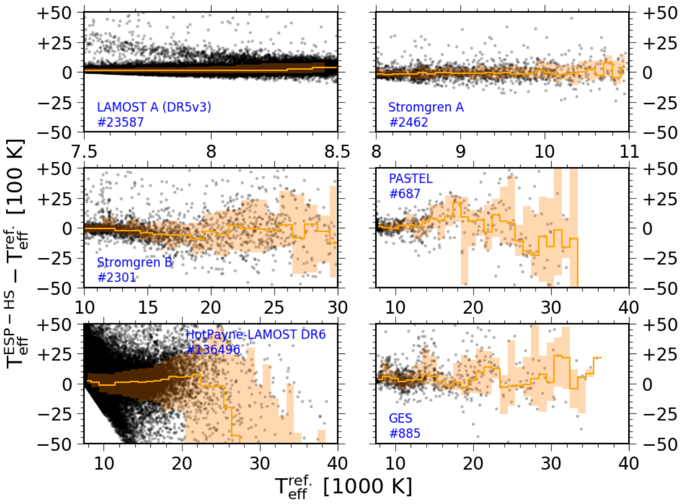

Fig. 15. Effective temperature difference between values obtained by ESP-HS and those from literature catalogues as a function of the literature $T_{\rm eff}$. We indicate the running median with the solid orange lines and the interquantile dispersions by the shaded orange regions. The blue numbers in each panel indicate the number of stars that is compared. We emphasise that the scales on the x- and y-axes are scaled by 1000 and 100 K, respectively. Reference estimates are from the LAMOST (DR5) A-type stars (LAMOST A (DR5v3), Luo et al. 2019), derived from Stromgren photometry (Stromgren A and Stromgren B) by adopting the updated calibration of Napiwotzki et al. (1993), from the Pastel catalogue (PASTEL; Soubiran et al. 2016), from the LAMOST OBA (DR6) catalogue (HotPayne LAMOST DR6; Xiang et al. 2022), and from the *Gaia* ESO survey (GES; Blomme et al. 2022).

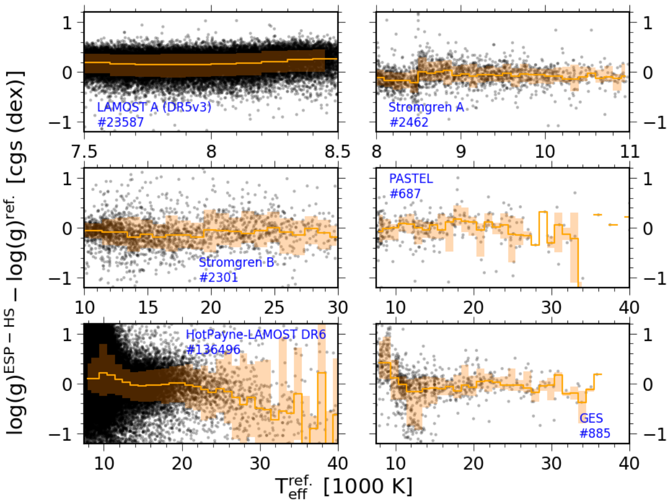

Fig. 16. Surface gravity difference between values obtained by ESP-HS and those from literature catalogues. Same conventions as Fig. 15.

ESP-HS. Stars hotter than 7500 K (O-, B-, and A-type stars) undergo a specific analysis by the ESP-HS module. It operates in two modes: simultaneous analysis of BP, RP, and RVS spectra (BP/RP+RVS), or BP and RP alone. ESP-HS first estimates the stellar spectral type[11] from its BP and RP spectra to further analyse O-, B-, and A-type stars alone (`astrophysical_parameters.spectraltype_esphs`: CSTAR, M, K, G, F, A, B, and O). Hot stars of these spectral types are inherently massive and short-lived according to stellar evolution, and consequently, they are young stars[12]. Hence, ESP-HS assumes a solar chemical composition, and therefore it does not provide any metallicity estimate. See the module details in the *Gaia* DR3 online documentation, Sect. 11.3.8. For stars hotter than 7500 K, the overlap between GSP-Phot and ESP-HS allows us to cross-validate our effective temperature estimates. We find that ESP-HS tends to provide higher $T_{\rm eff}$ than the GSP-Phot values due to different internal ingredients. We quantify the potential systematics from ESP-HS further with respect to catalogues in the literature. Figures 15 and 16 show the residuals relative to literature compilations for $T_{\rm eff}$ and $\log g$, respectively. Below 25 000 K, we obtain reasonable agreement of ESP-HS temperatures with the catalogues estimates. Overall, the dispersion in $T_{\rm eff}$ increases with temperature from ∼300 K for the A-type stars to 500–2000 K for B-type stars. Above 25 000 K, we find, relative to the $T_{\rm eff}$ versus spectral type scale of Weidner & Vink (2010), a systematic underestimation of our

---

[11] Originally produced by ESP-HS, the spectral type classification procedure moved to the ESP-ELS module for practical reasons.

[12] We assume that our data are dominated by disk stars and therefore ignore horizontal branch stars from the halo. ESP-HS does not include models for white dwarf atmospheres.

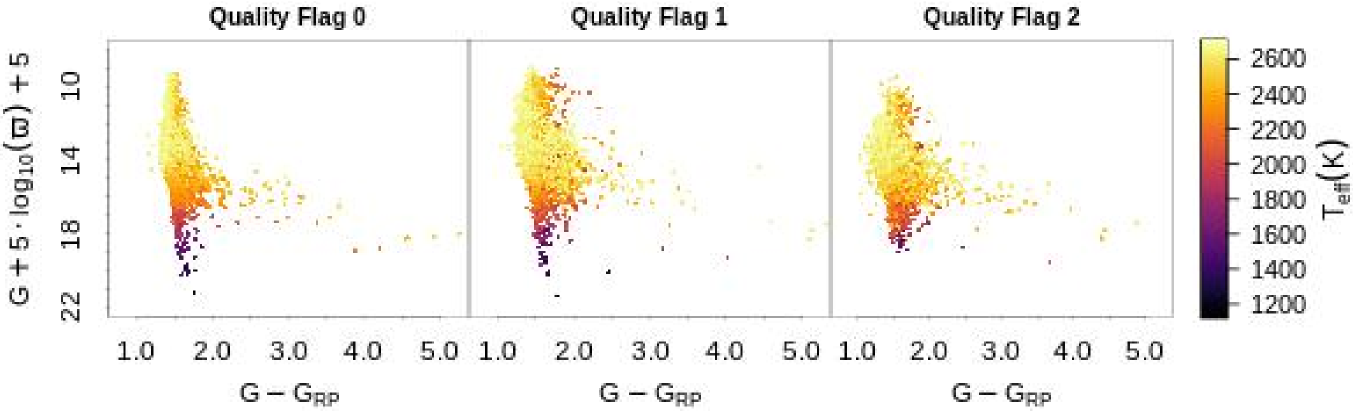

Fig. 17. Colour-absolute magnitude diagram of the UCD candidates from the ESP-UCD analysis. The y-axis uses the inverse parallax as a distance estimate, and we assumed negligible extinction. The colour code reflects the ESP-UCD $T_{\rm eff}$ estimate according to the scale on the right side.

temperatures by 1000 K to 5000 K for the Galactic O-type stars, while this can be up to 10 000 K for their LMC target samples. However, we also recall that this particular LMC sample is of subsolar metallicity, that is, beyond the model limits of ESP-HS. Similarly, the dispersion in $\log g$ increases from about 0.2 dex in the A-type star temperature range to ~0.4 dex for the O-type stars. More detailed numbers for the offset and dispersion of $T_{\rm eff}$ and $\log g$ relative to the catalogues considered in Figs. 15 and 16 are available in Gaia Collaboration (2022). We found that ESP-HS underestimated uncertainties by a factor of 5 to 10 in the BP/RP+RVS mode while reporting the correct order of magnitude in the BP/RP-only mode. We did not inflate the reported uncertainties in the *Gaia* DR3 catalogue accordingly. The first digit of `astrophysical_parameters.flags_esphs` reports which mode the ESP-HS estimates come from (i.e. 0: BP/RP+RVS, 1: BP/RP-only). We emphasise that we filtered out a significant number of poor fits of ESP-HS, but known outliers remain (e.g. $T_{\rm eff} > 50\,000$ K). In addition, ESP-HS processed white dwarfs (WD) despite not using a suitable library. Finally, some classes of stars intrinsically cooler than 7500 K (e.g. RR Lyrae stars) were misclassified as O, B, or A-type stars and ESP-HS analysed and reported about them assuming a correct classification.

ESP-UCD. At the faint end of the luminosity distribution, we transition between standard hydrogen-burning stars and brown dwarfs, which are not massive enough to sustain nuclear fusion. We define ultracool dwarfs (UCDs) as sources of spectral type M7 or later (Kirkpatrick et al. 1997) which corresponds to $T_{\rm eff} \leq 2656$ K according to the calibration by Stephens et al. (2009). Using a combination of parallaxes, colour indices, and RP spectra, we identified 94 158 UCD candidates in *Gaia* DR3 with $T_{\rm eff} < 2700$ K even though the *Gaia* instruments are suboptimal to observe these intrinsically faint sources. We note that unsurprisingly the flux in the BP band is negligible (or even absent) for these very red and faint sources. The adopted threshold (2700 K) is slightly hotter and more inclusive than the quoted 2656 K to take the $T_{\rm eff}$ estimate uncertainties into account. Creevey et al. (2023) detailed our characterisation module, the complete UCD selection criteria, our quality filters, and our training set definition. ESP-UCD produced effective temperatures for 94 158 UCD candidates in *Gaia* DR3, the vast majority of them (78 108) having $T_{\rm eff} > 2500$ K. However, *Gaia* DR3 provides temperature estimates from ESP-UCD (`astrophysical_parameters.teff_espucd`), but it does not include the corresponding $\log g$ or [M/H] estimates due to the poor performance of ESP-UCD on these properties and a severe lack of literature reference in this regime. We plan to publish them in *Gaia* DR4. ESP-UCD provides a flag (`astrophysical_parameters.flags_espucd`) to encode the quality of the data in one of three categories based on the Euclidean distance between a given RP spectrum and the closest template in the training set and the S/N of the integrated RP flux. Quality flag 0 corresponds to the best RP spectra distance below 0.005; quality 1 corresponds to sources with distances between 0.005 and 0.01 and $S/N > 30$ relative uncertainties $\sigma_{\rm RP}/f_{\rm RP} \leqq 0.03$; and finally quality flag 2 corresponds to sources with distances between 0.005 and 0.01 but $S/N < 30$ (the *Gaia* DR3 online documentation provides a more detailed description of the quality flags).

Figure 17 shows the colour-absolute magnitude diagram (CAMD) for all the UCD candidates we detected for the three ESP-UCD quality categories. We find good consistency in CAMD positions and the inferred effective temperatures: as expected for these stars, their temperatures strongly correlate with M_G. We note that Fig. 17 uses the inverse parallax as a good distance proxy to approximate M_G, because 95% of the sources have S/N $\varpi/\sigma_\varpi > 5$ (the median parallax S/N in the three quality categories 0, 1 and 2 are 25, 11, and 7.5, respectively). Overall, as the quality degrades, the vertical sequence spreads and becomes noisier with respect to the temperature scale.

More quantitatively, we compared our inferred temperatures with those of the *Gaia* UltraCool Dwarf Sample (GUCDS; Smart et al. 2017, 2019). We translated the GUCDS spectral types using the calibration by Stephens et al. (2009), and we found an rms of 103 K and a MAD of 88 K for the entire sample (see Fig. 18). We note that these statistics include low-metallicity and young sources. Figure 19 compares the ESP-UCD effective temperatures with SIMBAD spectral types when available. This sample includes and extends the GUCDS. We indicate the two spectral type-$T_{\rm eff}$ calibration relations by Stephens et al. (2009) for optical and infrared spectral types to provide a comparison reference. We used these two relations to define the empirical training set of the ESP-UCD module. We note that the spectral type M6 corresponds to an effective temperature ~2800 K. This temperature is hotter than the ESP-UCD parameter space limit. However, ESP-UCD attributed cooler $T_{\rm eff}$ values to some of these stars, which we published. They led to the apparent negative bias for the M6V bin in Fig. 19, however. Gaia Collaboration (2023f, Sect. 7) further explored the stellar population of UCDs in the Galaxy and their properties.

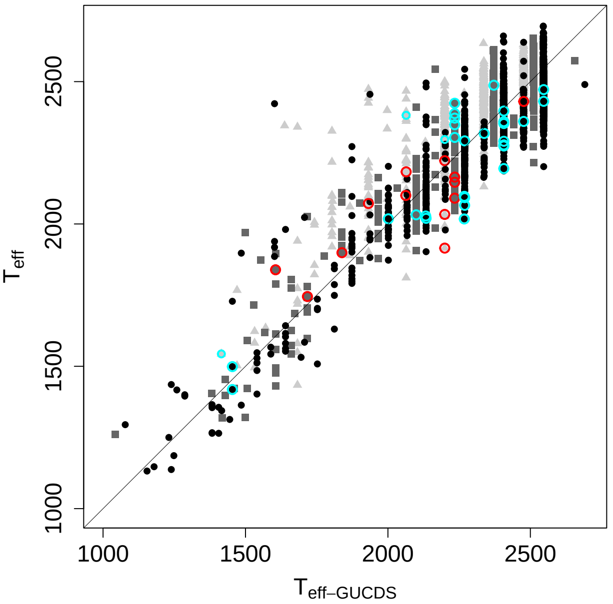

Fig. 18. Comparison of the effective temperatures (in Kelvin) between ESP-UCD estimates and those obtained by converting the GUCDS spectral types using the calibration by Stephens et al. (2009). Black circles correspond to quality 0, dark grey squares to quality 1, and light grey triangles to quality 2. Cyan symbols denote low-metallicity sources, and red symbols denote young sources.

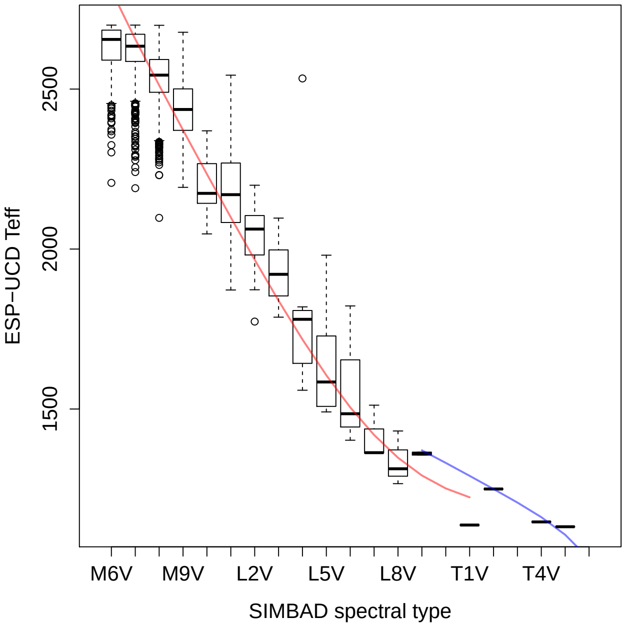

Fig. 19. Comparison of the SIMBAD spectral types with the ESP-UCD effective temperatures. The boxplots represent the medians, interquartile ranges, and outliers of each spectral type. The solid red and blue lines represent the calibrations from Stephens et al. (2009) for the M and L optical spectral types and the infrared spectral type-T sources, respectively.

3.2.2. Secondary atmospheric estimates: stellar classes, rotation, emission, and activity

Classification. There are four main stellar classifications from Apsis (see fields in Table D.3). First, DSC primarily distinguishes between extragalactic sources (quasars and galaxies) and stars (single stars, physical binaries, and white dwarfs). Users can classify sources using the probabilities from DSC of a source to belong to a given class. However, 99% of *Gaia* DR3 sources processed by Apsis are most certainly stars (or binaries). Hence the classification from DSC is not the most relevant for stellar objects (see Bailer-Jones et al. 2021; Creevey et al. 2023).

OA measures similarities between observed BP and RP spectra of different sources to produce an unsupervised classification using self-organising maps (SOMs; Kohonen 2001). One can use these maps to find similar groups of stars once labelled (details in Creevey et al. 2023) and peculiar or outlier sources (see Sect. 3.6). Finally, the user might prefer using the spectral types from ESP-HS and the classification of ESP-ELS for emission-line star types of stellar sources. This section focuses on the ESP-HS and ESP-ELS classification tailored to stellar objects.

Table 2. Number of identified candidates per ELS class and per quality flag (QF) value.

Class	`classlabel_espels_flag` (QF)						
	0	1	2	3	4	10–14	20–24
Be	3210	3118	2815	2332	1475	122	3879
Ae/Be	35	94	231	519	972	299	1754
T Tauri	914	2052	1594	1083	740	1180	27 594
dMe	0	1	5	54	178	43	380
PN	37	52	83	85	16	0	0
WC	106	13	8	7	2	0	0
WN	173	38	29	142	47	0	0

Notes. QF ≤ 2 implies a probability higher than 50%.

ESP-HS estimates the spectral type of a source from its BP and RP spectra. While primarily focused on hot stars, it provides the following main classes: CSTAR, M, K, G, F, A, B, and O. We find from a cross-match with the LAMOST OBA catalogue of Xiang et al. (2022) that ESP-HS obtained 62% of the Galactic A- and B-stars (assuming the other catalogue is complete). Conversely, we find only 186 (30%) of the 612 Galactic O-type stars published in the Galactic O-type Stars catalogue (GOSC; Maíz Apellániz et al. 2013). This low fraction reflects the persisting difficulties of deriving reliable hot-star APs from *Gaia* BP and RP spectra.

ESP-ELS identifies the BP and RP spectra that present emission features and classifies the corresponding target into one of the seven ELS classes listed in Table 2. We recall that ESP-ELS processed stars brighter than $G < 17.65$ mag (see Sect. 2). The ESP-ELS classification as ELS relies on detecting line emission and primarily on measuring the Hα pseudo-equivalent width (see below). We tagged particular failure modes with the quality flag (`astrophysical_parameters.classlabel_espels_flag`; see Table 2). Primarily, this flag takes values ranging from 0 (best) to 4 (worst) depending on the relative strength of the two most probable classes (i.e., ESP-ELS published random forest classifier class probability estimates in `astrophysical_parameters.classprob_espels_wcstar`, `classprob_espels_wnstar`, etc.). In addition, `astrophysical_parameters` indicates the GSP-Phot AP values we used to make the classification was removed by the final *Gaia* DR3 filtering or when these APs disagreed with the spectral type estimated by ESP-ELS. These two modes correspond to `classlabel_espels_flag` first bit 1 and 2, respectively (Table 2). We emphasise that the identification of Wolf-Rayet stars (WC and WN) and planetary nebula does not depend on any APs. All but 5 of the 136 detected WC stars have typical spectroscopic features. The missed Galactic WC stars taken from

Rosslowe & Crowther (2015)[13] are usually fainter than the processing limit of $G = 17.65$ mag or have a low-quality Hα pEW estimate (e.g. weaker emission lines with type WC8 or WC9). Half of the 431 WN star candidates do not show any typical emission line. Most of the known WNs in the literature are fainter than the processing limit of ESP-ELS.

Stellar rotation. While deriving the astrophysical parameters, ESP-HS also measures the line broadening on the RVS spectrum by adopting a rotation kernel. This by-product of the ESP-HS processing corresponds to a projected rotational velocity ($v \sin i$; `astrophysical_parameters.vsini_esphs`) obtained on co-added mean RVS spectra (Seabroke et al., in prep.). It therefore differs from `gaia_source.vbroad` obtained on epoch data by the radial velocity determination pipeline (Frémat et al. 2023). The ESP-HS estimate suffers from the same limitations as `vbroad`, mostly the limited resolving power of the RVS. This is increased by the poor $v \sin i$-related information for OBA stars in this wavelength domain. In addition, the determination of `vsini_esphs` is affected by the higher uncertainty of the epoch RV determination, expected for stars hotter than 10 000 K (Blomme et al. 2023), and by the use of a Gaussian *mean* ALong-scan LSF with a resolving power of 11 500 (Creevey et al. 2023, Sect. 2.2).

In Fig. 20 we present a comparison of the $v \sin i$ measurements by ESP-HS with those obtained in the framework of the LAMOST survey for OBA stars which presents the largest overlap with the results of ESP-HS compared to other surveys. The agreement rapidly decreases with magnitude and effective temperature, while the features that are most sensitive to rotational broadening disappear from the RVS domain. The half interquantile dispersion (i.e. 14.85%−15.15%) varies from 25 km s^{-1} to 40 km s^{-1} in the A-type $T_{\rm eff}$ domain when the magnitude G ranges from 8 to 12, respectively. At hotter temperatures, it varies from 60 km s^{-1} to 75 km s^{-1} at $G = 8$ and $G = 12$, respectively.

Hα emission. The ESP-ELS classification of a star as ELS primarily relies on measuring the Hα pseudo-equivalent width (pEW; `astrophysical_parameters.ew_espels_halpha`). However, measuring the Hα emission line is challenging due to the low resolving power of BP and RP spectra and the steep loss of transmission at this wavelength (blue side). Figure 21 compares our Hα pEW estimates to the values provided by various authors (Raddi et al. 2015; Dahm & Simon 2005; Vioque et al. 2018; Newton et al. 2017; Silaj et al. 2010; Manoj et al. 2006). We found a general consistency between the estimates, except for stars cooler than 4000 K, for which overlapping spectral molecular bands significantly alter the local continuum. We mitigated this effect using synthetic spectra and the GSP-Phot APs. However, the mismatches between the observed and theoretical spectra and some systematics in the APs we used to select the synthetic spectra caused us to misclassify active M dwarf and T Tauri stars. For the hotter targets, we attempted to link the ESP-ELS estimate, pEW(Hα), and the published measurements presented in Fig. 21 with the following linear relation:

$$\mathrm{EW}^{\mathrm{ref.}}(\mathrm{H}\alpha) = \alpha + \beta \times \mathrm{pEW}(\mathrm{H}\alpha), \qquad (1)$$

where Table 3 provides the coefficients, α and β, with their uncertainty. We indicate with the orange lines the fitted relations in Fig. 21.

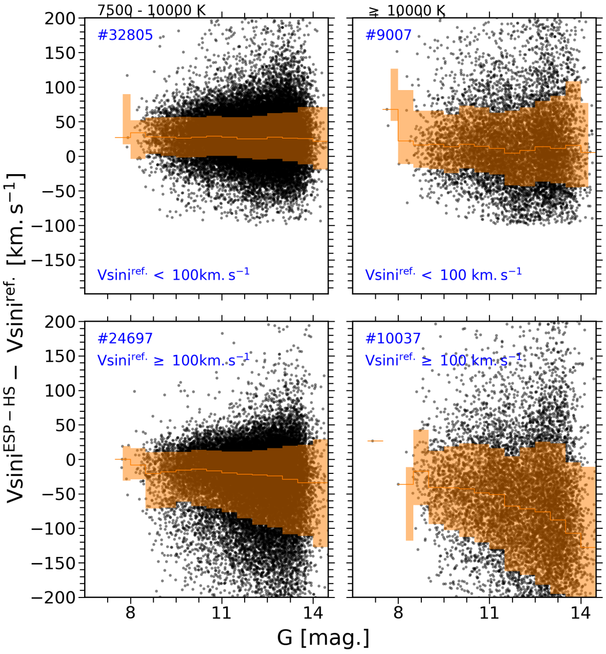

Fig. 20. Distribution with G magnitude of the differences between the LAMOST OBA results ($v \sin i^{\rm ref.}$) and ESP-HS $v \sin i$ measurement. Stars cooler and hotter than 10 000 K are plotted in the left and right panels, respectively. A distinction is also made between slow (upper panel) and rapid (lower panel) rotators. The running median is shown in orange, while the interquartile dispersion (at 14.85% and 85.15%) is represented by the orange shades.

Chromospheric activity index. The ESP-CS module computed an activity index `activityindex_espcs` from the analysis of the Ca II IRT (calcium infrared triplet) in the RVS spectra for 2 141 640 stars in *Gaia* DR3. ESP-CS defines cool stars as stars with $G \lesssim 15$ mag, $T_{\rm eff} \in [3000, 7000]$ K, $\log g \in [3.0, 5.5]$ dex, and $[\mathrm{M/H}] \in [-0.5, 1.0]$ dex. Stars with APs from GSP-Spec within these intervals undergo the analysis by ESP-CS. The activity index is the excess of the Ca II IRT lines from comparing the observed RVS spectrum with a purely photospheric model (assuming radiative equilibrium). The latter depends on a set of $T_{\rm eff}$, $\log g$, and [M/H] from either GSP-Spec or GSP-Phot (`activityindex_espcs_input` set to M1 or M2, respectively), and a line broadening estimate `gaia_source.vbroad` when available. We measured the excess equivalent width in the core of the Ca II IRT lines by computing the observed-to-template ratio spectrum in a $\pm\Delta\lambda = 0.15$ nm interval around the core of each of the triplet lines. This measurement translates the stellar chromospheric activity and, in more extreme cases, the mass accretion rate in pre-main-sequence stars. Lanzafame et al. (2023) detail the ESP-CS module, method, and scientific validation.

3.2.3. Chemical abundances

In *Gaia* DR3, GSP-Spec, most specifically, the Matisse-Gauguin algorithm, provides 13 chemical abundance ratios from 12 individual elements (N, Mg, Si, S, Ca, Ti, Cr, Fe, Ni, Zr, Ce, and Nd; with the FeI and FeII species) as well as equivalent-width estimates of the CN line at 862.9 nm. These chemical indexes rely on the line list and models from Contursi et al. (2021) and Recio-Blanco et al. (2023), respectively. For each of the 13 abundance estimates, GSP-Spec reports two quality flag bits, a

---

[13] `http://pacrowther.staff.shef.ac.uk/WRcat/`

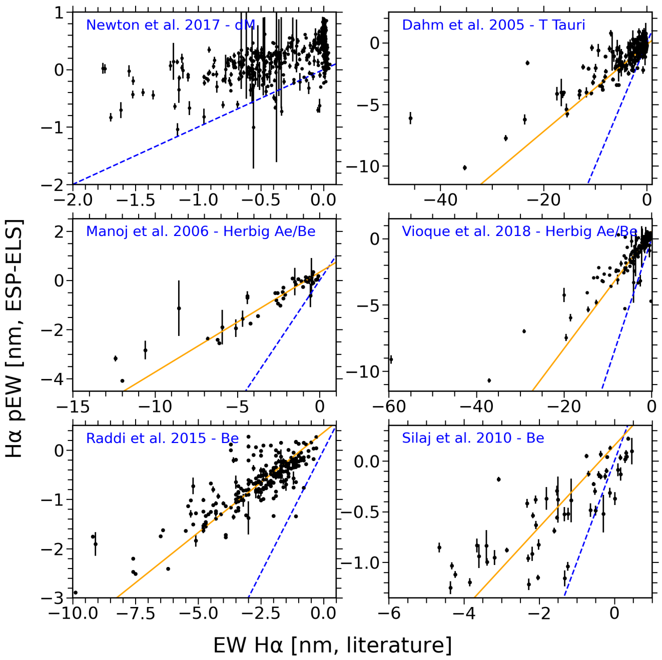

Fig. 21. Comparison of the Hα pseudo-equivalent width measured by ESP-ELS to the equivalent width published for different types of ELS classes. Each panel compares our estimates with the reference indicated in blue assuming their stellar class. The identity relation is given by the broken blue line and compared to a linear fit (orange line) through the data (Eq. (1) and Table 3).

Table 3. Coefficients of the line fitted (Eq. (1)) through the data points presented in Fig. 21.

Ref. for $\mathrm{EW}^{\mathrm{ref.}}(\mathrm{H}\alpha)$	α	β
Manoj et al. (2006)	-0.811 ± 0.165	$+2.464 \pm 0.198$
Dahm & Simon (2005)	$+0.407 \pm 0.075$	$+2.835 \pm 0.087$
Vioque et al. (2018)	-1.196 ± 0.133	$+2.266 \pm 0.141$
Raddi et al. (2015)	-0.886 ± 0.085	$+2.454 \pm 0.117$
Silaj et al. (2010)	-0.380 ± 0.241	$+2.480 \pm 0.371$

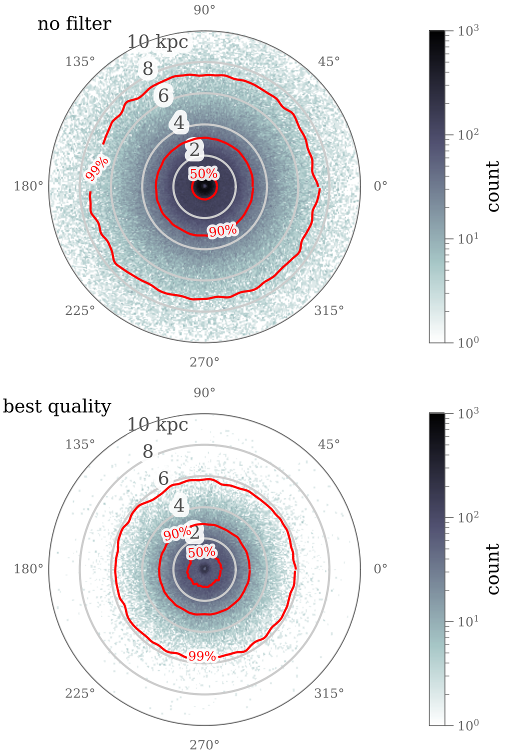

Fig. 22. Galactic top-down view of the processing coverage of GSP-Spec. The top and bottom panels show the entire and best quality samples, respectively. We projected the sources for which GSP-Spec provides any abundance estimates using the distances from GSP-Phot. The distribution centers around the Sun and the Galactic center is at (0, 8 kpc). The contours indicate the 50, 90, and 99% quantiles of the distribution, corresponding to ~1, 3, and 6 kpc, respectively.

confidence interval, the number of used spectral lines, and the line-to-line scatter (when there is more than one line). We note that the *Gaia* DR3 catalogue contains the [FeI/M] and [FeII/M] as all the other ratios, dictated by the parametrisation of the synthetic model grids in GSP-Spec. To obtain [FeI/H] and [FeII/H], one has to add the [M/H] of the star. Recio-Blanco et al. (2023) describe the definition and measurements of these chemical abundances and their quality flags in detail. Figure 22 shows the spatial extent of the abundance estimates in a top-down Galactic view. The coverage indicates that *Gaia* DR3 provides abundance estimates for a significant fraction of the stars observed by *Gaia* within 4 kpc as indicated by the 99% quantile contour. Figure 23 decompose the *Gaia* DR3 catalogue content into the individual abundance ratios for the best quality and whole sample. We note that Recio-Blanco et al. (2023) also provided intermediate selections. Gaia Collaboration (2023e) analysed the chemical abundance estimates in the context of the chemistry and Milky Way structure, stellar kinematics, and orbital parameters.

The validation of individual abundances is challenging as no fundamental standards exist for stars other than the Sun. A comparison with literature data requires particular care because of different zero-points and underlying assumptions (e.g., assumed solar-scaled composition).

We expect our derived abundances to have the usual limitations discussed in the literature stemming from model assumptions (e.g., 1- or 3-dimensional model atmospheres, hydrostatic, local thermodynamic equilibrium, the atomic line list) to observational effects (e.g. possible line blends, limited resolution of the RVS, and instrumental noise). These effects can lead to systematic offsets in the abundance determinations that depend on the atmospheric parameters. However, we were able to estimate (and correct) these systematic offsets using the GSP-Spec outputs alone and specific samples of stars. For instance, we selected stars from the immediate solar neighborhood (±250 pc from the Sun), with metallicities close to solar (±0.25) and velocities close to the local standard of rest ($\pm 25\,\mathrm{km\,s^{-1}}$). In this sample, any ratio of abundances (i.e.$[X_1/X_2]$ for two elements X_1 and X_2) deviating from zero (i.e. solar value) indicates systematics independent of the atmospheric parameters. In Recio-Blanco et al. (2023), we detail our samples and analysis, and we provide $\log g$-dependent calibration relations for 10 of the 13 chemical abundances in the form of polynomials (of the third or fourth order). In particular, Table 3 of Recio-Blanco et al. (2023) lists the coefficient values as well as the $\log g$ intervals over which the calibration is applicable (and comparison statistics with the literature). For instance, we selected sample stars from APOGEE DR17 (Abdurro'uf et al. 2022) and GALAH-DR3 (Buder et al. 2021) with GSP-Spec quality flags all equal to zero and literature uncertainties smaller than 500 K, 0.5 dex and 0.2 dex for T_{eff}, $\log g$ and [Mg/Fe]

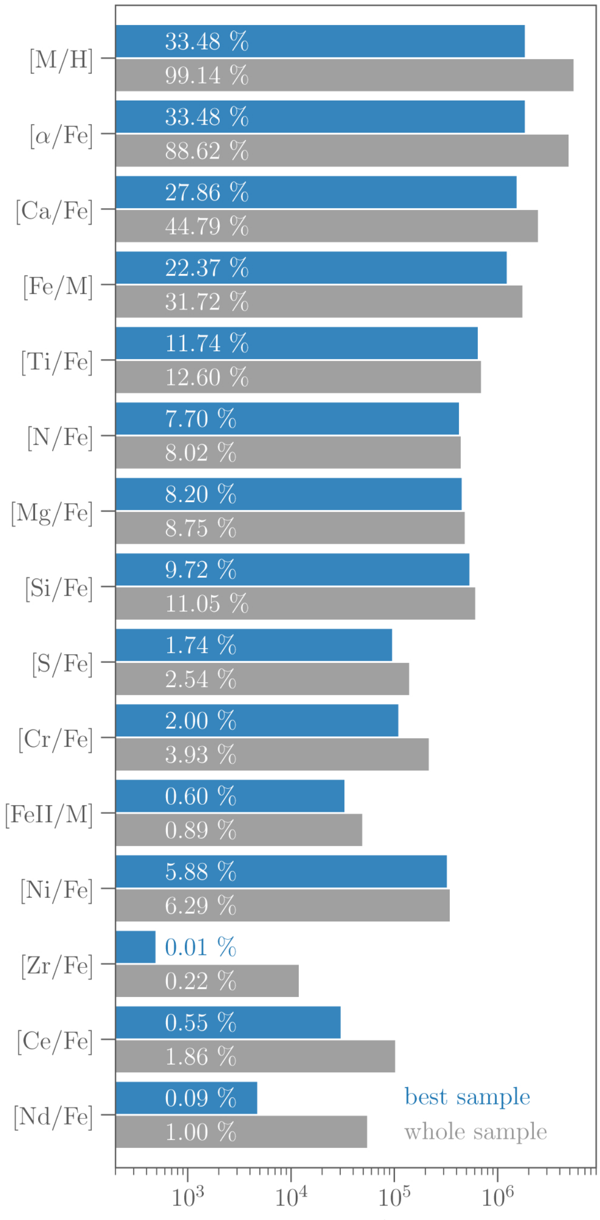

Fig. 23. Number of stars with individual abundance ratio estimates from GSP-Spec. The two sets represent the whole sample and the best quality sample (quality flags equal to zero) in grey and blue, respectively. We indicate at the top [M/H] and [α/Fe] for reference (discussed in Sect. 3.2.1). The percentages correspond to the fraction of estimates with respect to the 5 594 205 stars processed by GSP-Spec. On this scale, 1% corresponds to 40 000 stars.

or [FeI/H], respectively. This sample contains 1100 stars with [Mg/Fe] and 92 000 with [FeI/H] estimates. When comparing these with GSP-Spec abundances, we found a median abundance offset of −0.15 that dropped to 0.0 dex for [Mg/Fe] and −0.15 to 0.05 dex for [FeI/H] before and after applying those calibration relations, respectively (see Recio-Blanco et al. 2023, for further details).

3.3. Evolutionary APs

Gaia DR3 provides several parameters describing the evolution of a star. We grouped them into two sets. GSP-Phot and FLAME produce these parameters (see Table D.4). We emphasise that FLAME produces two sets of estimates: one using GSP-Phot APs and one using GSP-Spec obtained from the BP and RP, and RVS analysis, respectively, in addition to using photometry and distance (or parallax).

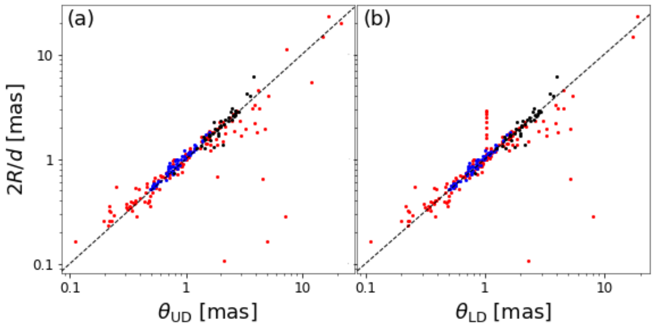

Fig. 24. Using GSP-Phot radius and distance estimates to predict measurements of angular diameters from ground-based interferometry by Boyajian et al. (2012a,b, 2013) (blue points), Duvert (2016) (red points), and van Belle et al. (2021) (black points). Panel a: comparison of $2R/d$ to estimates assuming a uniform disk. Panel b: comparison of $2R/d$ to estimates accounting for limb-darkened angular diameter.

We first discuss in Sect. 3.3.1 the observed parameters: luminosity L, absolute magnitude M_G, radius R, and gravitational redshift $rv_{\rm GR}$. These are relatively model independent, in contrast to the mass M, age τ, and evolutionary stage ϵ, which strongly depend on evolution models. We discuss them in Sect. 3.3.2.

3.3.1. Radius, luminosity, absolute magnitude, and gravitational redshift

Stellar radius. From the analysis of the BP and RP spectra, GSP-Phot estimates the stellar radii `astrophysical_parameters.radius_gspphot` and the distances `astrophysical_parameters.distance_gspphot`. We validate the ratio of twice the estimated radius to the estimated distance, $2R/d$, by comparing them with interferometric measurements of angular diameters. Figure 24 presents the excellent agreement with the samples from Boyajian et al. (2012a,b, 2013), Duvert (2016), and van Belle et al. (2021). We note that all of these targets are brighter than $G < 9.6$, and more than 90% of them have high-quality parallaxes with $\frac{\varpi}{\sigma_\varpi} > 20$. This means that the GSP-Phot results are very reliable (Andrae et al. 2023).

FLAME also provides radii estimates with a different approach based on the APs from either GSP-Phot or GSP-Spec combined with the *Gaia* photometry and parallaxes. The top panels in Fig. 25 compare `astrophysical_parameters.radius_flame` and `astrophysical_parameters.radius_flame_spec` with asteroseismic radii for giants from Pinsonneault et al. (2018). The agreement is at the 1% level with a scatter of 4%. Comparisons with other similar catalogues show agreement at the 1−2% level see further comparisons in the online documentation.

Bolometric luminosity. FLAME estimates the bolometric luminosities, L, using bolometric corrections based on GSP-Phot and GSP-Spec APs. We compared the L estimates with bolometric fluxes from Stevens et al. (2017). We selected a random subset of 90 000 main-sequence sources with *Gaia* DR3 parallaxes (panels from the second row in Fig. 25). We found that `astrophysical_parameters.lum_flame` and

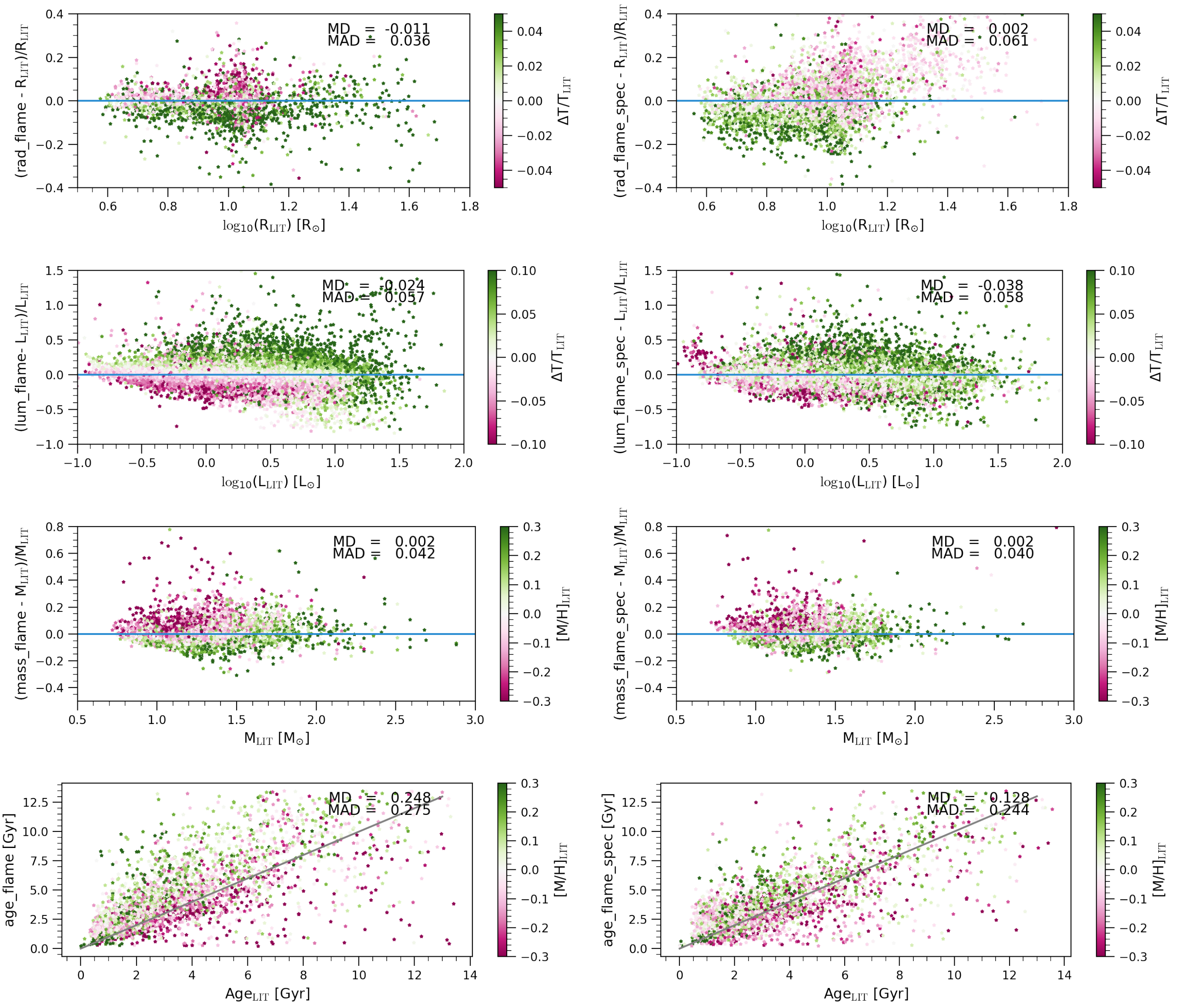

Fig. 25. Comparison of R, L, M, and age from FLAME to literature values. The left and right panels compares the estimates based on GSP-Phot and GSP-Spec from the `astrophysical_parameters` and `astrophysical_parameters_supp`, respectively. The top panel compares `radius_flame` and `radius_flame_spec` for giants with asteroseismic radii from Pinsonneault et al. (2018). The second panel compares main-sequence luminosities `lum_flame` and `lum_flame_spec` with those from Stevens et al. (2017) using a random selection of 90 000 stars. The third panel compares `mass_flame` and `mass_flame_spec` with masses from Casagrande et al. (2011), and the bottom panel compares `age_flame` and `age_flame_spec` from the same catalogue.

`astrophysical_parameters.lum_flame_spec` agree well with the literature with a median offset of 2–3% and a dispersion of about 5–6%. We also compared our estimates with other catalogues, such as Casagrande et al. (2011), with a median offset of +0.01 $L_{\odot}$ and similar dispersion.

Absolute magnitude M_G. Apsis provides two sets of absolute magnitudes: one from GSP-Phot obtained from the direct analysis of the BP and RP spectra and G magnitude (and parallax), and the other from FLAME if we use its luminosity L and the bolometric correction as follows: $M_G = 4.74 - 2.5\log_{10}(L/L_{\odot}) - \mathrm{BC}_G$. Figure 26 compares these two magnitude estimates. We find that most of the stars follow the bisector, indicating consistent results. However, we find a median absolute deviation of about 0.1 mag, and some artefacts. For instance, several vertical stripes appear (e.g. `mg_gspphot` = 3 mag), which might indicate anomalies due to GSP-Phot's models. In general, we find that FLAME tends to overestimate luminosity, leading to an underestimated M_G, when using parallaxes when the fractional uncertainties are about 15–20%. In contrast, we find a stronger agreement when FLAME uses `distance_gspphot` than when it uses `parallax` as a distance proxy, which is somewhat expected. `flags_flame` indicates the distances proxy that led to the luminosity estimates.

Gravitational redshift. FLAME produces another model-independent parameter, which is the gravitational redshift rv_{GR} (`astrophysical_parameters.gravredshift_flame` and `astrophysical_parameters_supp.gravredshift_flame_spec`). Typical values range from 0.05 to 0.8 km s^{-1}. Figure 27 compares `gravredshift_flame` and `gravredshift_flame_spec` We found a good consistency between the two flavors,

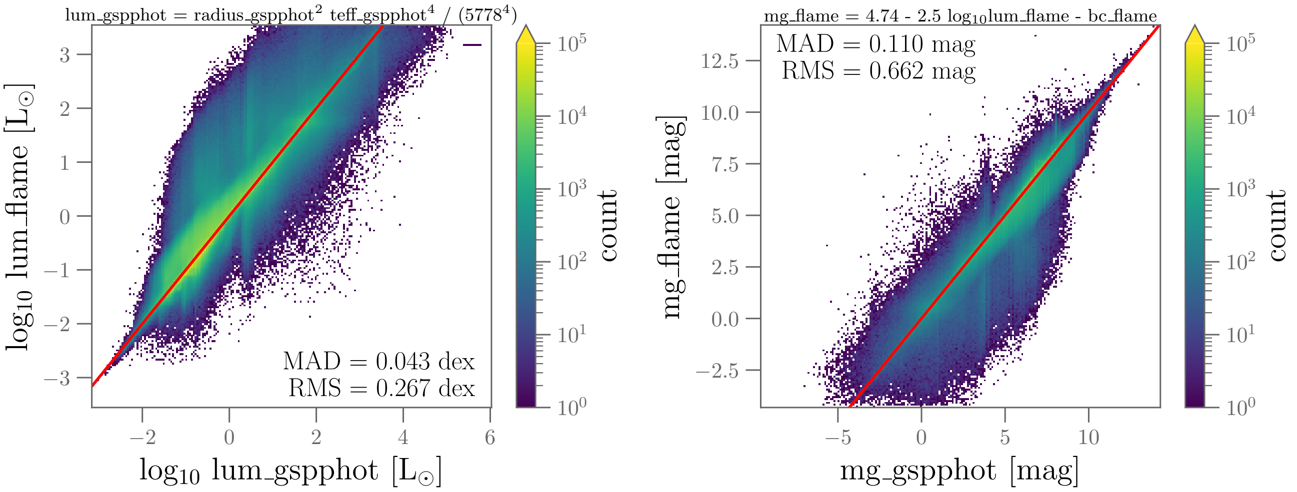

Fig. 26. Comparison of luminosities (left) and absolute magnitudes (right) from GSP-Phot and FLAME for all *Gaia* DR3 sources with estimates from both modules. Numbers quote the median absolute difference (MAD) and the root mean squared error (rms). We indicated the equations we used to construct the luminosities from GSP-Phot from the radius and temperatures, and the absolute magnitudes from FLAME from the luminosities and bolometric corrections.

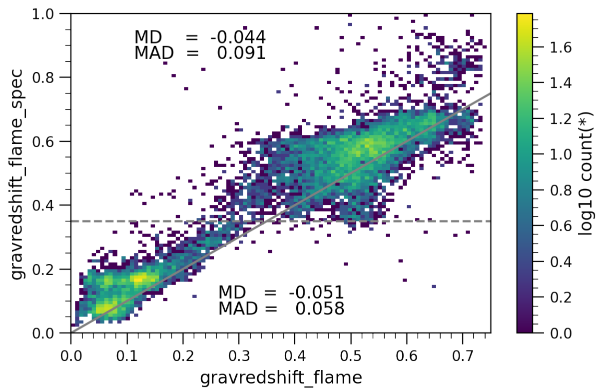

Fig. 27. Comparison between `gravredshift_flame` and `gravredshift_flame_spec` expressed in km s^{-1}. The bisector is indicated by the solid grey line. We indicate the mean offsets (MD) and absolute deviations (MAD) between the two clusters of values above and below 0.35 km s^{-1}. The differences reflect the different $\log g$ values from GSP-Spec and GSP-Phot along with the different radii derived by FLAME.

with median offset values of −0.05 km s^{-1}. This disagreement is a direct reflection of the different input data used to produce the value: $\log g$ and $T_{\rm eff}$ from GSP-Spec and GSP-Phot, and R from FLAME. Additionally, we selected solar-analogue stars from a random subset of 2 million stars from *Gaia* DR3, those for which GSP-Phot gave $T_{\rm eff}$ within 100 K, and $\log g$ within 0.2 dex of the solar values. This selection contained 46 667 stars, with a mean $rv_{\rm GR}$ of 588 ± 15 m s^{-1}, in agreement with the expected value of 600.4 ± 0.8 m s^{-1} for the Sun (Roca Cortés & Pallé 2014). We repeated this test for the GSP-Spec-based result, and we obtained a mean $rv_{\rm GR}$ of 590 ± 8 m s^{-1}. Although the second sample contained only 386 sources, we also found a good agreement with the known solar value.

3.3.2. Mass, age, and evolution stage

This section focuses on the most intrinsic evolution parameters: the mass M, age τ, and evolution stage ϵ. These are unique products of FLAME (with both GSP-Phot- and GSP-Spec-based flavors). These parameters are strongly model dependent as they directly relate to the stellar evolution models, here the BASTI models (Hidalgo et al. 2018). In addition, we emphasise that FLAME assumes solar metallicity when it estimates these parameters. Hence, we recommend using these estimates cautiously for stars with [M/H] < −0.5 dex.

Stellar masses. We compared the masses from FLAME with those from Casagrande et al. (2011) for main-sequence stars (see third panel in Fig. 25). Although we do not expect a significant influence, we note that Casagrande et al. (2011) also used the BASTI models in their analysis, but they used an older version from Pietrinferni et al. (2004). We find excellent agreement between the two estimates with an MAD of 0.002 $M_\odot$ with a scatter of 0.042 $M_\odot$. Overall, FLAME produces results that are comparable to literature results, with some outliers or disagreement with other catalogues that we traced back to the different input $T_{\rm eff}$ or $\log g$ estimates. In particular, one can reduce for giants these outliers when restricting the M estimates (`mass_flame`, `mass_flame_spec`) to only (i) $1.0 < M < 2.0\,M_\odot$ and (ii) $\tau > 1.5$ Gyr.

Stellar ages. Overall, we find an agreement between the ages from FLAME and the literature for non-evolved stars (i.e. main-sequence stars). The bottom panel of Fig. 25 compares the `astrophysical_parameters.age_flame` and `astrophysical_parameters_supp.age_flame_spec` with ages from Casagrande et al. (2011). In this comparison, we found a mean offset of about 0.1 to 0.3 Gyr with a dispersion about 0.25 Gyr. However, estimating ages for the giant stars reliably is more difficult because their ages strongly depend on their fitted mass. In addition, FLAME only relies on L and $T_{\rm eff}$ to obtain ages and masses, which has significant degeneracies. In addition, ages rely heavily on the solar abundance assumption in the FLAME processing.

One can trace most differences compared with the literature to the different input $T_{\rm eff}$ and L estimates. To support this statement, we compared FLAME ages to those we obtained with the SPinS public code (Lebreton & Reese 2020). We generated random sets of 600 stars with the SPinS code using the same *Gaia* DR3 APs that FLAME uses and compared the output

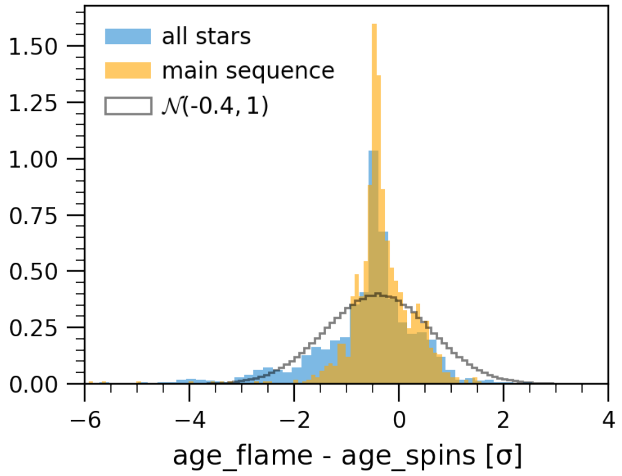

Fig. 28. Difference between `astrophysical_parameters.age_flame` with the age derived using the SPinS code normalised by their joint uncertainties. The Gaussian represents the ideal case but centred on the peak difference (-0.4σ) of the results using all stars irrespective of their evolutionary status. The input data are identical, and we assumed a solar-metallicity prior for both codes. We highlight the sample of MS stars discussed in Sect. 3.3.2.

ages in four different magnitude intervals. Figure 28 compares the estimates with `astrophysical_parameters.age_flame`. The agreement for the main-sequence stars is always at 1σ. The agreement is poorer for the evolved stars, but it remains within 3σ (see Creevey & Lebreton in prep., for more details).

Section 3.5 presents a further analysis of the masses and ages using clusters and further comparisons of mass and age with external data. We also present the analysis of the turn-off ages of some clusters in the online documentation; see online documentation.

Evolution stage. The ϵ parameter is an integer that takes values between 100 and 1300, representing the time step along a stellar evolutionary sequence. To first order, we tagged main-sequence stars with values between 100 and 420, subgiant stars with values between 420 and 490, and the giants above as defined in the BASTI models (Hidalgo et al. 2018). Figure 29 represents the evolution stage for members of four open star clusters (top panels; roughly solar metallicity) and four metal-poor globular clusters (bottom panels). We took the system members from Gaia Collaboration (2018b). These clusters were selected to contain a statistically significant number of stars in the three evolution phases estimated from FLAME. Overall, the main-sequence and giant evolution stages cover the expected colour-magnitude space. Although less numerous, the subgiant evolution stages are consistent with the expected colour-magnitude space. However, we also find discrepancies with expectations because the stellar models only cover the zero-age main sequence (ZAMS) to the tip of the red giant branch. The bottom panels in Fig. 29 clearly show horizontal giant branch stars that incorrectly labelled as main-sequence stars. Outside the ZAMS to the tip of the red giant branch phases, FLAME labels any star incorrectly. Again, the assumption of solar abundance in FLAME is challenged in these metal-poorer globular clusters.

As no other module produces M, age, or ϵ parameters, the only other method to assess their quality is to determine their consistency within other open clusters or wide binaries. We discuss this in Sect. 3.5.

3.4. Extinction, dust, and the ISM

When estimating the intrinsic stellar APs, it is also necessary to consider the effect of interstellar extinction on the observed SED, resulting in an estimation of the line-of-sight extinction for each star. We thus have extinction estimates from GSP-Phot, ESP-HS (for hot stars), and MSC (for double stars) as one of the spectroscopic parameters estimated from BP and RP spectra (A_0, A_G, $A_{\rm BP}$, $A_{\rm RP}$, and $E(G_{\rm BP}{-}G_{\rm RP})$). We also have an independent extinction estimate by GSP-Spec based on the analysis of the diffuse interstellar bands (DIB; see field details in Table D.6).

GSP-Phot. For all processed sources, GSP-Phot primarily estimates the monochromatic extinction A_0 at 541.4 nm (`astrophysical_parameters.azero_gspphot`) by fitting the observed BP and RP spectra, parallax, and apparent G magnitude. However, GSP-Phot also estimates the broadband extinctions A_G, $A_{\rm BP}$, and $A_{\rm RP}$, as well as $E(G_{\rm BP}{-}G_{\rm RP})$ obtained from the models (`astrophysical_parameters.ag_gspphot`, `abp_gspphot`, `arp_gspphot`, and `ebpminrp_gspphot` respectively). Extinction is a positive quantity, thus GSP-Phot imposes a non-negativity constraint on all estimates. Consequently, this can lead to a small systematic overestimation of extinction in truly low-extinction regions ($A_0 < 0.1$ mag)[14]. Andrae et al. (2023) demonstrated this effect for the Local Bubble where GSP-Phot estimates a mean extinction of $A_0 = 0.07$ mag instead of zero. A decreasing exponential approximates the distribution of GSP-Phot A_0 in the Local Bubble reasonably well, however, and it is also the maximum-entropy distribution of a non-negative random variate with a true value of zero. In other words, the exponential is equivalent to a Gaussian noise in more common contexts. Consequently, the standard deviation of this exponential distribution (identical to the mean value) provides an error estimate for A_0 of 0.07 mag. Similarly, Andrae et al. (2023) reported similar values of 0.07 mag for $A_{\rm BP}$, 0.06 mag for A_G, and 0.05 mag for $A_{\rm RP}$ within the Local Bubble. These values are in agreement with Leike & Enßlin (2019), who reported 0.02 mag. While one could allow small values of negative extinctions such that results for low-extinction stars may scatter symmetrically around zero, Andrae et al. (2023) showed that this is not sufficient in the case of StarHorse2021 (Anders et al. 2023), whose `av50` in the Local Bubble peaks around 0.2 mag twice as much as GSP-Phot. We found that StarHorse2021 extinction `av50` estimates appear globally higher than A_0 from GSP-Phot by 0.1 mag, which is likely a bias in the StarHorse2021 catalogue (see Anders et al. 2023, their Fig. 15). Andrae et al. (2023) also observed that in high-extinction regions `av50` can become significantly higher than A_0. It is currently unclear whether this is an overestimation by StarHorse2021 or an underestimation by GSP-Phot (or both).

Using solar-like stars, Gaia Collaboration (2023f) investigated the $G_{\rm BP}{-}W_2$ colour, which uses the *Gaia* and AllWise passbands for two reasons: (i) a colour is a quantity independent of distance, and (ii) as the extinction in the AllWISE W_2 band is negligible, we can safely associate any correlation with $G_{\rm BP}$ (i.e., a proxy for $A_{\rm BP}$). We find that the $G_{\rm BP}{-}W_2$ colour closely agrees with a linear trend with the GSP-Phot $A_{\rm BP}$ estimate to within 0.087 mag rms scatter, which is consistent with the 0.07 mag obtained for $A_{\rm BP}$ in the Local Bubble. We also found that the linear relation holds from the low-extinction to high-extinctions regimes. Additionally, Fig. 30 also shows good agreement of our A_0 estimates with our expectations in open clusters with only a mild overestimation of ~0.1 mag (see Sect. 3.5.1).

---

[14] The mean or median of a positive distribution is always strictly positive, but never null.

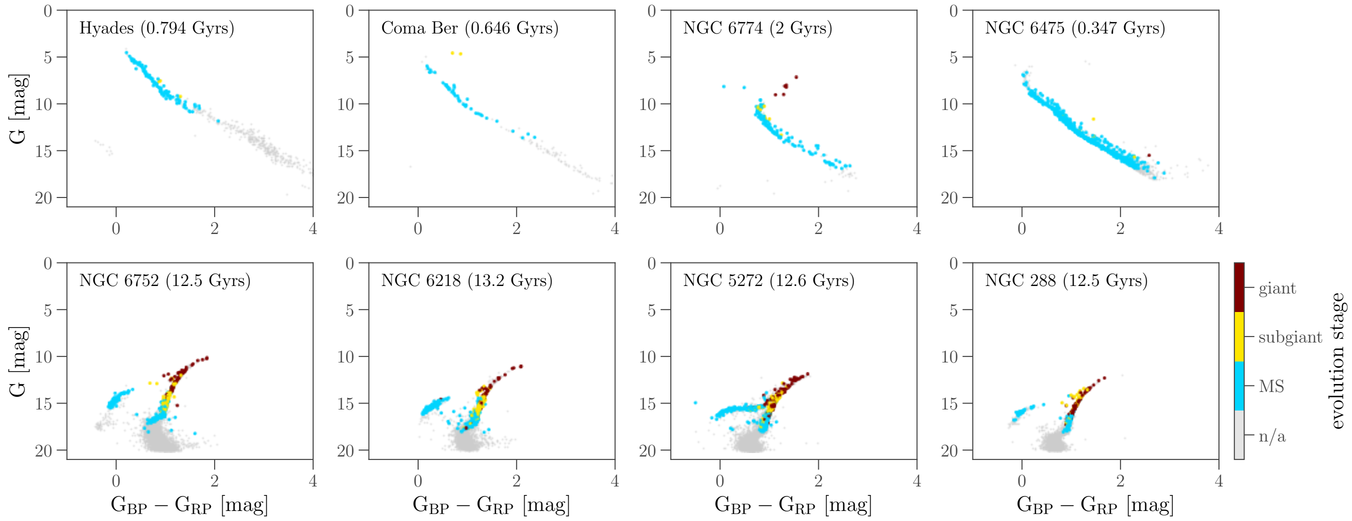

Fig. 29. Evolution stage distribution from FLAME in the CMDs of some selected clusters. The top panels represent the CMDs of four open clusters near solar metallicity, and the bottom panels are those of four low-metallicity globular clusters. Gaia Collaboration (2018b) provided the cluster members. We parsed the values of `astrophysical_parameters.evolstage_flame` into the three stages: main-sequence, subgiant, and giant colour-coded according to the scale on the right-hand side. We also indicate other members without phase estimates in grey.

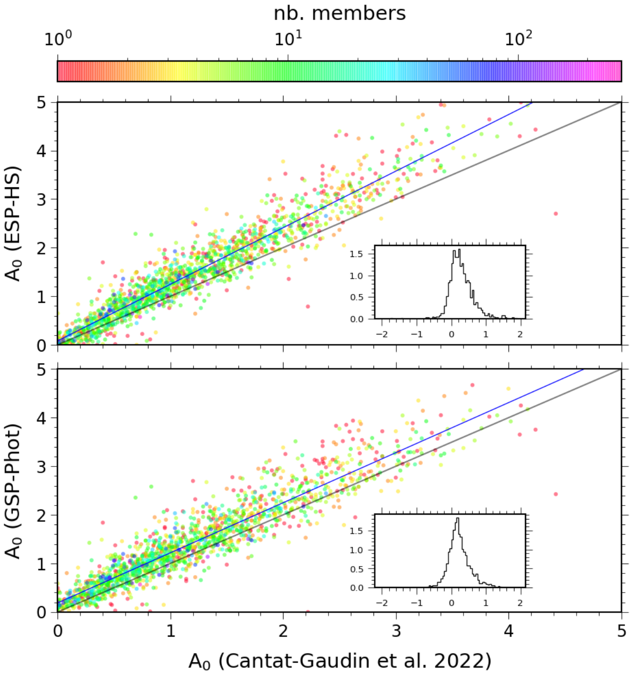

Fig. 30. Comparison of ESP-HS (top) and GSP-Phot (bottom) extinction estimates A_0 for hot stars in star clusters. We used the cluster members and mean extinctions from Cantat-Gaudin et al. (2020). We computed the GSP-Phot and ESP-HS median estimates using only stars with $T_{\mathrm{eff}} > 7500$ K. We colour-coded the data by the number of hot star members with estimates we found in the cluster (with respect to the colour bar at the top). In both panels, the grey lines represent the identity relation, and the blue lines a linear regression through the data points. The insets show the normalised distribution of the differences, A_0(GSP-Phot or ESP-HS)–A_0(literature).

TGE. GSP-Phot also provides the A_0 estimates used by TGE to produce an all-sky (two-dimensional) map of the total Galactic extinction, meaning the cumulative amount of extinction in front of objects beyond the edge of our Galaxy. TGE selects giant star tracers at the edge of the Milky Way, more specifically, stars with `gaia_source.classprob_dsc_combmod_star` > 0.5, `gaia_source.teff_gspphot` between 3000 and 5700 K, `gaia_source.ag_gspphot` between −10 and 4 mag, and distances from the Galactic plane beyond 300 pc using the `gaia_source.distance_gspphot`. When they are selected, TGE groups the tracers per HEALpix with levels adapted from 6 (∼0.08 deg^2) to 9 (∼0.01 deg^2) to have at least three stars per group. Finally, TGE estimates A_0 from the median and standard deviation of the ensemble of `gaia_source.azero_gspphot` values per defined HEALpix. We emphasise that TGE provides two tables: `total_galactic_extinction_map`, which contains the map with a variable HEALpix resolution (`healpix_level`) and `total_galactic_extinction_map_opt`, which contains the resampled information at HEALpix level 9. It is important to remark that TGE primarily uses `gaia_source.azero_gspphot`, which contains estimates with a mixture of atmosphere libraries, so-called best-fit estimates. Figure 31 compares the TGE estimates to those of GSP-Phot for the MARCS and PHOENIX atmosphere libraries providing APs for the giant stars. Although one could expect some AP variations from a set of atmosphere models to another, we find statistically no significant differences between the two libraries and TGE estimates. The high dispersion along the y-axis mostly reflects the low number of stars beyond 16 kpc from the Galactic center, especially with high extinction values. Delchambre et al. (2023) provided a more detailed description of the method and performance assessment of the TGE maps, especially comparisons with non-stellar tracers (e.g. Planck).

ESP-HS. For hot stars with $G < 17.65$ mag, ESP-HS also estimates A_0 by fitting the observed BP and RP spectra (`azero_esphs`). Like GSP-Phot, ESP-HS also provides A_G, and $E(G_{\mathrm{BP}}{-}G_{\mathrm{RP}})$. We compared the extinction A_0 from GSP-Phot and ESP-HS using star clusters for the hotter stars (Fig. 30). Both modules find consistent A_0 estimates when deriving extinctions higher than 0.3 mag. However, over this hot star sample, we find that GSP-Phot tend to constantly overestimate extinction by about 0.1 mag, and ESP-HS overestimate by a factor 1.2. Overall,

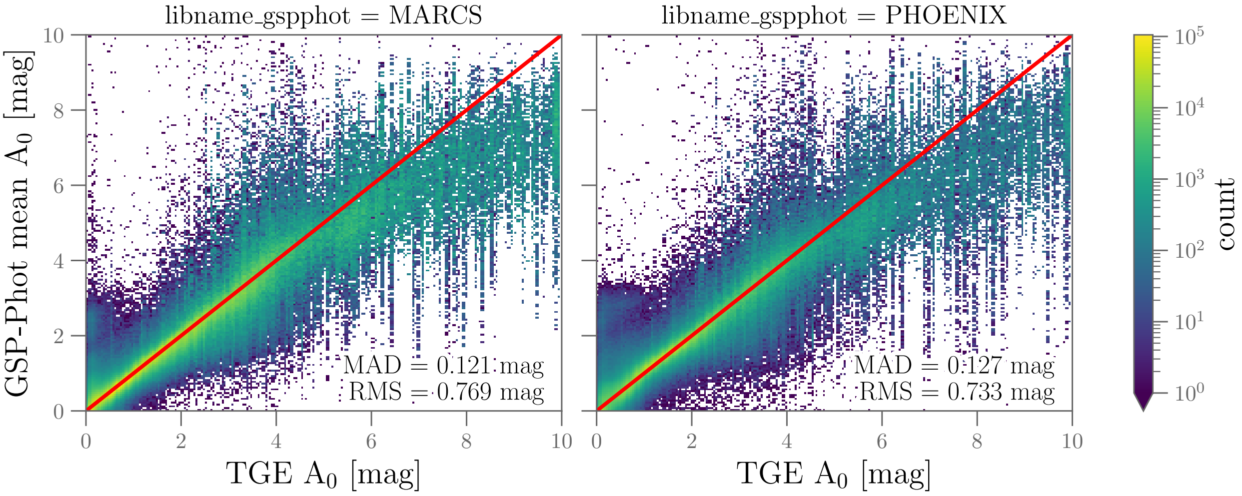

Fig. 31. Comparison of TGE and GSP-Phot extinction estimates A_0 limited to giant stars. We calculated the mean extinction `astrophysical_parameters.azero_gspphot` per healpix level 9 to compare to the TGE optimised map `total_galactic_extinction_map.a0`. We partially included the TGE tracer selection: $3000 <$ `teff_gspphot` < 5700 K and $-10 <$ `mg_gspphot` < 4 mag (we did not filter out based on distances). This represents 21 244 458 and 9 271 775 stars for the MARCS and PHOENIX library, respectively.

for all stars with GSP-Phot and ESP-HS estimates, we found a MAD of 0.120 mag, and rms of 0.380 mag. However, we emphasise that these differences, especially the rms statistics, also vary with the spectral libraries (`gaia_source.libname_gspphot` or `astrophysical_parameters.libname_gspphot`). When we restrict the comparison to the OB star library that best describes this temperature regime, we find an improved rms of 0.170 mag. This illustrates the importance of choosing or exploring which spectral library is appropriate for the sources of interest.

MSC. MSC also estimates the A_0 parameter by assuming that the BP and RP spectra represent a composite of an unresolved binary: two blended coeval stars at the same distance (`azero_msc`). The performance of MSC is similar to that of GSP-Phot (see Sect. 3.5.2).

GSP-Spec-DIBs. In addition to the stellar APs, GSP-Spec estimated the equivalent width of diffuse interstellar bands (DIBs) in the RVS spectra for 476 117 stars in *Gaia* DR3. The DIB spectral feature arises from largely unidentified molecules that are ubiquitously present in the interstellar medium (ISM). GSP-Spec measures the DIB profile parameters: the equivalent width (`astrophysical_parameters.dibew_gspspec`) and characteristic central wavelength (`astrophysical_parameters.dib_gspspec_lambda`) using a Gaussian profile fit for cool stars and a Gaussian process for hot stars. We described in detail the DIB measurements procedure in Recio-Blanco et al. (2023, Sect. 6.5) and further assessed their performance in Gaia Collaboration (2023c). We emphasise that one should restrict themselves to using the DIB estimates with quality flags `astrophysical_parameters.dibqf_gspspec` ≤ 2 (definition in Table 2 of Gaia Collaboration 2023c). Although one can question the standard analysis in this field, we applied the approach to compare our results with the literature. We estimated a linear relation between `dibew_gspspec` and `ebpminrp_gspphot` as

$$E(G_{\rm BP}{-}G_{\rm RP}) = 4.508(\pm 0.137) \times {\rm EW}_{862} - 0.027(\pm 0.047). \quad (2)$$

We identified the strong outliers to this relation as having an overestimated $E(G_{\rm BP}{-}G_{\rm RP})$ from GSP-Phot (linked to an incorrect temperature estimate; see Gaia Collaboration 2023c). GSP-Spec also measured DIBs for hot stars ($T_{\rm eff} > 7500$ K), providing us with a total of 1142 high-quality DIB measurements. We compared them with the extinction estimates from ESP-HS (`astrophysical_parameters.ebpminrp_esphs`) and found an excellent agreement with the relation we obtained above (see Fig. 9 of Gaia Collaboration 2023c). We further compared the DIB EW with the A_0 values of the TGE HEALPix level 5 map (`total_galactic_extinction_map`), where we found a strong linear correlation given by ${\rm EW} = 0.07 \times A_0 + 0.03$ up to $A_0 \sim 1.5$ mag, after which we found a shallower trend. We suspect that the slope change originates from TGE providing a total extinction far beyond the distance of stars with DIB λ862 measurements.

Finally, we estimated the standard quantity $E(B{-}V)$/EW of 3.105 ± 0.048. This lies in the range of the derived ratios in the literature (compilation in Table 3 in Gaia Collaboration 2023c).

3.5. Groups of stars

3.5.1. Clusters

Star clusters are very effective in assessing the qualities of stellar parameters, as proven in previous *Gaia* data releases. Open star clusters are coeval populations: they have the same age, same metallicity, about the same extinction, and distance.

Apsis processed all the stars independently, and, in particular, did not exploit the coevolution of stars. This section presents some key results concerning the global quality of the APs in star clusters. We provide additional validation, known issues, some calibration relations, and the optimal use of the quality flags in Andrae et al. (2023), Recio-Blanco et al. (2023) and Babusiaux et al. (2023).

We selected a sample of star clusters from the Cantat-Gaudin et al. (2020) catalogue. Gaia Collaboration (2023d) refined the cluster memberships using *Gaia* eDR3 astrometry. Our selection corresponds to about 230 000 stars: the number of stars per cluster varies significantly from 40 to more than 700, with an average of ~60 stars. Open clusters mostly contain main-sequence stars with a median $G = 15.6$ mag, but their populations vary significantly with the ages of the systems. We approximated the stellar population of each cluster by an isochrone to obtain reference estimates for $T_{\rm eff}$, $\log g$, mass, age, and distance. Additionally, we assumed homogeneity throughout the colour-magnitude diagram of A_0 and [M/H]. For the former, we avoided regions in which differential extinction is more likely to be present by excluding clusters younger than 100 Myr from our samples. We used the PARSEC isochrones[15] for this purpose, associated with the age, distance, extinction, and metallicity of the clusters from our literature catalogue. We summarise the statistical analysis of the accuracy of the relevant APs over the cluster members below.

We compare the atmospheric and evolution APs from GSP-Phot, GSP-Spec, and FLAME to the cluster isochrones. We emphasise that when analysing the GSP-Spec results, we selected stars with `astrophysical_parameters.flags_gspspec` with `f1,f2,f4,f5,f8` = 0. Table 4 presents the median and MAD of the residuals to the isochrones for $T_{\rm eff}$, $\log g$, A_G, M, and τ derived by GSP-Phot, GSP-Spec, and FLAME. We note that we compared A_G and τ with the literature values independently of the isochrones.

GSP-Phot. We found that $T_{\rm eff}$, $\log g$, A_G from GSP-Phot are in general agreement with expectations, but sometimes show high dispersions. It is important to note that we analysed the best library estimates (e.g. `astrophysical_parameters.teff_gspphot`), but the results may vary with different choices of library (e.g. `astrophysical_parameters_supp.teff_gspphot_marcs`). GSP-Phot performs better for $G < 16$ mag, where the S/N of the BP and RP spectra remains high ($S/N > 100$). Figure 32 illustrates our analysis with the example of Messier 67 (also known as NGC 2682). In this cluster, we found 4% of outliers defined as $\Delta T_{\rm eff}/T_{\rm eff} > 0.5$. But this fraction varies across the entire *Gaia* DR3 sample. Overall, we identified that GSP-Phot overestimated $T_{\rm eff}$ values for giants and underestimated them for supergiants (see Fig. 33). In detail, we find that the distribution of the GSP-Phot $\log g$ values has a long tail towards overestimating values on the main sequence. Still, in contrast, GSP-Phot underestimates gravity for hot stars and giants. We also note the issue with metallicity and the extinction estimates reported in Sect. 3.2.1. Messier 67 is at ~850 pc from us, a close distance that GSP-Phot a priori assumes to be mostly free of extinction. This prior leads to an underestimation of the reddening of these stars. As a result of preserving the observed stellar SEDs, GSP-Phot underestimate [M/H]. Andrae et al. (2023) discussed this extinction-distance prior and related issues in detail.

GSP-Spec. We also analysed the GSP-Spec APs and found that $\log g$ from GSP-Spec could show biases up to −0.3 dex compared to isochrone predictions (similarly to Sect. 3.2.1). In particular, we found a significant underestimation for hot stars, and we caution the user against using the GSP-Spec $\log g$ values for AGBs as we find them of poorer quality. We refer to Recio-Blanco et al. (2023) for the details and especially emphasise that these comparison results depend strongly on the quality flag selection. Recio-Blanco et al. (2023) also encouraged the user to define calibration relations for their specific uses.

FLAME. We also found that the FLAME APs are in good agreement when we restricted our analysis to the best-measured stars, those with `astrophysical_parameters.flags_flame` = 00,01. The fact that FLAME assumes solar metal metallicity produced poor τ and M estimates in low-metallicity clusters, unsurprisingly. However, in the solar metallicity regime, M is in good agreement with expectations (see Table 4). It also seems that FLAME overestimated τ for young stars and underestimated it for old stars, with the most significant discrepancies with the literature appearing for cool main-sequence stars.

Table 4. Parameter residuals to reference values in star clusters.

Parameter	Module	$M^{(1)}$	MAD$^{(2)}$	Units
$T_{\rm eff}$	GSP-Phot	34	400	K
$T_{\rm eff}$	GSP-Spec	6	160	K
$\log g$	GSP-Phot	0.01	0.22	dex
$\log g$	GSP-Spec	−0.30	0.44	dex
A_G	GSP-Phot	0.12	0.10	mag
M (spec)	FLAME	−0.02	0.10	$M_\odot$
M (phot)	FLAME	−0.11	0.14	$M_\odot$
τ,(spec)	FLAME	−0.40	0.60	Gyr
τ,(phot)	FLAME	0.30	0.40	Gyr

Notes. $^{(1)}$Median estimates of the residuals. $^{(2)}$Mean absolute deviation (MAD) of the residuals. `flags_gspspec` with `f1,f2,f4,f5,f8` = 0.

Using star clusters also has the advantage of assessing whether the reported uncertainties are of the correct order overall. FLAME reported underestimated uncertainties on M and τ derived either from GSP-Spec or GSP-Phot APs. Figure 34 demonstrates that the M residuals between GSP-Phot, and the isochrones disperse significantly more than the uncertainties (of the size of the symbols on average).

ESP-HS. A fraction of the OBA stellar population in open clusters went through the analysis by the ESP-HS module. Figure 35 illustrates a selection of cluster Kiel diagrams at different ages. By comparison to the PARSEC isochrones, we found estimates commonly to the right of the isochrones in the Kiel diagrams, suggesting somewhat older cluster ages than the literature references. These findings may relate to a systematic underestimation of $T_{\rm eff}$ and $\log g$. Although unlikely, the literature may underestimate the cluster ages. Still, more likely, our results may be affected by gravitational darkening due to axial rotation on the spectral energy distribution of OBA stars.

ESP-UCD. As we detailed in the online documentation, ESP-UCD detects significant overdensities at the positions of several clusters and star-forming regions. We used the BANYAN Σ (Gagné et al. 2018) to identify UCD members of nearby young associations within 150 pc from the Sun. Table C.1 contains the number of sources with a membership probability higher than 0.5 in each association and the effective temperature of the coolest UCD. We also include entries for associations beyond 150 pc derived from our clustering analysis using the OPTICS algorithm (Ankerst et al. 1999) in the space of Galactic coordinates, proper motions, and parallax. We did not use these stars to assess the performance of ESP-UCD, but we report our strong UCD candidates.

3.5.2. Unresolved binaries

In Apsis, the MSC module aims to distinguish between the two components of binaries by analysing their composite BP/RP spectra. It assumes that these sources are blended coeval stars (same distance, extinction, and metallicity). We were unable to create sufficiently high-quality synthetic models of BP and RP spectra of unresolved binaries; they did not fully model the instrumental (and data reduction) effects for these sources.

---

[15] PARSEC isochrones are available from `http://stev.oapd.inaf.it/cgi-bin/cmd`

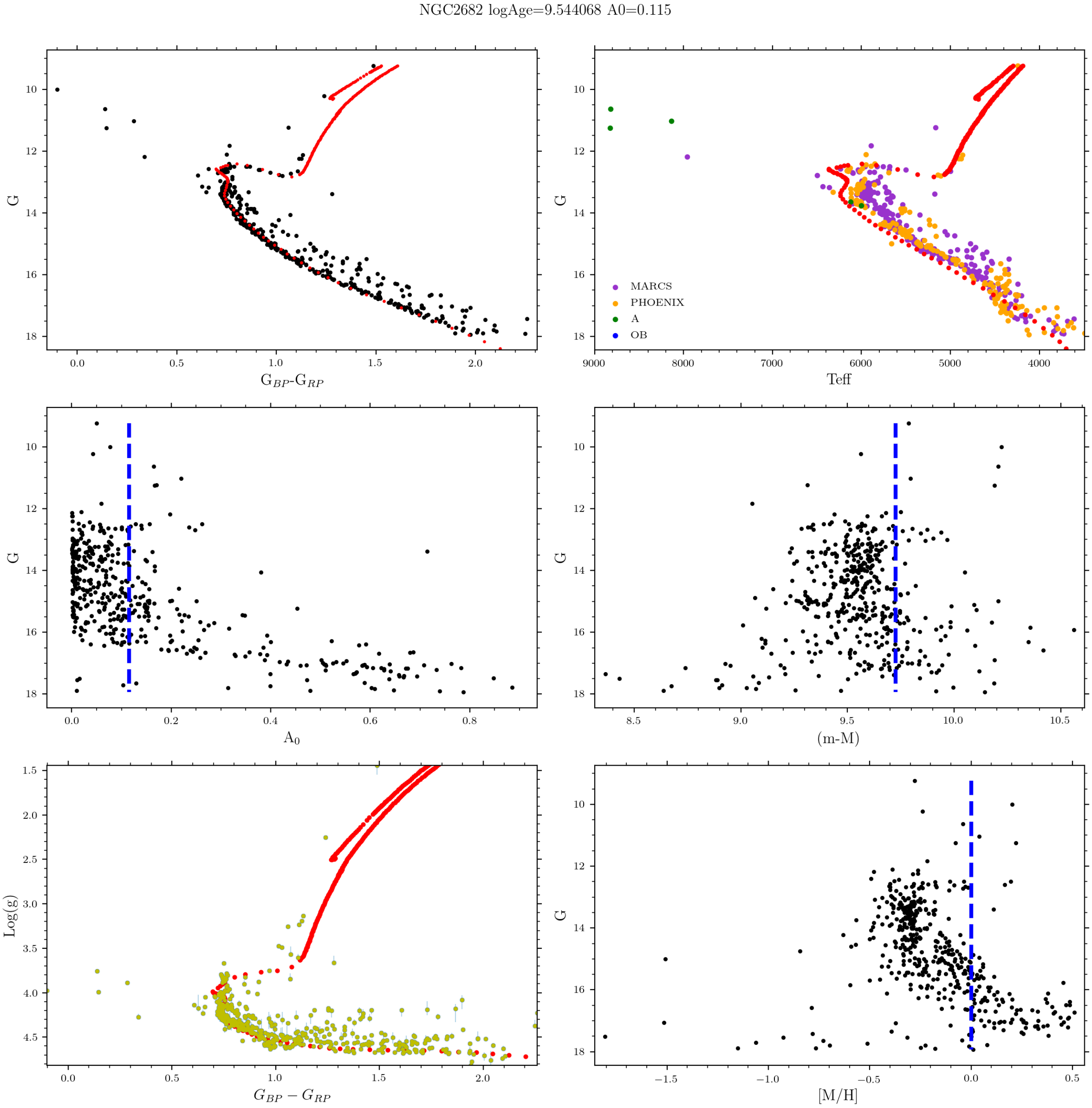

Fig. 32. Illustration of the performances of GSP-Phot in cluster Messier 67. The PARSEC isochrone (indicated by the red dots in the panels) indicates the reference of $\log_{10}(\mathrm{Age/yr}) = 9.5$, $A_0 = 0.11$ mag, and solar metallicity. We also indicate with the vertical dashed blue lines the reference A_0, distance modulus, and [M/H]. Top left panel: *Gaia* CMD of M 67. Top right panel: G vs. `gaia_source.teff_gspphot`. The colours indicate the different best libraries (`gaia_source.libname_gspphot`). Middle left panel: G vs. `gaia_source.azero_gspphot`. Middle right panel: G vs. distance modulus derived from `gaia_source.distance_gspphot`. Bottom left panel: `gaia_source.logg_gspphot` vs. G_{BP}–G_{RP} (yellow dots). Bottom right panel: G vs. `gaia_source.mh_gspphot`.

Instead, MSC implements an empirical set of models constructed from observed BP and RP spectra of spectroscopic binary stars (see Creevey et al. 2023, for details). As a result of the limited number of unresolved binaries for reference with APs, MSC adopted a strong [M/H] prior centred on solar values.

MSC analysed all sources with $G < 18.25$ mag and therefore inherently analyses single stars as well (assuming a binary source). Similarly, GSP-Phot takes all sources to be single stars. As internally MSC operates very similarly to GSP-Phot, we can compare their overlapping results more robustly than any other Apsis module. Figure 36 compares APs from MSC and GSP-Phot parameters with those from the binary sample of El-Badry et al. (2018). We find an expected negative bias in temperature and $\log g$ from GSP-Phot because it assumed that these sources are single stars. They correspond to a luminosity-weighted average between the primary and the secondary. Commonly, this leads to a lower T_{eff} and $\log g$ to reach the observed brightness of the binary system with a single star. We find that despite its strong solar metallicity prior, the posterior of [M/H] from MSC are broad. Overall MSC performs better than GSP-Phot on this particular sample of binaries.

The GALAH survey (Martell et al. 2017) provides another set of 11 263 spectroscopic binaries (Traven et al. 2020) with a component flux ratio lower than 5 (i.e. within the MSC parameter ranges). As above, we compared MSC with GSP-Phot on this sample, and we find that their APs

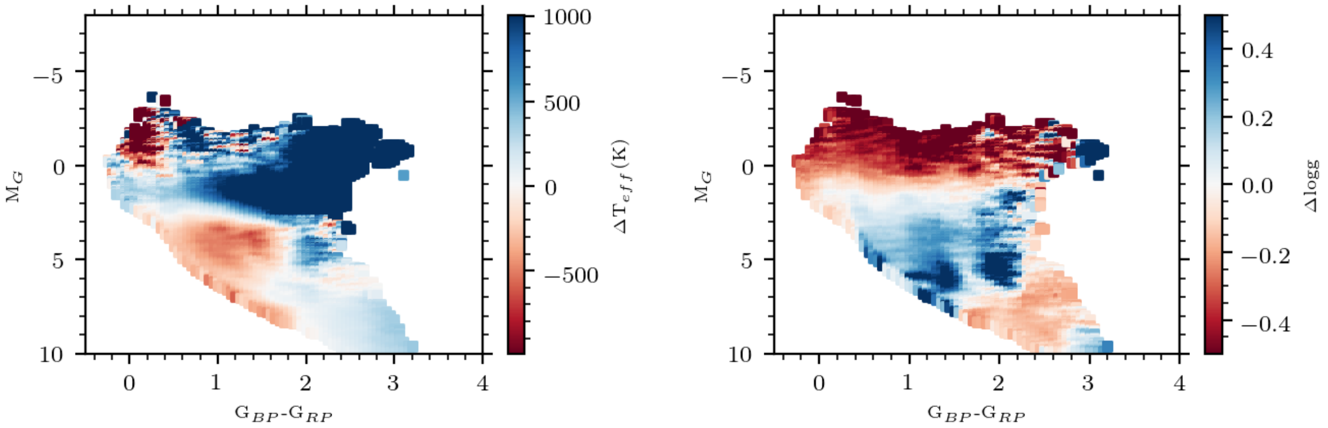

Fig. 33. Residuals of $T_{\rm eff}$ and $\log g$ estimates from GSP-Phot to isochrones of star clusters for the sample described in Sect. 3.5.1. We show the mean residuals of the members as a function of position in the M_G vs. ($G_{\rm BP}$–$G_{\rm RP}$) diagram. The y-axis is corrected for extinction and distance modulus using literature values. The colour indicates $\Delta(T_{\rm eff})$ and $\Delta(\log g)$ on the left and right panels, respectively.

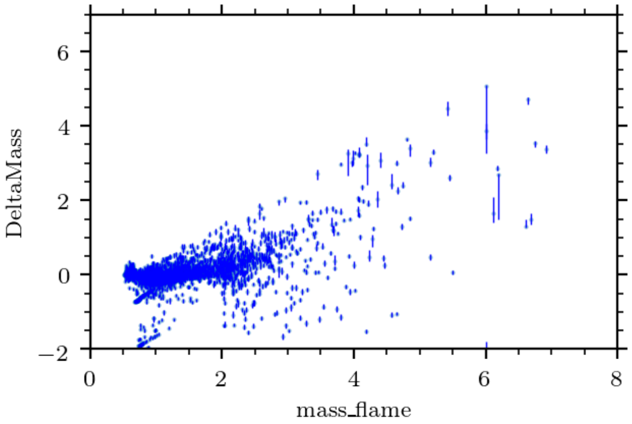

Fig. 34. Residuals, Delta(Mass) in $M_\odot$, between `mass_flame` and the isochrone predictions for star clusters taken from Cantat-Gaudin et al. (2020). We selected estimates with `flags_flame` = 00. Error bars indicate the uncertainties from FLAME.

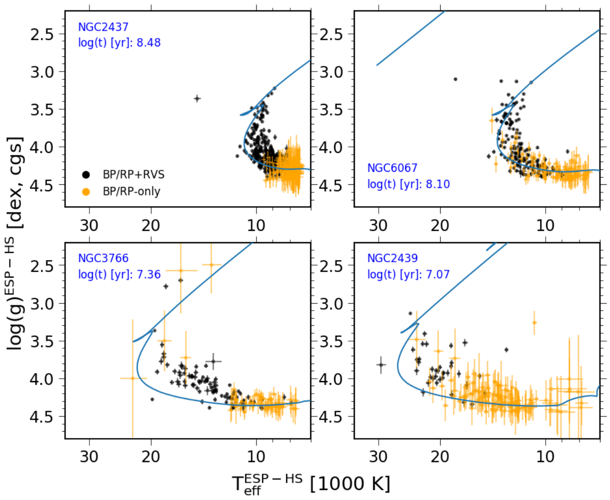

Fig. 35. Cluster Kiel diagrams of ESP-HS astrophysical parameters. The PARSEC isochrones (blue) correspond to the cluster age provided by Cantat-Gaudin et al. (2020) assuming solar metallicity. Estimates obtained in both ESP-HS processing modes are shown by black and orange disks with their corresponding uncertainties.

have comparable accuracies. Figure 37 compares the seven APs from MSC with those from GALAH. We note that the colour-coding in the plots indicates the goodness-of-fit (using `astrophysical_parameters.logposterior_msc`) rather than a source density. Except for A_0, the goodness-of-fit is best around the identity line. This behaviour confirms that MSC fits the composite spectra of binaries well when the MCMC procedure converges. The goodness-of-fit also indicates that MSC did not converge properly for many sources.

We can flag poor convergence as sources with low `logposterior_msc` values. Finding a unique threshold for all science applications is challenging. However, Table 5 provides the evolution of the residual statistics with the GALAH sample when changing the goodness-of-fit threshold. By construction, the residuals and the overall biases improve as the threshold increases, but we remove a significant number of sources from the sample. Regardless of this filtering, MSC tends to overestimate $\log g_1$, $\log g_2$, and [M/H] for the GALAH sample. We suspect that the prior of MSC favors solar metallicity leads to overestimating [M/H]. As a result, to match the BP and RP spectra, MSC compensates the high [M/H] by decreasing the intrinsic luminosity, requiring higher $\log g$ values. However, we cannot exclude a biases in the GALAH data, as suggested by the fact that the GALAH catalogue provides significantly lower [M/H] for the binaries than for their single stars (Traven et al. 2020, Sect. 8.3). This open issue is also supported by the discrepancies with APs reported by the APOGEE binary sample (El-Badry et al. 2018), with 26 sources in common.

We also found chemically homogeneous spectroscopic parameters from *Gaia* for the components of wide binaries when compared with high-resolution data from Hawkins et al. (2020). In their sample of 25 wide binaries, 20 had a metallicity difference lower than 0.05 dex, while the remaining 5 showed deviations of ~0.1 dex. From Table 3 of Hawkins et al. (2020), we selected the 20 homogeneous binaries (excluding WB02, WB05, WB09, WB16, and WB21) and compared the metallicities from Apsis for each of the two components[16]. We did not apply any calibration to the data. These are dwarf stars with $T_{\rm eff}$ between 5000 and 6400 K and

---

[16] The *Gaia* DR2 source IDs listed in Table 3 of Hawkins et al. (2020) are the same as the *Gaia* DR3 source IDs, except for WB13B, which has DR3 source ID 3230677874682668672.

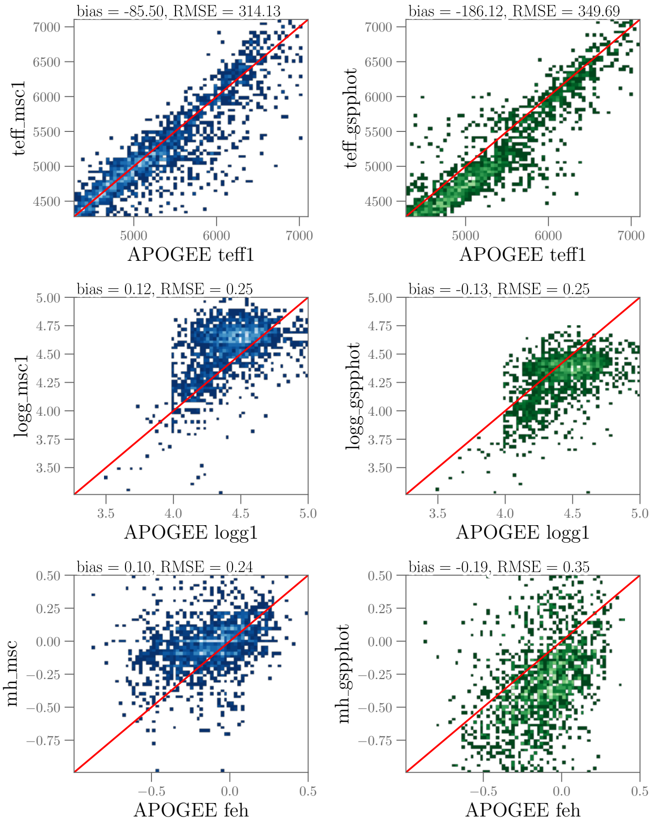

Fig. 36. MSC and GSP-Phot inferred values on the y-axis vs. APOGEE (El-Badry et al. 2018) literature values on the x-axis for sources with a common parameter range (including a flux ratio lower than 5). The three parameters $T_{\rm eff}$, $\log g$, and [M/H] with their respective 1:1 line are shown in orange. We applied the GSP-Phot postprocessing, a cut on `fidelity_v2` > 0.5 (Rybizki et al. 2022), and a cut on `logposterior_msc` > −1000.

metallicities above −0.8 dex. For 16 out of the 20 homogeneous binaries according to Hawkins et al. (2020), the metallicities from GSP-Phot (`astrophysical_parameters.mh_gspphot`) agree within 0.15 dex. For the remaining 4 binaries, they deviate by 0.2 to 0.3 dex (WB08, WB13, WB18, and WB22). Eighteen of the 20 binaries have metallicity determinations from GSP-Spec (`astrophysical_parameters.mh_gspspec`) for both components, and all except 2 agree within 0.15 dex. The exceptions are WB14 with a difference of 0.16 dex, and WB15 with a difference of 0.5 dex. WB15 also has a difference in $\log g$ (`astrophysical_parameters.logg_gspspec`) of 1.1 dex, whereas the two components should have equal surface gravity according to Hawkins et al. (2020). This indicates that the *Gaia* metallicities are reliable (at least in a statistical sense) in the parameter space covered by the binary sample.

We further explored the possibility of cleaning the MSC results by excluding sources with possible spurious astrometric solutions. It is expected that *Gaia* astrometry may be affected by binarity. We applied the method from Rybizki et al. (2022), and we kept sources with `fidelity_v2` > 0.5. After this selection, the GALAH sample shrank from 11 263 to 9836 sources. The rms for the distance comparison improved from 617 to 429 pc, and its bias from −184 to −157 pc (when we assumed the inverse parallax as the true distance). It also improved the statistics of the other parameters and overall the agreement with GSP-Phot APs.

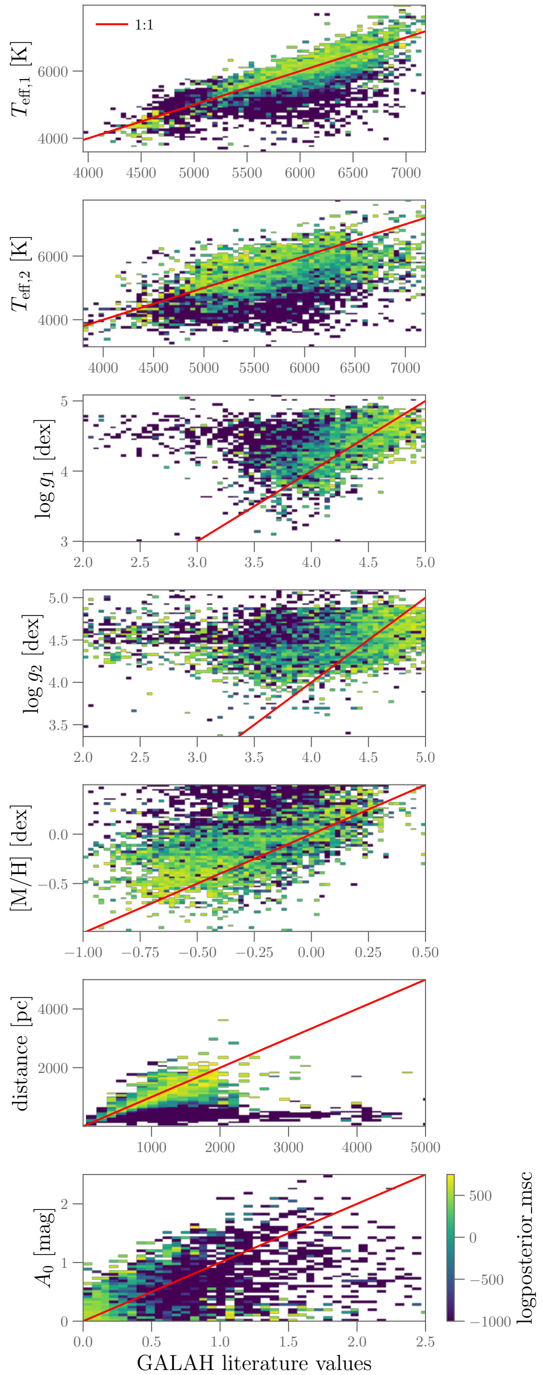

Fig. 37. Comparison of MSC APs with the GALAH catalogue for 11 263 binary stars. From top to bottom, we compare the seven parameters inferred by MSC on the y-axes with the GALAH literature values on the x-axes: $T_{\rm eff}$ and $\log g$ for the two components of the system, [M/H], distance, and A_0 (Table D.2 lists the corresponding catalogue field names). In each panel, we indicate the 1:1 line for reference, and the colour corresponds to the average `astrophysical_parameters.logposterior_msc` of all stars per bin. We provide associated statistics in Table 5.

Overall, the performance of MSC remains challenging to estimate. Only a few reference catalogues exist, and they rarely provide statistically significant samples (many thousands) with APs. In addition, one needs to use the astrometric measurements of binary systems with caution. We expect *Gaia* DR4 to provide a significant improvement in the future.

Table 5. MSC versus GALAH sample bias and rms comparison for different `logposterior_msc` percentile cut-offs.

Percentiles count	0 11 263	5 10 699	16 9461	50 537	84 1814	95 567
Parameters ↓			Sample rms			
$T_{\rm eff,1}$ [K]	387	348	273	192	144	135
$T_{\rm eff,2}$ [K]	632	592	536	417	310	258
$\log g_1$ [dex]	0.40	0.35	0.33	0.29	0.25	0.24
$\log g_2$ [dex]	0.58	0.54	0.50	0.45	0.38	0.36
[M/H] [dex]	0.30	0.29	0.27	0.24	0.22	0.21
Distance [pc]	617	553	277	152	47	25
A_0 [mag]	0.27	0.24	0.21	0.19	0.15	0.13
			Sample bias			
$T_{\rm eff,1}$ [K]	–139	–118	–72	–6	21	10
$T_{\rm eff,2}$ [K]	–418	–392	–350	–245	–144	–60
$\log g_1$ [dex]	0.24	0.22	0.20	0.17	0.13	0.12
$\log g_2$ [dex]	0.35	0.33	0.30	0.24	0.17	0.15
[M/H] [dex]	0.21	0.20	0.19	0.19	0.19	0.18
Distance [pc]	–184	–148	–95	–49	–16	–9
A_0 [mag]	–0.01	0.00	0.01	0.02	0.01	0.01

3.6. Identification and analysis of peculiar cases (outliers)

Galactic sources dominate the content of *Gaia* DR3. Sources with BP and RP spectra are essentially intermediate-mass stars of FGK spectral types with $G < 17.65$ mag, with the addition of a set of UCDs and extragalactic objects (see Fig. 1). Outliers in this context mean objects that are not similarly consistent with the rest of the sample. The similarity in this context relates to the distance metric implemented in the clustering algorithm in the OA module summarised below.

On the one hand, Apsis provides multiple classifications and flags that one can use to identify outliers (see Table D.3). For instance, one can remove stars with emission lines using ESP-ELS parameters, or one can generate a pure sample of solar analogues by combining APs and flags from GSP-Phot and GSP-Spec (see Gaia Collaboration 2023f, and other examples herein). However, these derive from supervised classifications and comparisons against models, which limits discoveries of peculiar objects.

On the other hand, the OA (outlier analysis) software is an Apsis module that aims at identifying groups of similar objects in the *Gaia* DR3 sample exclusively according to their BP and RP spectra. The approach of OA to unsupervised clustering is entirely empirical by implementing self-organising maps (Kohonen 2001). One can further explore the resulting clusters, label them, and identify new classes of objects. However, OA analyses only 10% of the sources that DSC processed, those with the lowest DSC combined probabilities of membership to astronomical classes. They represent about 56 million sources in *Gaia* DR3. We note that the analysis scope will expand in *Gaia* DR4.

To compare the results from OA to those of DSC, we identified OA clusters associated with the DSC classes (see Sect. 11.3.12 in the online documentation for further details). Table 6 presents the resulting confusion matrix between DSC and OA. We find an agreement of 83% between the two classifications for galaxies, but an agreement of only 35% for quasars, where OA confused them with stars and white dwarfs. We assume that the extragalactic classification from DSC is accurate, as shown in Delchambre et al. (2023). We note that DSC includes astrometric information in its analysis which OA does not. It is thus not surprising to find significant differences. These results show that the two classifications are complementary.

One way to analyse OA neurons (or clusters) is to compare their prototype spectra with templates. We constructed our templates from averaged spectra with reliable spectral classifications in the literature, mainly from APOGEE-DR17 and GALAH-DR3. The online documentation (Sect. 11.3.12) details our procedure. Based on these stellar templates, OA attributed spectral labels (A-, F-, G-, K-, and M-type stars) to its relevant clusters. We compared these labels to the GSP-Phot temperatures (`teff_gspphot`). We cast the $T_{\rm eff}$ scale of GSP-Phot stars into O ($T_{\rm eff} \geq 30\,000$ K), B ($10\,000 \leq T_{\rm eff} < 30\,000$ K), A ($7300 \leq T_{\rm eff} < 10\,000$ K), stars), F ($5950 \leq T_{\rm eff} < 7300$ K), G ($5200 \leq T_{\rm eff} < 5950$ K), K ($3760 \leq T_{\rm eff} < 5200$ K), and M ($T_{\rm eff} < 3760$ K), and we constructed the confusion matrix shown in Table 7, which shows the agreement between the two modules.

Overall, the agreement between both classifications is very high. However, we found 51 O-type stars, 6 B-type stars, and 10 A-type stars from GSP-Phot that OA classified as late-type stars. Figure 38 shows 18 BP/RP spectra from stars labelled M-type by OA but with GSP-Phot $T_{\rm eff} > 30\,000$ K. The SEDs of all these objects peak around 850 nm, as is typically expected for cool stars. As a result of visual inspection, OA identified erroneous $T_{\rm eff}$ labels from GSP-Phot.

On the one hand, *Gaia* DR3 shows a richness and variety of information about Milky Way stars. On the other hand, the different interpretations and inconsistencies in the analysis we provide in the catalogue mean that we are to proceed with caution.

4. Candidates for deeper science analyses

We provide a list of six example use cases for deeper scientific analysis.

The first case is the identification of sources within some AP ranges. One should use the confidence intervals to find all sources of interest. For instance, Gaia Collaboration (2023d) select upper main-sequence stars based on their apparent colours. Gaia Collaboration (2023f) defined various golden samples of stars using our APs, stars with the most accurate and precise astrophysical parameters: for example FGK star samples supporting many Galactic surveys, solar analogs, ultra-cool dwarfs, carbon stars, and OBA stars challenging our stellar evolution and atmosphere models.

The second case is constructing the chemodynamical distribution of stars in some region of space. For instance, Gaia Collaboration (2023e) analysed the chemical patterns in the positions and orbital motions of stars to reveal the flared structure of the Milky Way disk and the various orbital substructures associated with chemical patterns.

The third case is constructing the three-dimensional spatial properties of the ISM. Using published extinctions and distances, Dharmawardena et al. (2022) inferred the individual structure of the Orion, Taurus, Perseus, and Cygnus X star-forming regions and found coherent ISM filaments that may link the Taurus and Perseus regions. One could easily replace those estimates with those (or a subset of those) we presented. Similarly, Gaia Collaboration (2023c) explores the ISM kinematics using our DIB measurements.

A fourth case is the age dating of wide binaries in the field. If an MS star has a white dwarf (WD) companion and a known distance, the age of this binary system can be determined precisely from the WD cooling sequence as long as the MS companion gives the chemical composition, which much harder to obtain

Table 6. DSC versus OA class label confusion matrix for the sample in common.

DSC	OA					Total
	STAR	WD	QSO	GAL	UNK	
STAR	21 073 253 (40%)	1 735 025 (3%)	11 834 708 (22%)	12 709 682 (24%)	5 942 859 (11%)	53 295 527
WD	38 651 (42%)	47 418 (51%)	2881 (3%)	0 (0%)	3236 (4%)	92 186
QSO	617 511 (29%)	453 890 (21%)	763 200 (35%)	48 658 (2%)	275 657 (13%)	2 158 916
GAL	30 351 (4%)	2542 (0%)	73 493 (9%)	708 253 (83%)	36 488 (4%)	851 127
UNK	4110 (22%)	1320 (7%)	6481 (35%)	4183 (22%)	2510 (13%)	18 604

Table 7. GSP-Phot versus OA stellar type confusion matrix for the sample in common.

GSP-Phot	OA					Total
	STAR-A	STAR-F	STAR-G	STAR-K	STAR-M	
STAR-O	146 (56%)	61 (24%)	1 (0%)	1 (0%)	50 (19%)	259
STAR-B	4082 (92%)	339 (8%)	6 (0%)	3 (0%)	3 (0%)	4433
STAR-A	23 836 (99%)	250 (1%)	22 (0%)	10 (0%)	0 (0%)	24 118
STAR-F	4868 (4%)	126 786 (95%)	1719 (1%)	215 (0%)	34 (0%)	133 622
STAR-G	0 (0%)	5955 (13%)	37 699 (83%)	1697 (4%)	172 (0%)	45 523
STAR-K	0 (0%)	0 (0%)	5694 (2%)	241 823 (64%)	131 517 (35%)	379 034
STAR-M	0 (0%)	0 (0%)	0 (0%)	11 613 (2%)	624 845 (98%)	636 458

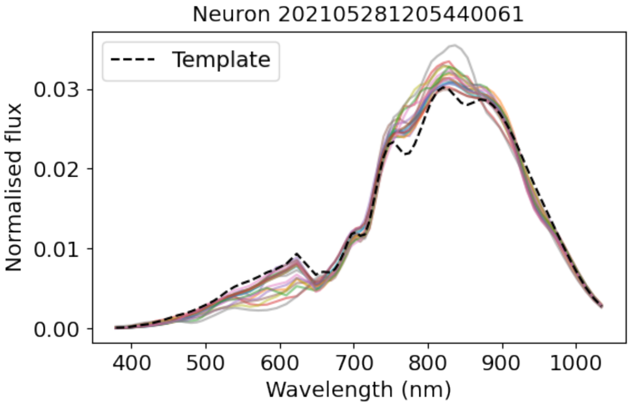

Fig. 38. BP and RP spectra of 18 stars labelled as M-type by OA, but with $T_{\rm eff} > 30\,000$ K from GSP-Phot. The dashed line indicates the best stellar template for this cluster, corresponding to a M-type star.

from the WD directly (e.g. Fouesneau et al. 2019; Qiu et al. 2021).

A fifth case is providing the largest uniformly derived set of APs that one could use to calibrate theoretical or data-driven stellar models. For instance, Green et al. (2021) developed a data-driven modelling technique to map stellar parameters (e.g. $T_{\rm eff}$, $\log g$, [M/H]) accurately to spectrophotometric space, supporting more accurate 3D mapping of the Milky Way.

A sixth application could be understanding the details of star formation and the dynamical evolution of star clusters. For instance, Fig. 39 compares the FLAME (current) mass estimates with a simulation of stars drawn for a universal initial mass function (IMF; assumed here to be one following Kroupa 2001). This simulation is created by sampling the mass function (over the given mass range) for each cluster with their respective given number of *Gaia*-identified members with mass estimates. Although we make a comparison of current with initial stellar masses, the agreement is very good overall. The lower-mass end is affected by how many low-mass stars *Gaia* can extract from these clusters and thus cannot be well reproduced without a selection function. The upper-mass end agrees perfectly with our predictions from a single IMF. We note that FLAME cannot predict masses above 10 $M_\odot$ with its current models. This analysis could support the study of cluster evaporation and mass segregation when also accounting for stellar mass loss.

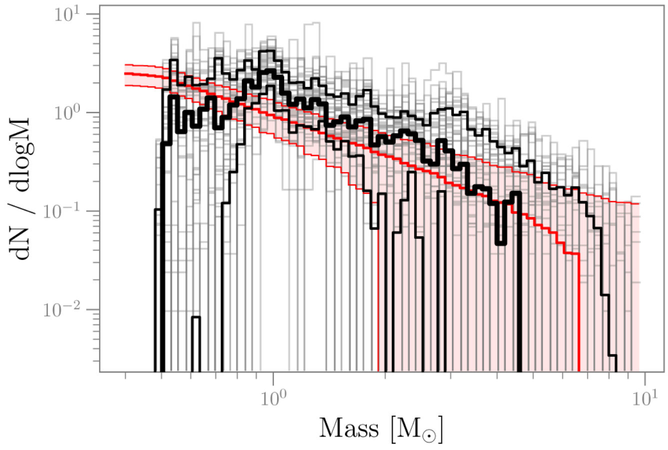

Fig. 39. Mass distribution from FLAME compared with a Kroupa (2001) IMF. For each of the 44 open clusters from Gaia Collaboration (2018b), we plot (grey) the recovered mass distributions from FLAME estimates. We highlight the overall median and [16, 84th] percentile interval in black. For reference, we plot in red the expected shape of masses drawn from a Kroupa IMF accounting for the limited number of identified members. Because of the noise of low number-statistics, we expect significant scatter from cluster to cluster. The low-mass end is affected by the *Gaia* selection function.

This list is not exhaustive. The previous *Gaia* data releases led to thousands of studies ranging from Solar System objects to discovering new streams and merger episodes that shaped our Galaxy.

5. Limitations

We recall the following assumptions and limitations of our *Gaia* DR3 catalogue. We produced APs that summarised many-dimensional posterior distributions using only quantile numbers such as mean, median, and percentile values (computed on one-dimensional marginal distributions). It is rarely possible to recover the complexity of the posterior distributions per object. One can query the MCMC chains published by GSP-Phot and MSC. These summary statistics cannot capture the full complexity of these distributions. One should not ignore the confidence intervals.

Most sources in *Gaia* DR3 have substantial fractional parallax uncertainties. Hence, the spectro-photometric data (BP/RP) often dominate the inference of our distances and APs. However, the parallax remains generally sufficient to limit the degeneracies of dwarfs versus giants.

The poorer the data, the more strongly our prior dominates our estimates. Our priors vary significantly from one Apsis module to the next. None of the modules included a three-dimensional extinction map or a detailed Milky Way model. One should expect significant differences with other AP catalogues when the prior dominates. However, in reality, if the actual stellar population, extinction, or reddening distributions are very different from those of Galactic models, these differences may partially indicate these deviations.

To derive stellar APs, we implicitly assumed that all *Gaia* sources are single stars in the Galaxy (apart from MSC). These estimates are most likely incorrect for any non-single star (binaries, extended sources, or extragalactic sources).

Furthermore, our stellar models also had intrinsic limitations in the range of parameters they covered. For instance, our models did not include specific physics inherent to WDs, AGBs, and HB stars.

Finally, by design, we inferred properties for each source independently. If a set of stars is known to be in a cluster, they have a similar distance, extinction, chemical patterns, and age. It constitutes a prior that one should exploit to infer the properties of the individual stars more accurately than what we have done here.

6. Summary

We have produced a catalogue of distances, astrophysical, and dust extinction parameters using the *Gaia* BP, RP, RVS spectra, integrated G photometry, and parallaxes available with *Gaia* DR3. More specifically, we provide:

- 470 million distances, $T_{\rm eff}$, $\log g$, and [M/H] estimates using BP/RP;
- 6 million using RVS $T_{\rm eff}$, $\log g$, [M/H], and [α/Fe] estimates;
- 470 million radius estimates;
- 140 million mass, and 120 million age estimates;
- 5 million chemical abundance ratios;
- half a million diffuse interstellar band analysis parameters;
- 2 million stellar activity indices;
- 200 million Hα equivalent widths;
- and further stellar classification with 220 million spectral types and 50 thousand emission-line stars.

We only presented a high-level overview of the validation and performance of these data products. We detail some of these tests and results in Creevey et al. (2023), Delchambre et al. (2023), Andrae et al. (2023), Recio-Blanco et al. (2023), Lanzafame et al. (2023), Babusiaux et al. (2023), and in the online documentation. Our tests comprised confirming the astrophysical consistency of our data through HR or Kiel diagrams, for example, which help to point out weaknesses in our analyses or failure in specific regions of the stellar parameter spaces. In addition, we compared our estimates with literature data to assess the performance of Apsis. The complexity and spread of our products often caused us to restrict our tests to subsamples and extrapolate our conclusions.

We emphasise that we did not calibrate Apsis APs to mimic external catalogues. Many of these external catalogues are not consistent with each other. As we do not know the true absolute scale of each AP dimension, we sometimes used external catalogues to obtain statistical relations to anchor our APs in a common ground. We recommend using these relations, but we did not apply them before the publication and instead provide the community with internally consistent APs.

First and foremost, our models have limitations in the range of parameters they can handle. We made assumptions that we discussed in Sect. 5.

Our data necessarily demanded several extreme simplifications and assumptions. Therefore, one should use the data with great care. We recommend always using the flags and filters, defined in Appendix A.

Our catalogue increases the availability of APs in the literature while offering results based on assumptions that differ from previous works. These works helped to validate our results. In addition, it provides the community with values of reference to explore and understand the content of *Gaia* DR3 better.

Gaia DR3 is not an incremental improvement of the *Gaia* data. It multiplies the quantities of multi-messenger information of *Gaia* with new data products (e.g. BP, RP, RVS, and APs). We increased the volume of sources with APs by a factor of 5, but we also increased the number of APs from 2 to ~40. *Gaia* DR3 represents a significant step forward to anchor all current and future spectroscopic surveys in a common ground, and it provides the most comprehensive view of our Galaxy.

Acknowledgements. We would like to thank the anonymous referee for the review of the manuscript and Carme Jordi for the very constructive feedback during the internal review of the manuscript. 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. This research has used NASA's Astrophysics Data System, the VizieR catalogue access tool (CDS, Strasbourg, France). This publication made extensive use of the online authoring Overleaf platform (`https://www.overleaf.com/`). The data processing and analysis made use of matplotlib (Hunter 2007), NumPy (Harris et al. 2020), the IPython package (Pérez & Granger 2007), Vaex (Breddels & Veljanoski 2018), TOPCAT (Taylor 2005), pyphot (`http://github.com/mfouesneau/pyphot`), QuantStack xtensor (`http://xtensor.readthedocs.io/`) and R (R Core Team 2013). Part of the calculations have been performed with the high-performance computing facility SIGAMM, hosted by the Observatoire de la Côte d'Azur. C.B.J., R.A., M.F., and J.R. acknowledge the support from the DLR (German space agency) via grants 50QG0602, 50QG1001, 50QG1403, and 50QG2102. A.B., A.R.B., A.C., G.C., G.K., Pd.L., C.O., P.A.P., M.S., and F.T. acknowledge financial supports from the french space agency (CNES), Agence National de la Recherche (ANR 14-CE33-014-01) and Programmes Nationaux de Physique Stellaire & Cosmologie et Galaxies (PNPS & PNCG) of CNRS/INSU. A.G., A.K., B.E., J.S. and U.H. acknowledge financial supports from the Swedish National Space Agency (SNSA). L.S. acknowledges financial support by the Spanish Ministerio de Ciencia e Innovación through grant PID2020-112949GB-I00. Y.F. and A.L. acknowledge the BELgian federal Science Policy Office (BELSPO) through various PROgramme de Développement d'Expériences scientifiques (PRODEX) grants.

References

Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35
Anders, F., Chiappini, C., Rodrigues, T. S., et al. 2017, A&A, 597, A30
Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2023, A&A, 658, A91
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8
Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, A&A, 674, A27 (*Gaia* DR3 SI)
Ankerst, M., Breunig, M. M., Peter Kriegel, H., & Sander, J. 1999, Optics: Ordering Points to Identify the Clustering Structure (ACM Press), 49

Babusiaux, C., Fabricius, C., Khanna, S., et al. 2023, A&A, 674, A32 (*Gaia* DR3 SI)

Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147

Blomme, R., Daflon, S., Gebran, M., et al. 2022, A&A, 661, A120

Blomme, R., Frémat, Y., Sartoretti, P., et al. 2023, A&A, 674, A7 (*Gaia* DR3 SI)

Boyajian, T. S., McAlister, H. A., van Belle, G., et al. 2012a, ApJ, 746, 101

Boyajian, T. S., von Braun, K., van Belle, G., et al. 2012b, ApJ, 757, 112

Boyajian, T. S., von Braun, K., van Belle, G., et al. 2013, ApJ, 771, 40

Bragaglia, A., & Tosi, M. 2006, AJ, 131, 1544

Breddels, M. A., & Veljanoski, J. 2018, A&A, 618, A13

Brott, I., & Hauschildt, P. H. 2005, in The Three-Dimensional Universe with Gaia, eds. C. Turon, K. S. O'Flaherty, & M. A. C. Perryman, ESA Spec. Publ., 576, 565

Buder, S., Sharma, S., Kos, J., et al. 2021, MNRAS, 506, 150

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

Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86

Casagrande, L., Schoenrich, R., Asplund, M., et al. 2011, VizieR Online Data Catalogue: J/A+A/530/A138

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:`1612.05560`]

Contursi, G., de Laverny, P., Recio-Blanco, A., & Palicio, P. A. 2021, A&A, 654, A130

Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, A&A, 674, A26 (*Gaia* DR3 SI)

Dahm, S. E., & Simon, T. 2005, AJ, 129, 829

Datson, J., Flynn, C., & Portinari, L. 2015, A&A, 574, A124

De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, A&A, 674, A2 (*Gaia* DR3 SI)

Delchambre, L., Bailer-Jones, C. A. L., Bellas-Velidis, et al. I., et al. 2023, A&A, 674, A31 (*Gaia* DR3 SI)

Dharmawardena, T. E., Bailer-Jones, C. A. L., Fouesneau, M., & Foreman-Mackey, D. 2022, A&A, 658, A166

Dowler, P., Bonnarel, F., Michel, L., & Demleitner, M. 2015, IVOA DataLink Version 1.0, IVOA Recommendation 17 June 2015

Ducourant, C., Krone-Martins, A., Galluccio, L., et al. 2023, A&A, 674, A11 (*Gaia* DR3 SI)

Duvert, G. 2016, VizieR Online Data Catalog: II/345

Eisenhardt, P. R. M., Marocco, F., Fowler, J. W., et al. 2020, ApJS, 247, 69

El-Badry, K., Ting, Y.-S., Rix, H.-W., et al. 2018, MNRAS, 476, 528

Eyer, L., Audard, M., Holl, B., et al. 2023, A&A, 674, A13 (*Gaia* DR3 SI)

Fouesneau, M., Rix, H.-W., von Hippel, T., Hogg, D. W., & Tian, H. 2019, ApJ, 870, 9

Fouesneau, M., Andrae, R., Dharmawardena, T., et al. 2022, A&A, 662, A125

Frémat, Y., Royer, F., Marchal, O., et al. 2023, A&A, 674, A8 (*Gaia* DR3 SI)

Gagné, J., Mamajek, E. E., Malo, L., et al. 2018, ApJ, 856, 23

Gaia Collaboration 2022, Gaia DR3 Online Documentation, `https://gea.esac.esa.int/archive/documentation/GDR3/`

Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1

Gaia Collaboration (Brown, A. G. A., et al.) 2018a, A&A, 616, A1

Gaia Collaboration (Babusiaux, C., et al.) 2018b, A&A, 616, A10

Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1

Gaia Collaboration (Vallenari, A., et al.) 2023a, A&A, 674, A1 (*Gaia* DR3 SI)

Gaia Collaboration (Bailer-Jones, C. A. L., et al.) 2023b, A&A, 674, A41 (*Gaia* DR3 SI)

Gaia Collaboration (Schultheis, M. et al.) 2023c, A&A, 674, A40 (*Gaia* DR3 SI)

Gaia Collaboration (Drimmel, R., et al.) 2023d, A&A, 674, A37 (*Gaia* DR3 SI)

Gaia Collaboration (Recio-Blanco, A., et al.) 2023e, A&A, 674, A38 (*Gaia* DR3 SI)

Gaia Collaboration (Creevey, O. L., et al.) 2023f, A&A, 674, A39 (*Gaia* DR3 SI)

Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25

Green, G. M., Rix, H.-W., Tschesche, L., et al. 2021, ApJ, 907, 57

Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357

Hawkins, K., Lucey, M., Ting, Y.-S., et al. 2020, MNRAS, 492, 1164

Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, A&A, 582, A49

Hidalgo, S. L., Pietrinferni, A., Cassisi, S., et al. 2018, ApJ, 856, 125

Huang, Y., Beers, T. C., Wolf, C., et al. 2022, ApJ, 925, 164

Huber, D., Zinn, J., Bojsen-Hansen, M., et al. 2017, ApJ, 844, 102

Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90

Johansson, J., Thomas, D., & Maraston, C. 2010, MNRAS, 406, 165

Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, AJ, 160, 120

Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, A&A, 558, A53

Kirkpatrick, J. D., Henry, T. J., & Irwin, M. J. 1997, AJ, 113, 1421

Kohonen, T. 2001, Self-Organizing Maps (New York: Springer-Verlag)

Kroupa, P. 2001, MNRAS, 322, 231

Lanz, T., & Hubeny, I. 2003, ApJS, 146, 417

Lanz, T., & Hubeny, I. 2007, ApJS, 169, 83

Lanzafame, A. C., Brugaletta, E., Frémat, Y., et al. 2023, A&A, 674, A30 (*Gaia* DR3 SI)

Lebreton, Y., & Reese, D. R. 2020, A&A, 642, A88

Leike, R. H., & Enßlin, T. A. 2019, A&A, 631, A32

Lindegren, L., Bastian, U., Biermann, M., et al. 2021, A&A, 649, A4

López-Sanjuan, C., Yuan, H., Vázquez Ramió, H., et al. 2021, A&A, 654, A61

Luo, A. L., Zhao, Y. H., Zhao, G., et al. 2019, VizieR Online Data Catalog: V/164

Maíz Apellániz, J., Sota, A., Morrell, N. I., et al. 2013, Massive Stars: From alpha to Omega, 198

Manoj, P., Bhatt, H. C., Maheswar, G., & Muneer, S. 2006, ApJ, 653, 657

Manteiga, M., Ordóñez, D., Dafonte, C., & Arcay, B. 2010, PASP, 122, 608

Martell, S. L., Sharma, S., Buder, S., et al. 2017, MNRAS, 465, 3203

Montegriffo, P., De Angeli, F., Andrae, R., et al. 2023, A&A, 674, A3 (*Gaia* DR3 SI)

Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682

Napiwotzki, R., Schoenberner, D., & Wenske, V. 1993, A&A, 268, 653

Newton, E. R., Irwin, J., Charbonneau, D., et al. 2017, ApJ, 834, 85

Pérez, F., & Granger, B. E. 2007, Comput. Sci. Eng., 9, 21

Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168

Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, ApJS, 239, 32

Pourbaix, D., Tokovinin, A. A., Batten, A. H., et al. 2004, A&A, 424, 727

Qiu, D., Tian, H.-J., Wang, X.-D., et al. 2021, ApJS, 253, 58

Raddi, R., Drew, J. E., Steeghs, D., et al. 2015, MNRAS, 446, 274

R Core Team 2013, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria

Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, A&A, 585, A93

Recio-Blanco, A., de Laverny, P., Palicio, P. A., et al. 2023, A&A, 674, A29 (*Gaia* DR3 SI)

Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3

Rimoldini, L., Holl, B., Gavras, P., et al. 2023, A&A, 674, A14 (*Gaia* DR3 SI)

Roca Cortés, T., & Pallé, P. L. 2014, MNRAS, 443, 1837

Rosslowe, C. K., & Crowther, P. A. 2015, MNRAS, 447, 2322

Rybizki, J., Demleitner, M., Bailer-Jones, C., et al. 2020, PASP, 132, 074501

Rybizki, J., Green, G. M., Rix, H.-W., et al. 2022, MNRAS, 510, 2597

Serenelli, A., Johnson, J., Huber, D., et al. 2017, ApJS, 233, 23

Shulyak, D., Tsymbal, V., Ryabchikova, T., Stütz, C., & Weiss, W. W. 2004, A&A, 428, 993

Silaj, J., Jones, C. E., Tycner, C., Sigut, T. A. A., & Smith, A. D. 2010, ApJS, 187, 228

Smart, R. L., Marocco, F., Caballero, J. A., et al. 2017, MNRAS, 469, 401

Smart, R. L., Marocco, F., Sarro, L. M., et al. 2019, MNRAS, 485, 4423

Soubiran, C., Le Campion, J.-F., Brouillet, N., & Chemin, L. 2016, A&A, 591, A118

Starkenburg, E., Martin, N., Youakim, K., et al. 2017, MNRAS, 471, 2587

Steinmetz, M., Matijevič, G., Enke, H., et al. 2020, AJ, 160, 82

Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154

Stevens, D. J., Stassun, K. G., & Gaudi, B. S. 2017, AJ, 154, 259

Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29

Traven, G., Feltzing, S., Merle, T., et al. 2020, A&A, 638, A145

Tucci Maia, M., Ramírez, I., Meléndez, J., et al. 2016, A&A, 590, A32

van Belle, G. T., von Braun, K., Ciardi, D. R., et al. 2021, ApJ, 922, 163

Vioque, M., Oudmaijer, R. D., Baines, D., Mendigutía, I., & Pérez-Martínez, R. 2018, A&A, 620, A128

Weidner, C., & Vink, J. S. 2010, A&A, 524, A98

Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924

Wu, Y., Du, B., Luo, A., Zhao, Y., & Yuan, H. 2014, in Statistical Challenges in 21st Century Cosmology, eds. A. Heavens, J. L. Starck, & A. Krone-Martins, IAU Symp., 306, 340

Xiang, M., Rix, H.-W., Ting, Y.-S., et al. 2022, A&A, 662, A66

Yu, J., Huber, D., Bedding, T. R., et al. 2018, ApJS, 236, 42

---

[1] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany

[2] Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium

[3] Observational Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden

[4] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, Allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[5] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[6] INAF – Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
[7] Dpto. de Inteligencia Artificial, UNED, c/ Juan del Rosal 16, 28040 Madrid, Spain
[8] INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy
[9] Dipartimento di Fisica e Astronomia "Ettore Majorana", Università di Catania, Via S. Sofia 64, 95123 Catania, Italy
[10] CIGUS CITIC – Department of Computer Science and Information Technologies, University of A Coruña, Campus de Elviña s/n, A Coruña 15071, Spain
[11] INAF – Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese, TO, Italy
[12] Institut d'Astrophysique et de Géophysique, Université de Liège, 19c, Allée du 6 Août, 4000 Liège, Belgium
[13] Department of Astrophysics, Astronomy and Mechanics, National and Kapodistrian University of Athens, Panepistimiopolis, Zografos, 15783 Athens, Greece
[14] National Observatory of Athens, I. Metaxa and Vas. Pavlou, Palaia Penteli, 15236 Athens, Greece
[15] Aurora Technology for European Space Agency (ESA), Camino Bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
[16] Telespazio for CNES Centre Spatial de Toulouse, 18 Avenue Édouard Belin, 31401 Toulouse Cedex 9, France
[17] CNES Centre Spatial de Toulouse, 18 Avenue Édouard Belin, 31401 Toulouse Cedex 9, France
[18] Thales Services for CNES Centre Spatial de Toulouse, 18 Avenue Édouard Belin, 31401 Toulouse Cedex 9, France
[19] European Space Agency (ESA), European Space Astronomy Centre (ESAC), Camino Bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
[20] ATG Europe for European Space Agency (ESA), Camino Bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
[21] Aix-Marseille Univ., CNRS, CNES, LAM, Marseille, France
[22] Dpto. de Matemática Aplicada y Ciencias de la Computación, Univ. de Cantabria, ETS Ingenieros de Caminos, Canales y Puertos, Avda. de los Castros s/n, 39005 Santander, Spain
[23] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[24] Centre for Astrophysics Research, University of Hertfordshire, College Lane, AL10 9AB Hatfield, UK
[25] APAVE SUDEUROPE SAS for CNES Centre Spatial de Toulouse, 18 Avenue Édouard Belin, 31401 Toulouse Cedex 9, France
[26] Data Science and Big Data Lab., Pablo de Olavide University, 41013 Seville, Spain
[27] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 Place Jules Janssen, 92190 Meudon, France
[28] Université Rennes, CNRS, IPR (Institut de Physique de Rennes) – UMR 6251, 35000 Rennes, France
[29] Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen Ø, Denmark
[30] DXC Technology, Retortvej 8, 2500 Valby, Denmark
[31] CIGUS CITIC, Department of Nautical Sciences and Marine Engineering, University of A Coruña, Paseo de Ronda 51, 15071 A Coruña, Spain
[32] IPAC, California Institute of Technology, Mail Code 100-22, 1200 E. California Blvd., Pasadena, CA 91125, USA
[33] European Organisation for Astronomical Research in the Southern Hemisphere, Alonso de Córdova 3107, Vitacura, 19001 Casilla, Santiago de Chile, Chile
[34] IRAP, Université de Toulouse, CNRS, UPS, CNES, 9 Av. colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France
[35] Institute of Global Health, University of Geneva, 24 Rue du Général-Dufour, 1211 Genève 4, Geneva, Switzerland
[36] Applied Physics Department, Universidade de Vigo, 36310 Vigo, Spain
[37] Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98bis Bd. Arago, 75014 Paris, France

Appendix A: Recommended caution and corrections

We recommend the following corrections for the *Gaia* DR3 AP products:

- GSP-Phot provides a Python tool that implements empirical calibration models of its stellar parameters. It currently provides a metallicity, [M/H], and an effective temperature, $T_{\rm eff}$, model. We trained these calibration models on literature catalogues (e.g LAMOST DR6) based upon machine-learning algorithms. These models are not simple equations, and therefore, we provide a wrapper for the users so any update in the models will be transparently propagated (see Andrae et al. 2023).

- We recommend checking for potential outliers in GSP-Phot APs by using the fractional parallax uncertainties (σ_ϖ/ϖ). This also helps identify inference priors and assumptions about the Milky Way structure that matter.

- GSP-Spec provides an extensive flag definition detailed in Recio-Blanco et al. (2023).

- GSP-Spec also recommend polynomial functions of $\log g$ to rescale the various abundance ratios ([M/H], [α/Fe], [Mg/H], etc.). These relations are polynomial equations with coefficients given in Table 3 of Recio-Blanco et al. (2023).

Appendix B: Example queries

In this section, we list query examples that we used to produce various figures in the manuscript.

• Data for Figs. 1 and 3. The following query took about 9 hours to run. Selecting random subsets could extract statistically equivalent smaller datasets and run faster.

```
select
  round((gaia.phot_g_mean_mag + 5 *
      log10(parallax/100)) * 10) / 10 as gmag,
  round(gaia.bp_rp * 10) / 10 as bp_rp,
  count(*) as n,
  sum(IF_THEN_ELSE(gaia.has_xp_continuous =
      ’true’, 1, 0)) as has_xp,
  sum(IF_THEN_ELSE(gaia.has_rvs = ’true’, 1, 0))
      as has_rvs,
  count(aps.classprob_dsc_combmod_star) as dsc,
  count(aps.teff_gspphot) as gspphot,
  count(aps.teff_gspspec) as gspspec,
  count(aps.classlabel_espels) as espels,
  count(aps.teff_esphs) as esphs,
  count(aps.teff_espucd) as espucd,
  count(aps.activityindex_espcs) as espcs,
  count(aps.radius_flame) as flame,
  count(aps.teff_msc1) as msc,
  count(aps.neuron_oa_id) as oa
from gaiadr3.gaia_source as gaia
inner join gaiadr3.astrophysical_parameters as aps
   on aps.source_id = gaia.source_id
group by bp_rp, gmag
order by bp_rp, gmag
```

• The Kiel diagram from GSP-Phot used in Fig.2 and Fig. 11.

```
select
   floor(log10(teff_gspphot) / 0.05) * 0.05 as
       logT,
   floor(logg_gspphot / 0.05) * 0.05 as logg,
   count(*) as n
from gaiadr3.gaia_source
group by logT, logg
```

• Data for Fig. 7. The following query took about 20 minutes to run. We note that Cantat-Gaudin et al. (2020) shared their cluster catalogue through the Archive

```
select
   cl_members.source_id, cl_members.cluster,
   cl_prop.agenn as logage, cl_prop.avnn as av,
   cl_prop.dmnn, cl_prop.distpc,
   gaia.distance_gspphot, gaia.azero_gspphot
from user_tcantatg.members_2681_ocs as cl_members
inner join user_tcantatg.clusters_dr3_astrometry
    as cl_prop
   on cl_members.cluster = cl_prop.oc
inner join gaiadr3.gaia_source as gaia
   on gaia.source_id = cl_members.source_id
```

• The queries for Fig. 14. As one cannot use the string comparison in the selection clause, we needed to run two queries.

```
select round(log10(teff_gspphot) * 100) * 0.01 as
    logteff_gspphot,
      round(log10(teff_gspspec) * 100) * 0.01 as
          logteff_gspspec,
      count(*) as n
from gaiadr3.astrophysical_parameters
group by logteff_gspphot, logteff_gspspec
```

```
select round(log10(teff_gspphot) * 100) * 0.01 as
    logteff_gspphot,
      round(log10(teff_gspspec) * 100) * 0.01 as
          logteff_gspspec,
      count(*) as n
from gaiadr3.astrophysical_parameters
where flags_gspspec is not NULL
and flags_gspspec like ’0000000000000%’
group by logteff_gspphot, logteff_gspspec
```

• The queries for Fig. 26. We generated the quantities directly during the query.

```
select
   round((2 * log10(aps.radius_gspphot) + 4 *
       log10(aps.teff_gspphot / 5778.)) * 100.) /
       100. as loglum_gspphot,
   round(log10(aps.lum_flame) * 100) / 100. as
       loglum_flame,
   count(*) as n
from gaiadr3.astrophysical_parameters as aps
group by loglum_gspphot, loglum_flame
order by loglum_gspphot, loglum_flame
```

```
select
   round(aps.mg_gspphot * 100.) / 100. as
       mg_gspphot,
   round((4.74 - 2.5 * log10(aps.lum_flame) -
       aps.bc_flame) * 100) / 100. as mg_flame,
   count(*) as n
from gaiadr3.astrophysical_parameters as aps
group by mg_gspphot, mg_flame
order by mg_gspphot, mg_flame
```

• The queries for Fig. 31, in which we used the tracer selection from TGE to select the giant stars from GSP-Phot.

```
select
    count(*) as n, GAIA_HEALPIX_INDEX(9,
        aps.source_id) as hpx9,
    avg(aps.azero_gspphot) as azero_mean_gspphot,
        stddev(aps.azero_gspphot) as
        azero_std_gspphot,
    aps.libname_gspphot,
```

```
    num_tracers_used as n_tge, a0 as
        azero_mean_tge, a0_uncertainty as
        azero_std_tge
from gaiadr3.astrophysical_parameters as aps
inner join
    gaiadr3.total_galactic_extinction_map_opt
    on healpix_id = GAIA_HEALPIX_INDEX(9,
        aps.source_id)
where aps.azero_gspphot is not NULL
and teff_gspphot between 3000 and 5700 and
    mg_gspphot between -10 and 4
group by hpx9, libname_gspphot
```

Appendix C: Candidate UCDs in young associations

Table C.1 lists the young associations for which we identified candidate UCD members using BANYAN Σ (Gagné et al. 2018) or the OPTICS clustering algorithm (Ankerst et al. 1999).

Table C.1. Number of UCD candidates in nearby young associations according to BANYAN Σ and our clustering analysis using the OPTICS algorithm.

Association	# UCDs	Min $T_{\rm eff}$ (K)	Method
CARN	61	1250	BANYAN Σ
ARG	424	1494	BANYAN Σ
ABDMG	155	1557	BANYAN Σ
BPMG	47	1874	BANYAN Σ
THA	42	1882	BANYAN Σ
UCL	575	1991	BANYAN Σ
COL	39	2006	BANYAN Σ
HYA	52	2070	BANYAN Σ
CAR	20	2105	BANYAN Σ
OCT	153	2110	BANYAN Σ
LCC	241	2166	BANYAN Σ
ROPH	63	2168	BANYAN Σ
USCO	508	2176	BANYAN Σ
TWA	11	2189	BANYAN Σ
TAU	214	2190	BANYAN Σ
THOR	11	2233	BANYAN Σ
CRA	7	2262	BANYAN Σ
PL8	20	2279	BANYAN Σ
IC2391	20	2321	BANYAN Σ
PLE	97	2331	BANYAN Σ
IC2602	12	2360	BANYAN Σ
UCRA	45	2362	BANYAN Σ
EPSC	5	2374	BANYAN Σ
VCA	4	2385	BANYAN Σ
XFOR	2	2391	BANYAN Σ
118TAU	5	2415	BANYAN Σ
CBER	7	2456	BANYAN Σ
NGC 1333 + IC 348	488	2230	OPTICS
Serpens	420	2185	OPTICS
Chameleon	69	2228	OPTICS
γ2 Vel	266	2387	OPTICS
Orion	1083	2156	OPTICS

Notes. BANYAN Σ identifiers, see Gagné et al. (2018), OPTICS clustering algorithm, see (Ankerst et al. 1999).

Appendix D: AP estimates, producers, and where to find them

In this section, we compile the various estimates of stellar parameters from *Gaia* DR3, list the Apsis module producing them, and indicate the table and field that store the values in the *Gaia* catalogue.

Table D.1 lists the distance estimates discussed in Sect. 3.1. Tables D.2 and D.3 list the primary and secondary atmospheric estimates, respectively (discussed in Sect. 3.2). Table D.4 lists the abundances estimates (discussed in Sect. 3.2.3). Table D.5 lists the parameters characterising the evolutionary state of a star. Finally, Table D.6 lists the extinction parameters and the diffuse interstellar band properties (discussed in Sect. 3.4).

Table D.1. Distance estimates in *Gaia* DR3.

distance	GSP-Phot	gaia_source.distance_gspphot
		astrophysical_parameters.distance_gspphot
		astrophysical_parameters_supp.distance_gspphot_a
		astrophysical_parameters_supp.distance_gspphot_marcs
		astrophysical_parameters_supp.distance_gspphot_ob
		astrophysical_parameters_supp.distance_gspphot_phoenix
	MSC	astrophysical_parameters.distance_msc

Table D.2. Primary atmospheric estimates in *Gaia* DR3: $T_{\rm eff}$, $\log g$, [M/H], and [α/Fe].

$T_{\rm eff}$	GSP-Phot	gaia_source.teff_gspphot
		astrophysical_parameters.teff_gspphot
		astrophysical_parameters_supp.teff_gspphot_a
		astrophysical_parameters_supp.teff_gspphot_marcs
		astrophysical_parameters_supp.teff_gspphot_ob
		astrophysical_parameters_supp.teff_gspphot_phoenix
	GSP-Spec	astrophysical_parameters.teff_gspspec
		astrophysical_parameters_supp.teff_gspspec_ann
	ESP-HS	astrophysical_parameters.teff_esphs
	ESP-UCD	astrophysical_parameters.teff_espucd
	MSC	astrophysical_parameters.teff_msc1
		astrophysical_parameters.teff_msc2
$\log g$	GSP-Phot	gaia_source.logg_gspphot
		astrophysical_parameters.logg_gspphot
		astrophysical_parameters_supp.logg_gspphot_a
		astrophysical_parameters_supp.logg_gspphot_marcs
		astrophysical_parameters_supp.logg_gspphot_ob
		astrophysical_parameters_supp.logg_gspphot_phoenix
	GSP-Spec	astrophysical_parameters.logg_gspspec
		astrophysical_parameters_supp.logg_gspspec_ann
	ESP-HS	astrophysical_parameters.logg_esphs
	MSC	astrophysical_parameters.logg_msc1
		astrophysical_parameters.logg_msc2
[M/H]	GSP-Phot	gaia_source.mh_gspphot
		astrophysical_parameters.mh_gspphot
		astrophysical_parameters_supp.mh_gspphot_a
		astrophysical_parameters_supp.mh_gspphot_marcs
		astrophysical_parameters_supp.mh_gspphot_ob
		astrophysical_parameters_supp.mh_gspphot_phoenix
	GSP-Spec	astrophysical_parameters.mh_gspspec
		astrophysical_parameters_supp.mh_gspspec_ann
[α/Fe]	GSP-Spec	astrophysical_parameters.alphafe_gspspec
		astrophysical_parameters_supp.alphafe_gspspec_ann

Table D.3. Secondary atmospheric estimates in *Gaia* DR3: classes, rotation, emission, and activity.

classification	DSC	`astrophysical_parameters.classprob_dsc_allosmod_star`
		`astrophysical_parameters.classprob_dsc_combmod_binarystar`
		`astrophysical_parameters.classprob_dsc_combmod_star`
		`astrophysical_parameters.classprob_dsc_combmod_whitedwarf`
		`astrophysical_parameters.classprob_dsc_specmod_binarystar`
		`astrophysical_parameters.classprob_dsc_specmod_star`
		`astrophysical_parameters.classprob_dsc_specmod_whitedwarf`

	ESP-HS	`astrophysical_parameters.spectraltype_esphs`

	ESP-ELS	`astrophysical_parameters.classlabel_espels`
		`astrophysical_parameters.classprob_espels_bestar`
		`astrophysical_parameters.classprob_espels_dmestar`
		`astrophysical_parameters.classprob_espels_herbigstar`
		`astrophysical_parameters.classprob_espels_pne`
		`astrophysical_parameters.classprob_espels_ttauristar`
		`astrophysical_parameters.classprob_espels_wcstar`
		`astrophysical_parameters.classprob_espels_wnstar`

rotation	ESP-HS	`astrophysical_parameters.vsini_esphs`
Chromospheric activity	ESP-ELS	`astrophysical_parameters.ew_espels_halpha`
	ESP-CS	`astrophysical_parameters.activityindex_espcs`

Table D.4. 13 chemical abundance ratios from 12 individual elements (N, Mg, Si, S, Ca, Ti, Cr, Fe, Ni, Zr, Ce, and Nd; with the FeI and FeII species) and the CN equivalent width in *Gaia* DR3.

chemical abundances	GSP-Spec	`astrophysical_parameters.fem_gspspec`
		`astrophysical_parameters.feiim_gspspec`
		`astrophysical_parameters.cafe_gspspec`
		`astrophysical_parameters.cefe_gspspec`
		`astrophysical_parameters.crfe_gspspec`
		`astrophysical_parameters.mgfe_gspspec`
		`astrophysical_parameters.ndfe_gspspec`
		`astrophysical_parameters.nfe_gspspec`
		`astrophysical_parameters.nife_gspspec`
		`astrophysical_parameters.sfe_gspspec`
		`astrophysical_parameters.sife_gspspec`
		`astrophysical_parameters.tife_gspspec`
		`astrophysical_parameters.zrfe_gspspec`
		`astrophysical_parameters.cn0ew_gspspec`

Table D.5. Evolution parameter estimates in *Gaia* DR3.

Luminosity L	FLAME	astrophysical_parameters.lum_flame astrophysical_parameters_supp.lum_flame_spec
absolute magnitude M_G	GSP-Phot	astrophysical_parameters.mg_gspphot astrophysical_parameters_supp.mg_gspphot_a astrophysical_parameters_supp.mg_gspphot_marcs astrophysical_parameters_supp.mg_gspphot_ob astrophysical_parameters_supp.mg_gspphot_phoenix
radius R	GSP-Phot	astrophysical_parameters.radius_gspphot astrophysical_parameters_supp.radius_gspphot_a astrophysical_parameters_supp.radius_gspphot_marcs astrophysical_parameters_supp.radius_gspphot_ob astrophysical_parameters_supp.radius_gspphot_phoenix
	FLAME	astrophysical_parameters.radius_flame astrophysical_parameters_supp.radius_flame_spec
age	FLAME	astrophysical_parameters.age_flame astrophysical_parameters_supp.age_flame_spec
mass M	FLAME	astrophysical_parameters.mass_flame astrophysical_parameters_supp.mass_flame_spec
evolution stage ϵ	FLAME	astrophysical_parameters.evolstage_flame astrophysical_parameters_supp.evolstage_flame_spec

Table D.6. Extinction and DIB parameter estimates in *Gaia* DR3.

monochromatic at 541.4 nm A_0	GSP-Phot	astrophysical_parameters.azero_gspphot astrophysical_parameters_supp.azero_gspphot_a astrophysical_parameters_supp.azero_gspphot_marcs astrophysical_parameters_supp.azero_gspphot_ob astrophysical_parameters_supp.azero_gspphot_phoenix
	ESP-HS	astrophysical_parameters.azero_esphs
	MSC	astrophysical_parameters.azero_msc
in G band A_G	GSP-Phot	astrophysical_parameters.ag_gspphot astrophysical_parameters_supp.ag_gspphot_a astrophysical_parameters_supp.ag_gspphot_marcs astrophysical_parameters_supp.ag_gspphot_ob astrophysical_parameters_supp.ag_gspphot_phoenix
	ESP-HS	astrophysical_parameters.ag_esphs
in BP band A_{BP}	GSP-Phot	astrophysical_parameters.abp_gspphot astrophysical_parameters_supp.abp_gspphot_a astrophysical_parameters_supp.abp_gspphot_marcs astrophysical_parameters_supp.abp_gspphot_ob astrophysical_parameters_supp.abp_gspphot_phoenix
in RP band A_{RP}	GSP-Phot	astrophysical_parameters.arp_gspphot astrophysical_parameters_supp.arp_gspphot_a astrophysical_parameters_supp.arp_gspphot_marcs astrophysical_parameters_supp.arp_gspphot_ob astrophysical_parameters_supp.arp_gspphot_phoenix
in BP-RP colour $E(G_{\mathrm{BP}}-G_{\mathrm{RP}})$	GSP-Phot	astrophysical_parameters.ebpminrp_gspphot astrophysical_parameters_supp.ebpminrp_gspphot_a astrophysical_parameters_supp.ebpminrp_gspphot_marcs astrophysical_parameters_supp.ebpminrp_gspphot_ob astrophysical_parameters_supp.ebpminrp_gspphot_phoenix
DIB central wavelength, EW, and complexity	GSP-Spec	astrophysical_parameters.dib_gspspec_lambda astrophysical_parameters.dibew_gspspec astrophysical_parameters.dibp0_gspspec astrophysical_parameters.dibp2_gspspec