A&A 674, A26 (2023)
https://doi.org/10.1051/0004-6361/202243688
© The Authors 2023

Astronomy & Astrophysics
Special issue

Gaia Data Release 3

Gaia Data Release 3

Astrophysical parameters inference system (Apsis). I. Methods and content overview

O. L. Creevey[1,⋆], R. Sordo[2], F. Pailler[3], Y. Frémat[4], U. Heiter[5], F. Thévenin[1], R. Andrae[6], M. Fouesneau[6], A. Lobel[4], C. A. L. Bailer-Jones[6], D. Garabato[7], I. Bellas-Velidis[8], E. Brugaletta[9], A. Lorca[10], C. Ordenovic[1], P. A. Palicio[1], L. M. Sarro[11], L. Delchambre[12], R. Drimmel[13], J. Rybizki[6], G. Torralba Elipe[7], A. J. Korn[5], A. Recio-Blanco[1], M. S. Schultheis[1], F. De Angeli[14], P. Montegriffo[15], A. Abreu Aramburu[16], S. Accart[3], M. A. Álvarez[7], J. Bakker[17], N. Brouillet[18], A. Burlacu[19], R. Carballo[20], L. Casamiquela[18,21], A. Chiavassa[1], G. Contursi[1], W. J. Cooper[22,13], C. Dafonte[7], A. Dapergolas[8], P. de Laverny[1], T. E. Dharmawardena[6], B. Edvardsson[24], Y. Le Fustec[19], P. García-Lario[25], M. García-Torres[26], A. Gomez[7], I. González-Santamaría[7], D. Hatzidimitriou[27,8], A. Jean-Antoine Piccolo[3], M. Kontiza[27], G. Kordopatis[1], A. C. Lanzafame[9,28], Y. Lebreton[29,30], E. L. Licata[13], H. E. P. Lindstrøm[13,31,32], E. Livanou[27], A. Magdaleno Romeo[19], M. Manteiga[33], F. Marocco[34], D. J. Marshall[35], N. Mary[36], C. Nicolas[3], L. Pallas-Quintela[7], C. Panem[3], B. Pichon[1], E. Poggio[1,13], F. Riclet[3], C. Robin[36], R. Santoveña[7], A. Silvelo[7], I. Slezak[1], R. L. Smart[13], C. Soubiran[18], M. Süveges[37], A. Ulla[38], E. Utrilla[10], A. Vallenari[2], H. Zhao[1], J. Zorec[39], D. Barrado[40], A. Bijaoui[1], J.-C. Bouret[41], R. Blomme[4], I. Brott[42], S. Cassisi[43,44], O. Kochukhov[5], C. Martayan[45,4], D. Shulyak[46,47], and J. Silvester[5,23]

(Affiliations can be found after the references)

Received 31 March 2022 / Accepted 31 May 2022

ABSTRACT

Gaia Data Release 3 contains a wealth of new data products for the community. Astrophysical parameters are a major component of this release, and were produced by the Astrophysical parameters inference system (Apsis) within the *Gaia* Data Processing and Analysis Consortium (DPAC). The aim of this paper is to describe the overall content of the astrophysical parameters in *Gaia* DR3 and how they were produced. In Apsis, we use the mean BP/RP and mean RVS spectra along with astrometry and photometry, and we derive the following parameters: source classification and probabilities for 1.6 billion objects; interstellar medium characterisation and distances for up to 470 million sources, including a 2D total Galactic extinction map; 6 million redshifts of quasar candidates; 1.4 million redshifts of galaxy candidates; and an analysis of 50 million outlier sources through an unsupervised classification. The astrophysical parameters also include many stellar spectroscopic and evolutionary parameters for up to 470 million sources. These 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 (up to 5 million), diffuse interstellar band analysis (0.5 million), activity indices (2 million), Hα equivalent widths (200 million), and further classification of spectral types (220 million) and emission-line stars (50 000). This paper is the first in a series of three papers, and focusses on describing the global content of the parameters in *Gaia* DR3. The accompanying Papers II and III focus on the validation and use of the stellar and non-stellar products, respectively. This catalogue is the most extensive homogeneous database of astrophysical parameters to date, and is based uniquely on *Gaia* data. It will only be superseded by *Gaia* Data Release 4, and will therefore remain a key reference over the next four years, providing astrophysical parameters independent of other ground- and space-based data.

Key words. methods: data analysis – catalogs – ISM: general – stars: fundamental parameters – Galaxy: stellar content – galaxies: fundamental parameters

1. Introduction

Physical characterisation of astrophysical objects is a key input for understanding the structure and evolution of astrophysical systems. By physical characterisation, we mean intrinsic properties for a stellar object such as its effective temperature $T_{\rm eff}$, age, and chemical element composition, as well as other inferred properties such as redshifts of distant sources and object classification. We collectively refer to all of these parameters as astrophysical parameters (APs). In the context of *Gaia* (Gaia Collaboration 2016, 2018, 2023g), APs are complementary to multi-dimensional position and velocity information for achieving a better understanding of the dynamical evolution of the Milky Way. Characterisation of a significant sample of the stars of our Galaxy also allows studies of individual stellar populations, stellar systems including planets, and a better understanding of the structure and properties of stars themselves. *Gaia* also observes objects both within our own Solar System and beyond the Milky Way, and characterising these objects in a homogeneous way promises to open new windows (Gaia Collaboration 2023b; Tanga et al. 2023; Ducourant et al. 2023).

The *Gaia* Data Processing and Analysis Consortium (*Gaia*-DPAC) is tasked with the analysis of *Gaia* data to provide a catalogue of astrometric, photometric, and spectroscopic data to the public. The role of the Coordination Unit 8 (CU8), 'Astrophysical Parameters', is to provide a catalogue of derived APs to the community based on the *mean* astrometric, photometric,

⋆ Corresponding author: O. L. Creevey, e-mail: ocreevey@oca.eu

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

and spectroscopic data. The Astrophysical Parameters Inference System (Apsis) is the pipeline that was designed and is executed at the Data Processing Center CNES (DPCC), Toulouse, France, which produces APs for all sources in the *Gaia* catalogue. These APs are not only destined for *Gaia* releases, but they are also used internally in DPAC systems, for example for determining the radial velocity (RV) template in the RV data reduction and analysis (Sartoretti et al. 2018).

The CU8 Apsis pipeline was first described in Bailer-Jones et al. (2013) before the launch of *Gaia*. Apsis comprises 13 modules that use different input data and/or models to produce APs for substellar objects, stars, galaxies, and quasars. In *Gaia* Data Release 2 (DR2), only two of the thirteen modules processed data to produce five stellar parameters ($T_{\rm eff}$, extinction A_G, colour-excess $E(G_{\rm BP}-G_{\rm RP})$, radius R, luminosity L) based on parallaxes and integrated photometry (Andrae et al. 2018). Now, in *Gaia* Data Release 3 (DR3), all of the 13 Apsis modules processed data and have contributed to the catalogue to provide 43 primary APs along with auxiliary parameters that appear in a total of 538 archive fields.

APs produced by CU8 appear in ten tables of the *Gaia* archive, with a subset of these also appearing in `gaia_source`. These data comprise both individual parameters (in four tables) and multi-dimensional data (in six tables). The individual parameters are properties such as atmospheric parameters, evolutionary parameters, chemical element abundances, and extinction parameters for stars, along with class probabilities and redshifts of distant sources. The multi-dimensional data comprise a self-organising map (SOM) of outliers, with prototype spectra, a 2D total Galactic extinction map at four healpix levels, as well as an optimal-level map, and Markov chain Monte Carlo samples for two of the Apsis modules.

The goal of the present paper is to describe the production and content of the CU8 data products available in the *Gaia* DR3 archive. More details on the validation and use of the stellar and non-stellar products can be found in the accompanying Papers II and III, respectively. More complete descriptions of specific products and methods can be found in the following papers: Delchambre (2018), Andrae et al. (2023), Lanzafame et al. (2023) and Recio-Blanco et al. (2023), and the official online documentation[1]. The AP content of *Gaia* DR3 represents one of the most extensive homogeneous databases of APs to date for exploitation in many domains of astrophysics; see for example Gaia Collaboration (2023b,c,d,e,f).

The paper is structured as follows: Sects. 2–4 describe the input data, provide an overview of the methods used in Apsis, and describe the stellar models, respectively. Section 5 describes the general content and scope of the ten *Gaia* DR3 archive tables with CU8 parameters and contains useful reference tables for guidance, while Sect. 6 describes all of the APs from CU8 grouped by astrophysical category. An overview of the validation process is described in Sect. 7, while readers are referred to the accompanying Papers II and III for detailed validation of the Apsis results. In Sect. 8 we describe the main caveats and known issues, and we conclude in Sect. 9. The Appendix contains additional information on the empirical methods that were employed, use of the multi-dimensional tables, selection function information, and some tools that have been made available to the community to aid in the exploitation of these products.

[1] https://gea.esac.esa.int/archive/documentation/GDR3/index.html

2. Input data

The results from Apsis in *Gaia* DR3 are based solely on *Gaia* input data, and these are described in this section. Figure 1 illustrates the input data that are used by the different modules in Apsis.

2.1. Input astrometry and photometry

We used the proper motions and parallaxes from *Gaia* in the processing of some of the Apsis modules. As some stellar-based modules are sensitive to the parallax zero point, we implemented the systematic correction to the parallaxes as proposed by Lindegren et al. (2021), who report biases that vary with magnitude, colour, and ecliptic latitude.

Some of the Apsis modules use integrated photometry in the G, $G_{\rm BP}$, and/or $G_{\rm RP}$ bands, using the zero points provided directly by the Coordination Unit 5 (CU5) in *Gaia* eDR3 (Riello et al. 2021). They also recommended correcting some of the *Gaia* eDR3 photometry, and this was implemented in our processing. This same correction to the eDR3 photometry has been fixed in *Gaia* DR3 (Gaia Collaboration 2021a). Figure 2 shows the distribution in G magnitude of all of the products from CU8 from the four individual parameter tables.

2.2. RVS spectra

Some of the products from CU8 are based on the Radial Velocity Spectrometer (RVS) spectra that are processed by Coordination Unit 6 (CU6, Seabroke et al., in prep.). The CU6 pipeline provides wavelength-calibrated epoch spectra using standard spectroscopic techniques. The mean spectrum can therefore be obtained by a simple stacking of the spectra. CU8 processed these mean spectra as provided by CU6. A fraction of these spectra result from a deblending process of overlapping sources. All spectra were corrected for the radial velocities of the stars; they were also cosmic-ray-clipped, normalised at the local (pseudo-) continuum ($T_{\rm eff} \geq 3500$ K), and re-sampled from 846 to 870 nm with a spacing of 0.01 nm. The median resolving power $R = \lambda/\Delta\lambda = 11\,500$ (Cropper et al. 2018). Figure 3 shows examples of input spectra (black) of different $T_{\rm eff}$ and identifies the main spectroscopic features. Some fits to models are also shown in orange. These figures are further described in Sects. 3 and 6. A more detailed description of the RVS data and their treatment is provided by Seabroke et al. (in prep.).

While over 37 million combined spectra were available to CU8, the median signal-to-noise-ratio (S/N) is ~6.5. The Apsis modules processing RVS data then applied their own S/N thresholds and quality checks before processing. Therefore, in practice, only about 10 million spectra were processed, while, after applying the module-specific post-processing filters, approximately 6.3 million of these led to published astrophysical parameters. The G magnitude range covered by the remaining data varies from 2 to 15.2 mag.

2.3. BP and RP spectra

Most of the Apsis modules produce APs based on the mean blue and red prism spectra (BP and RP, respectively), which are available for all of the sources in *Gaia* DR3. Examples of these prism spectra are shown in Fig. 4 for stars with different $T_{\rm eff}$ and extinction (A_0). These mean low-resolution spectra ($20 \leq R \leq 60$ for BP, $30 \leq R \leq 50$ for RP, see Fig. 18 of Montegriffo et al. 2023) allow us to extract the atmospheric parameters ($T_{\rm eff}$, log g,

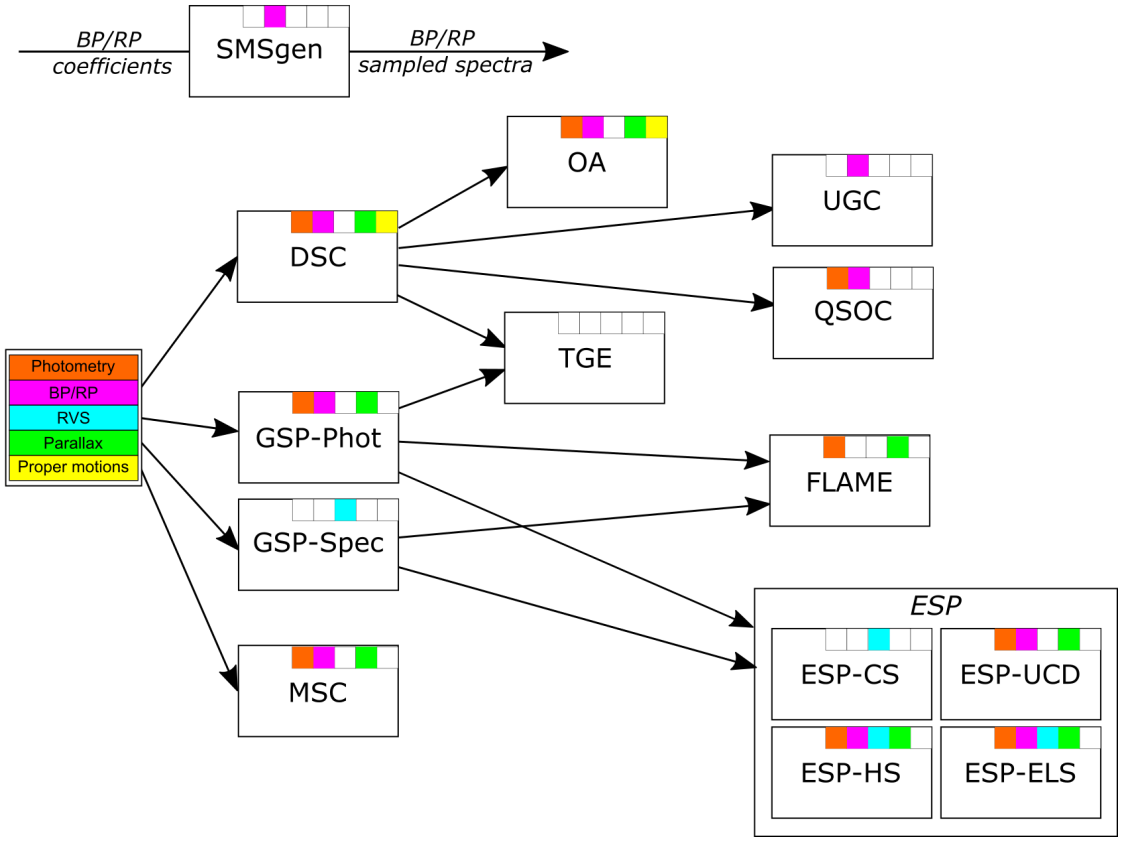


Fig. 1. Apsis workflow showing the input data (colour coded) used by the 13 modules producing APs in *Gaia* DR3 along with the dependencies among these modules (arrows). The input BP/RP spectra in Apsis are in the form of sampled spectra produced by SMSGen; see Fig. 5.

A_G, [M/H]), but are also of sufficient resolution to explore specific features such as the Hα line for emission-line stars (ELSs) and extragalactic objects. The RP spectra of very cool stars also show molecular absorption bands from TiO and VO (e.g., Reiners et al. 2007). The BP and RP spectra are processed by CU5 and then adapted within the Apsis pipeline in the form of sampled spectra, as explained below.

2.3.1. Production of data by CU5

The production of internally calibrated mean BP/RP spectra by CU5 is described in detail in Carrasco et al. (2021) and De Angeli et al. (2023). We emphasise that these mean BP/RP spectra are averaged over time, which means that any intrinsic variability of sources is lost. One should be aware of this point when using APs from Apsis for stars with important variability. Due to varying geometry over the field of view, occasional suboptimal centring of the window on the observed target and variations of the instrument response, and optics across the focal plane and in time, the epoch spectra of all transits of a given source cannot simply be stacked[2]. Instead, they need to be carefully calibrated, resulting in each epoch spectrum having its own pixel sampling. The combined mean spectrum is a continuous mathematical function that can be evaluated at any pixel position. For this, CU5 adopts a linear basis representation in terms of Gauss-Hermite polynomials. The resulting expansion coefficients and their covariance matrix are the fundamental CU5 data products and are the input to the Apsis pipeline.

2.3.2. Sampled Mean Spectrum Generator

The modules in the Apsis pipeline use the internally calibrated mean BP/RP spectra in the format of sampled spectra (integrated flux vs. pixel). Computing these sampled spectra from the CU5 coefficients is the task of the Sampled Mean Spectrum Generator (SMSGen). To this end, SMSGen takes the CU5 definition of the basis functions and integrates the spectral flux densities for a fixed wavelength grid. This wavelength grid defines 120 pixels for each BP and RP spectrum that cover the range of non-zero transmission in each spectrum as shown in Fig. 5[3]. Here, we use the most recent eDR3 passbands[4]. The wavelength sampling is approximately uniform in pixel space but non-uniform in wavelength. SMSGen then numerically integrates the flux densities in order to obtain integrated fluxes in each pixel. We note that BP/RP spectra can exhibit non-zero flux in pixels which have no transmission because of the LSF smearing effect of the BP/RP prisms, although this is negligible in practice. In any case, Apsis modules using BP/RP spectra discard several pixels at the edges that typically have very low flux and very low S/N.

The sampling process of the CU5 basis functions as well as the flux integration are strictly linear operations. Consequently, SMSGen can easily propagate the CU5 uncertainty estimates on the coefficients into uncertainties of the sampled BP/RP

[2] Other normal ageing effects impacting all observations are the contamination level, the radiation damage, the changes in small-scale effects such as hot or cold columns, the changes in CCD, and filter response.

[3] The sampling scheme is available at https://www.cosmos.esa.int/web/gaia/dr3-aps-wavelength-sampling

[4] https://www.cosmos.esa.int/web/gaia/edr3-passbands

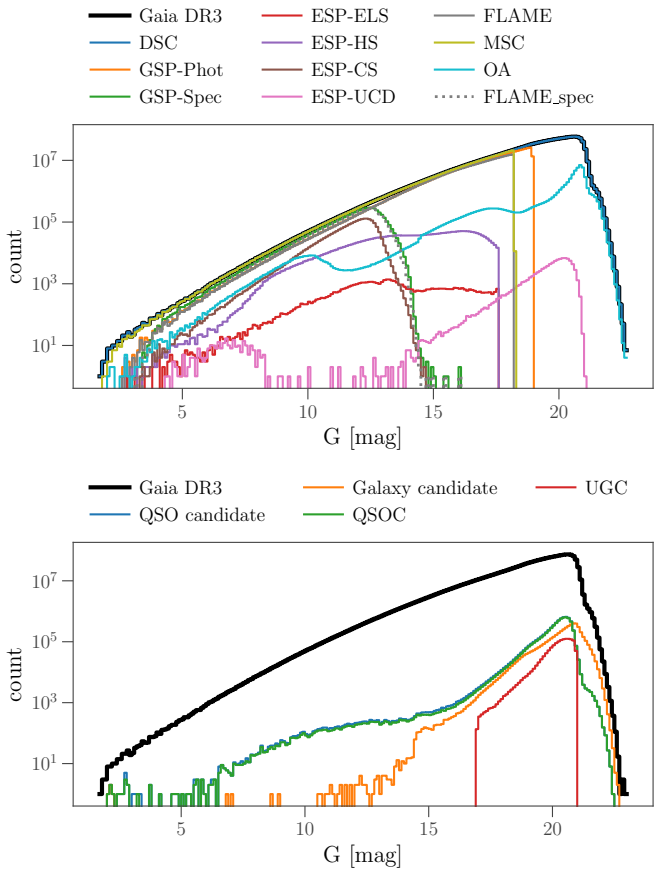


Fig. 2. *G*-magnitude distribution of the sources processed by CU8. *Top*: distribution of sources that appear in the `astrophysical_parameters` and `astrophysical_parameters_supp` tables grouped by module, illustrated by the different colours. The `astrophysical_parameters_supp` table contains results from GSP-Phot, GSP-Spec, and FLAME only, and only those from FLAME are indicated in this top panel by the dashed lines, because the distributions in both tables are identical for GSP-Phot and GSP-Spec. *Bottom*: sources with a CU8 result in the `qso_candidate` or `galaxy_candidate` tables (blue/orange) and the sources in those tables with a redshift from QSOC (green) or UGC (red).

spectrum. However, Apsis modules currently ignore any correlations between pixels, and so SMSGen only provides standard deviations for the flux uncertainties of each pixel. This approximation ensures lower computational cost, which was a limiting factor during CU8 operations. Unfortunately, as illustrated in Fig. 6, notable long-range correlations between pixels do exist in BP and RP spectra (see Babusiaux et al. 2023). Ignoring these correlations therefore causes several Apsis modules to systematically underestimate the uncertainties in their parameters, although most modules have inflated their uncertainties to account for this effect.

3. Parameter estimation methods

The Apsis chain produces all of the data from CU8 for *Gaia* DR3. Apsis is composed of 14 modules, 13 of which produce data for the release. All of the modules are described individually and in more technical detail in Sect. 3 of Chapter 11 *Astrophysical Parameters* in the online documentation for *Gaia* DR3. The first module providing the BP/RP spectra in the CU8 format (SMSGen) is summarised above in Sect. 2.3.2. Here, we provide a brief overview of the other modules in order for the reader to gain a basic understanding of the underlying methods, along with the dependencies among modules and dependencies on models and training data. Both Fig. 1 and Table 1 provide an overview of these details, which together describe the different categories of parameters, the object type, the CU8 and non-CU8 input data, the dependencies, the models and training data that are used, the approximate number of sources in *Gaia* DR3, and their *G* magnitude range for which a result can be found; see also Fig. 2 for the distribution of *G* magnitude. In addition, Fig. 7 shows a Hertzsprung-Russell diagram (HRD) illustrating the parameter spaces in which the different stellar-based modules are applied. The background HRD is a representative random sample of 10 million $T_{\rm eff}$ and M_G from Apsis.

3.1. Discrete Source Classifier

The Discrete Source Classifier (DSC; Sect. 11.3.2 of the online documentation; Delchambre et al. 2023; Bailer-Jones 2021) classifies sources probabilistically into five classes, namely quasar, galaxy, star, white dwarf, and physical binary star, although it is primarily intended to identify extragalactic sources. DSC comprises three classifiers: (1) Specmod, an ExtraTrees method using the BP/RP spectrum; (2) Allosmod (Bailer-Jones et al. 2019), a Gaussian Mixture Model using several photometric and astrometric features; and (3) Combmod, which combines the probabilities from the other two classifiers. The classes are defined empirically. DSC incorporates a global class prior that reflects the intrinsic rareness of extragalactic objects. All classifiers produce posterior class probabilities.

3.2. Outlier analysis

The Outlier Analysis (OA) module (Sect. 11.3.12 of the online documentation), aims to complement the overall classification performed by the DSC module by processing those sources with the lowest combined classification probabilities from DSC. In order to analyse outliers, the OA performs an unsupervised classification (clustering) by means of SOMs (Kohonen 2001), grouping similar objects according to their BP/RP spectra. Each group of similar objects is referred to as a neuron. In addition, the OA characterises each neuron by reporting statistics of various parameters within them, such as magnitudes, Galactic latitudes, parallaxes, and number of transits.

3.3. Unresolved Galaxy Classifier

The Unresolved Galaxy Classifier (UGC; Sect. 11.3.13 of the online documentation; Delchambre et al. 2023; Gaia Collaboration 2023b) is designed to estimate the redshift of unresolved galaxies observed by *Gaia*. The module processes every source that has a combined probability of greater than or equal to 0.25 of being a galaxy according to the DSC, that is, `classprob_dsc_combmod_galaxy` ≥ 0.25, and which has a magnitude within the range $13 \leq G \leq 21$ (after postprocessing there are no results with $G < 15$). The UGC predicts the redshift of the source by applying a supervised machine learning model based on support vector machines (SVM, Cortes & Vapnik 1995) to its sampled BP/RP spectrum. The module is trained on a set of *Gaia* spectra of galaxies with redshifts provided by an external catalogue (see Appendix A.2) and predicts redshifts in the range $0.0 \leq z \leq 0.6$.

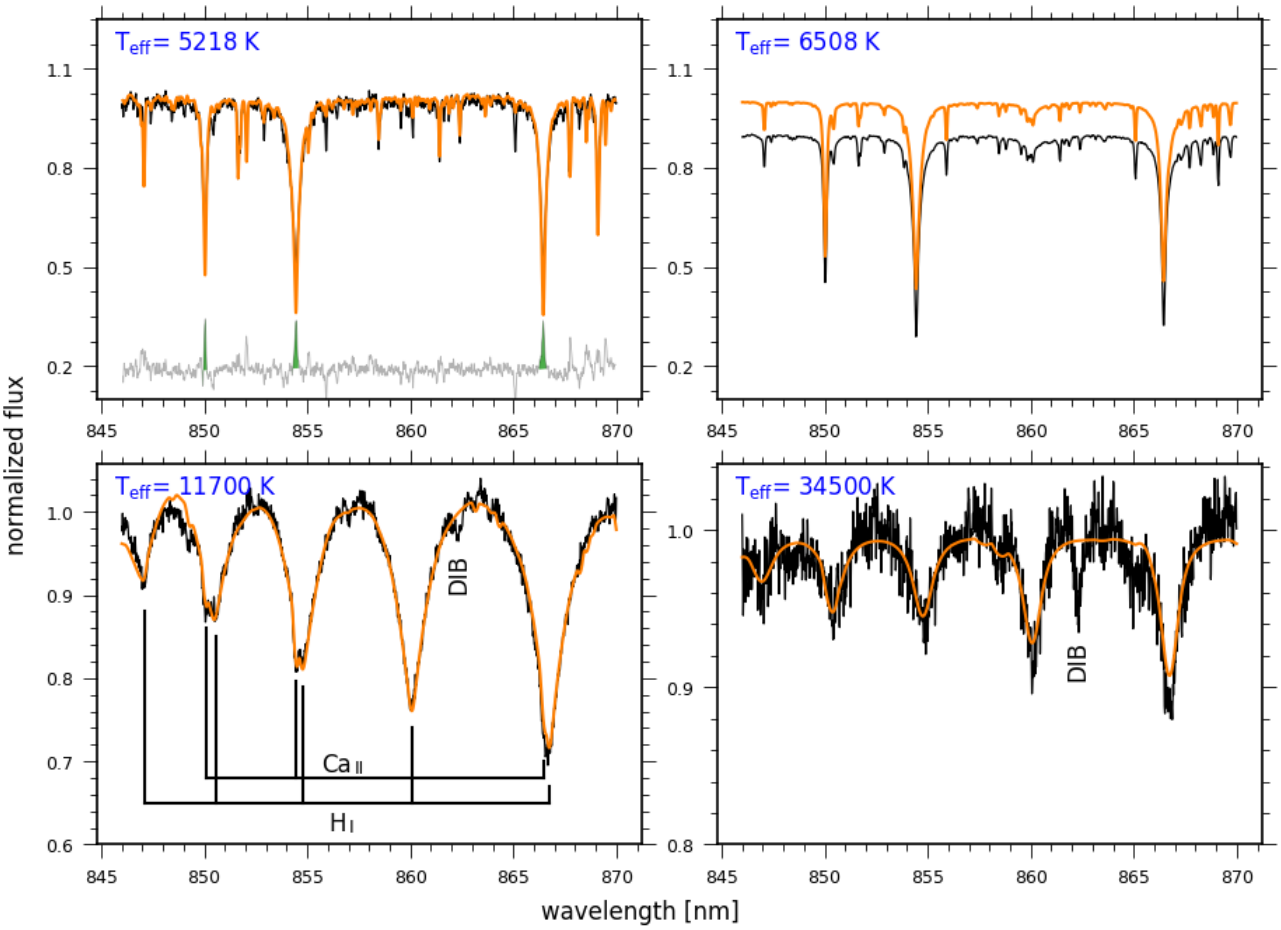


Fig. 3. Examples of the observed RVS spectra (black curve) analysed by various modules of the Apsis pipeline. The effective temperatures estimated by GSP-Spec (*upper panels*) and by ESP-HS (*lower panels*) are given in blue, while the best-fitting synthetic spectrum is shown in orange. *Upper left panel*: adopting the APs by GSP-Spec (orange spectrum), ESP-CS derives an activity index from the residuals (grey lines: residuals vertically shifted by +0.2) summed up around the calcium triplet line cores (shaded green area). *Upper right panel*: synthetic and observed (shifted by −0.1 for readability) spectrum corresponding to the GSP-Spec APs. The spectrum is then used to derive chemical abundances. *Lower panels*: determination of APs of stars hotter than 7500 K, by analysing the RVS and BP/RP data and assuming a solar chemical composition using ESP-HS. We overplot the λ862 nm DIB which is also measured by GSP-Spec.

3.4. QSO Classifier

The Quasi-Stellar Objects Classifier (QSOC; Sect. 11.3.14 of the online documentation; Delchambre et al. 2023; Gaia Collaboration 2023b) is designed to determine the redshift of the sources that are classified as quasars by the DSC module, though it uses a loose cut of `classprob_dsc_combmod_quasar` $\geq$ 0.01 in order to be as complete as possible. The method is based on a chi-square approach whereby the cross-correlation function between a rest-frame quasar template and an observed BP/RP spectrum is evaluated at a range of trial redshifts. The module predicts redshifts in the range $0.0826 < z < 6.1295$ and also provides an uncertainty and quality measurements from which flags are derived.

3.5. General Stellar Parametrizer from photometry

The General Stellar Parametrizer from photometry, GSP-Phot (Sect. 11.3.3 of the online documentation; Andrae et al. 2023; Liu et al. 2012; Bailer-Jones 2011) estimates effective temperature $T_{\rm eff}$, logarithm of surface gravity log g, metallicity [M/H], absolute magnitude M_G, radius R, distance r, line-of-sight extinctions A_0, A_G, $A_{\rm BP}$, and $A_{\rm RP}$, and the reddening $E(G_{\rm BP} - G_{\rm RP})$ by forward-modelling the BP/RP spectra, apparent G magnitude, and parallax using a Markov chain Monte Carlo (MCMC) method. To this end, GSP-Phot employs PARSEC 1.2S Colibri S37 models (Tang et al. 2014; Chen et al. 2015; Pastorelli et al. 2020, and references therein) in a forward-model interpolation in order to obtain self-consistent temperatures, surface gravities, metallicities, radii, and absolute magnitudes. For full details, we refer readers to Andrae et al. (2023). GSP-Phot results come from four stellar synthetic spectra 'libraries' using different grids of atmospheric models (MARCS, PHOENIX, A stars, OB stars, see Table 1) that cover different temperature ranges. A 'best' library is recommended according to the library that achieves the highest mean log-posterior value averaged over the MCMC samples.

3.6. General Stellar Parametrizer from spectroscopy

The General Stellar Parametrizer from spectroscopy (GSP-Spec; Sect. 11.3.4 of the online documentation; Recio-Blanco et al. 2023) takes the combined RVS spectra of single stars and estimates stellar atmospheric parameters ($T_{\rm eff}$, log g, [M/H], [α/Fe]), individual chemical abundances ([N/Fe], [Mg/Fe],

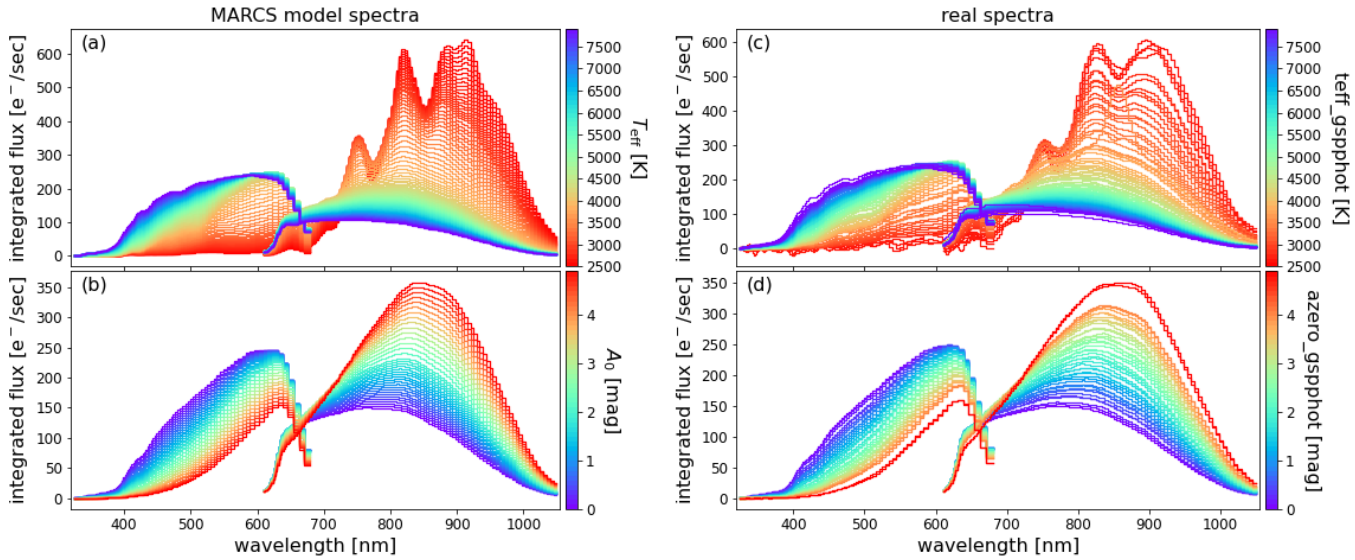


Fig. 4. Example BP/RP model spectra (*left*) and real spectra (*right*). All BP/RP spectra have been rescaled to an apparent magnitude of $G = 15$ in order to make their flux levels comparable. Panels (*a*) and (*c*) show the variation with $T_{\rm eff}$, and panels (*b*) and (*d*) show the variation with A_0. Panels (*a*) and (*b*) show synthetic BP/RP spectra based on MARCS models (see Sect. 4.1). Panels (*c*) and (*d*) show BP/RP spectra obtained by *Gaia* where the APs were produced by the GSP-Phot module in the Apsis pipeline. BP spectra approximately cover the wavelength range from 325 nm to 680 nm and RP spectra from 610 nm to 1050 nm; see Fig. 5.

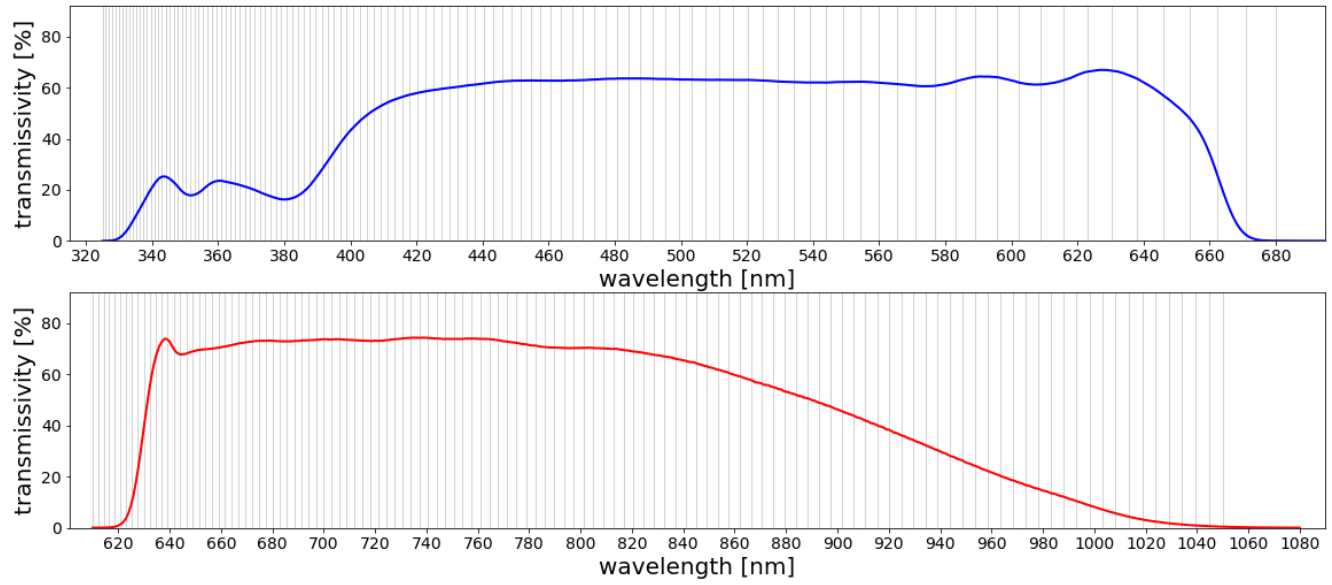


Fig. 5. CU8 sampling scheme for BP/RP spectra. Vertical grey lines show the wavelengths of 121 pixel edges defining 120 pixels for BP (*top panel*) and RP (*bottom panel*). The blue and red lines show the BP and RP transmission curves from *Gaia* eDR3, respectively.

[Si/Fe], [S/Fe], [Ca/Fe], [Ti/Fe], [Cr/Fe], [Fe/M], [FeII/M], [Ni/Fe], [Zr/Fe], [Ce/Fe], and [Nd/Fe]), diffuse interstellar band (DIB) parameters, and a CN under- and over-abundance proxy with auxiliary parameters. No additional information (astrometric, photometric, or BP/RP data) is considered, allowing a purely spectroscopic treatment. GSP-Spec uses specific synthetic spectra grids computed from MARCS models; see Sect. 4.1, and two different algorithms, Matisse-Gauguin and an artificial neural network (ANN), which are described in Recio-Blanco et al. (2016) and see also Recio-Blanco et al. (2006) for the Matisse algorithm. Both algorithms are applied for atmospheric parameter estimates. Individual abundances and DIB parameters are estimated only from the Matisse-Gauguin algorithm using the approaches described in Recio-Blanco et al. (2016) and Zhao et al. (2021), respectively.

3.7. Extended Stellar Parametrizer for emission-line stars

The Extended Stellar Parametrizer for emission-line stars (ESP-ELS; Sect. 11.3.7 of the online documentation) identifies the BP/RP spectra of ELSs brighter than magnitude $G = 17.65$. It then proposes a class label chosen among the following: Be, Herbig Ae/Be, Wolf Rayet (WC or WN), T Tauri, active M dwarf (dMe) stars, and planetary nebulae (PNe). Figure 8 shows typical BP/RP spectra of some of these classes. The module uses three Random Forest classifiers (RFCs; Sect. A.4), and

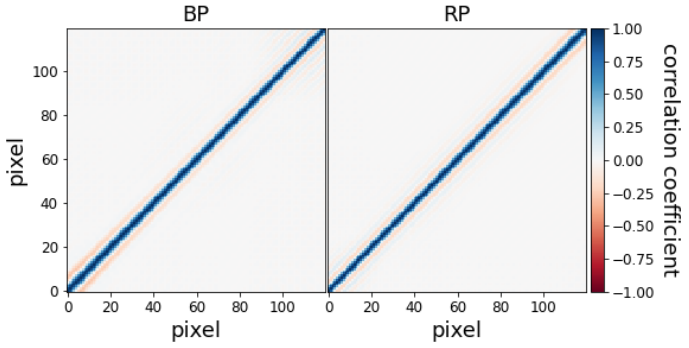


Fig. 6. Random example (`source_id`=5336426878835464960) of correlation matrices of pixel flux uncertainties for BP (*left panel*) and RP (*right panel*) for the CU8 sampling scheme shown in Fig. 5.

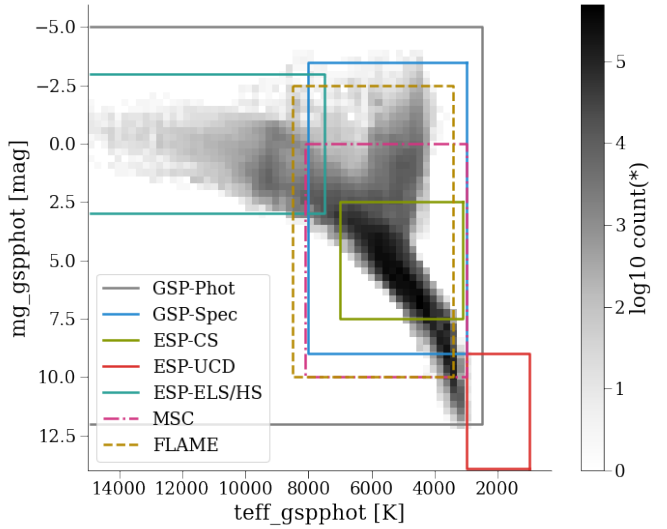


Fig. 7. HRD showing the parameter spaces covered by the different stellar modules in Apsis. The solid and dashed-dotted lines represent the modules that derive spectroscopic parameters. GSP-Phot and GSP-Spec are the general stellar parametrisers that use BP/RP and RVS data, respectively. The *esp* modules work in specific stellar regimes: the ultra-cool dwarfs (ESP-UCD), cool stars (ESP-CS), hot stars (ESP-HS), and emission-line stars (ESP-ELS). ESP-HS, ESP-ELS, and GSP-Phot provide results on stars up to 50 000 K (not shown). The green dashed-dotted line shows the regime of MSC which analyses the BP/RP spectrum as a combination of two components of an unresolved binary. The dashed line shows the parameter space of FLAME that derives evolutionary parameters only. The grey data are the stellar parameters from GSP-Phot.

a measure of the pseudo-equivalent width (pEW) of the Hα line. A first classifier (ELSRFC1) trained on synthetic BP/RP spectra is used to get a first coarse temperature estimate and assigns one of the following spectral type tags to each target: O, B, A, F, G, K, M, or CSTAR (candidate carbon star; see also Gaia Collaboration 2023c). Only non-'CSTAR' targets that received a spectral type tag are further processed by the module. The second RFC (ELSRFC2) identifies the spectra of PN and of Wolf Rayet WC and WN stars. All the targets that are not identified as PN, WC, or WN are further processed. If significant Hα emission is suspected based on the pEW value, a third RFC (ELSRFC3) is applied to the data in order to identify Be, Herbig Ae/Be, T Tauri, and dMe stars. In this process, the astrophysical parameters derived by GSP-Phot are used to help disentangle the candidate members of the four classes.

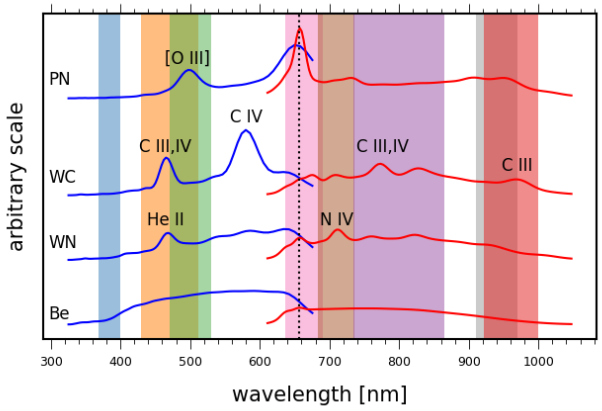


Fig. 8. BP (blue) and RP (red) spectra typical of PNe, WC or WN stars, and Be stars. The wavelength position of the strongest features is noted, while the Hα line is represented by the vertical broken line. The wavelength domains considered for ELS classification are shown with colour shades; see Sect. 6.1.3 and Appendix A.4.

3.8. Extended Stellar Parametrizer for hot stars

The Extended Stellar Parametrizer for hot stars (ESP-HS; Sect. 11.3.8 of the online documentation) derives $T_{\rm eff}$, log g, A_0, A_G, $E(G_{\rm BP}-G_{\rm RP})$, and $v \sin i$ (broadening) for stars with $T_{\rm eff}$ between 7500 K and 50 000 K, based on either BP/RP+RVS spectra or BP/RP alone by assuming solar composition for stars with $G \leq 17.65$. The target selection is based on receiving an A, B, or O spectral type tag derived by ESP-ELS (see Sect. 3.7). The BP/RP spectra (over the range from 340 to 800 nm) are compared to synthetic spectra processed by SMSGen and rebinned into 40 wavelength bins, and fit in a multi-step χ^2-minimisation. The flux uncertainties were multiplied by a factor of five to account for the amplitude of the systematic differences found between the observations and the simulations based on synthetic spectra.

We note that gravitational darkening due to rapid rotation in hot stars is expected to affect the parameter determination based on BP/RP and/or RVS spectra (e.g., Frémat et al. 2005). However, it is beyond the scope of the automatic pipeline to take these effects into account.

3.9. Extended Stellar Parametrizer for cool stars

The Extended Stellar Parametrizer for cool stars (ESP-CS; Sect. 11.3.9 of the online documentation) computes a chromospheric activity index from the analysis of the calcium infrared triplet (Ca II IRT) in the RVS spectra. The activity index is derived by comparing the observed RVS spectrum with a purely photospheric model (assuming radiative equilibrium) with $T_{\rm eff}$, log g, and [M/H] from either GSP-Spec or GSP-Phot, and from `vbroad` when available from CU6 (provided in `gaia_source`); see Fig. 3 top left panel. An excess equivalent width factor in the core of the Ca II IRT lines, which is computed on the observed-to-template ratio spectrum in a $\pm\Delta\lambda = 0.15$ nm interval around the core of each of the triplet lines, is taken as an index of the stellar chromospheric activity or, in more extreme cases, of the mass accretion rate in pre-main sequence stars.

3.10. Extended Stellar Parametrizer for ultra cool dwarfs

The Extended Stellar Parametrizer for ultra cool dwarfs (ESP-UCD; Sect. 11.3.10 of the online documentation) provides $T_{\rm eff}$ of

Gaia sources cooler than 2500 K. This is an arbitrary definition that includes stellar objects and brown dwarfs. In practice, $T_{\rm eff}$ predictions of up to 2700 K have been included in the catalogue in order to accommodate uncertainties. As UCDs are detected at very short distances, typically less than 200 pc, extinction should be very small and therefore we ignored this parameter for these objects in *Gaia* DR3. The ESP-UCD module consists of a Gaussian Process regression module that takes RP spectra as input and assigns $T_{\rm eff}$ estimates. The RP spectra used as input to the ESP-UCD module were reconstructed from the continuous representation using a truncation procedure described in Carrasco et al. (2021). We use $a = 3$ where a is the threshold coefficient in Eq. (27) of Carrasco et al. (2021).

3.11. Final Luminosity Age Mass Estimator

The Final Luminosity Age Mass Estimator (FLAME; Sect. 11.3.6 of the online documentation; Creevey & Lebreton 2022) is designed to produce the stellar mass and evolutionary parameters for each *Gaia* source that has been analysed by GSP-Phot and/or GSP-Spec; therefore FLAME produces two results for some sources. The FLAME parameters comprise the radius R, luminosity L, and gravitational redshift $rv_{\rm GR}$, along with the mass M, age τ, and evolutionary stage ϵ. FLAME uses as input data $T_{\rm eff}$, log g, and [M/H] from the GSP-Phot 'best library' and, when available, these same parameters from GSP-Spec Matisse-Gauguin, along with a distance estimate, G-band photometry (Sect. 2), and extinction from GSP-Phot. A bolometric correction is evaluated on a grid of models; see Sect. 4.3. To infer M, τ, and ϵ, the BaSTI[5] (Hidalgo et al. 2018) solar-metallicity stellar evolution models are employed, which consider a mass range of 0.5–10 $M_{\odot}$ and evolution stages from the zero-age main sequence (ZAMS) until the tip of the red giant branch (RGB).

3.12. Multiple Star Classifier

The Multiple Star Classifier (MSC; Sect. 11.3.5 of the online documentation) infers stellar parameters by assuming the BP/RP is a composite spectrum of an unresolved coeval binary system and that the two components have a flux ratio in the BP/RP spectrum of between 1 and 5. The primary is defined as the brighter source in the BP+RP spectrum total flux. The MSC uses an empirical BP/RP model (Appendix A.5) within an MCMC method to sample the posterior over its parameter space: $T_{\rm eff}$ and log g of its primary and secondary components, as well as a common metallicity, extinction, and distance. The MSC produces results for all sources with BP/RP spectra, a parallax, and $G \leq 18.25$.

3.13. Total Galactic Extinction

The total Galactic extinction (TGE; Sect. 11.3.11, Delchambre et al. 2023) module uses a subset of giants with extinction estimates provided by GSP-Phot as extinction tracers to construct all-sky maps at various resolutions of the total foreground extinction from the Milky Way. The maps specify the median extinction A_0 of the tracers per HEALPix, where A_0 is the extinction parameter of the adopted extinction curve of Fitzpatrick (1999); see Sect. 4.2 for details. Sky coverage is 97.2% at HEALPix level 6 (0.84 square degrees per HEALPix), with missing extinction estimates for some HEALPixes at Galactic latitude $|b| < 5°$. Sky coverage is less at higher resolution because of the limited number of tracers per HEALPix.

[5] http://basti-iac.oa-abruzzo.inaf.it

4. Models and training data

The Apsis modules require models and training data to infer APs. In this section, we describe these auxiliary data.

4.1. Synthetic spectra

For the estimation of stellar APs, extensive synthetic spectral libraries based on atmospheric models were computed for the G, $G_{\rm BP}$, and $G_{\rm RP}$ filter ranges and the BP/RP and RVS wavelength ranges. These libraries were used to simulate *Gaia*-observed spectra through the *Gaia* instrument models, with noise and extinction added (see Sect. 4.6 and Montegriffo et al. 2023).

4.1.1. Synthetic fluxes for BP/RP

Stellar fluxes have been simulated using standard 1D stellar-atmosphere codes, covering all spectral types of normal stars. Several grids were produced by different code families, each different in physics and assumptions, with large overlaps in the parameter space. The providers of these libraries were free to compute models following their own expertise and preferences while paying attention to the challenges of the respective stellar types (e.g., dust formation, molecular absorption, treatment of convection, chemical peculiarities, departures from local thermodynamic equilibrium (LTE), and stellar winds). For example, models for OB-type stars take into account non-LTE effects both in the computation of the model and of the spectrum. For the MARCS models (Gustafsson et al. 2008), the chemical abundances compared to the Sun have been varied over several orders of magnitude by enhancing or reducing all metals (atomic mass $A > 4$) with α-elements roughly following the Galactic trend changing linearly from $[\alpha/{\rm Fe}] = 0.0$ at $[{\rm Fe/H}] = 0.0$ (solar) to $[\alpha/{\rm Fe}] = 0.4$ below $[{\rm Fe/H}] = -1.0$. Some differences in the assumed solar reference composition exist between individual libraries, reflecting the choices of the modellers at the time of computation. Cool stars ($T_{\rm eff} < 4500$ K) with prominent molecular bands are sensitive to different assumptions concerning the chemical mixture. The assumed composition should therefore be considered when comparing results derived using different libraries at this low $T_{\rm eff}$.

Spacing between grid points also varies, both between and within libraries, and can be as low as 25 K in $\Delta T_{\rm eff}$ for the MARCS models (see Table 2). GSP-Phot relies on linear interpolation between grid points (for computational cost reasons). As the spectral flux does not change linearly with changes in the parameters (see e.g., Zwitter et al. 2004), finer grids will result in better performance than coarser ones.

An overview of the parameter space, the number of models, and the stellar model providers is given in Table 2, and some examples of synthetic spectra are shown in Fig. 9 for different objects. While several libraries cover the physical parameter space of horizontal-branch stars, only the ESP-HS module provides APs for these (Sect. 6.2.3). Libraries 'HotSpot' and 'WD' were finally not used for the production of the data in DR3.

The computation of each of the libraries requires basic information such as input stellar parameters, key individual abundances, and mass fractions of H, He, metals, and so on. For the MARCS, PHOENIX, and A and OB libraries, these parameter

Table 1. Input data, models, training data, data products, and dependencies of the Apsis algorithms.

Module	Category	Object	Input data		Inference model		Number	G
		type	Non-Apsis	Apsis	Models	Empirical	(millions)	
DSC	Probabilities, Classification	SEB	XP, ϖ, pm, G			ExtraTrees/ Gauss.Mix	1591	< 21
OA	Classification	O	XP	(DSC)		SOM	56	<21
UGC	Galaxy redshift	E	XP	(DSC)		SVM	2	<21
QSOC	Quasar redshift	E	XP	(DSC)			6	<21
GSP-Phot	Spectroscopic, Interstellar, evolution	SI	XP, ϖ, G G_{BP},G_{RP}		MARCS/PHOENIX OB/A/PARSEC		470	<19
GSP-Spec	Spectroscopic, Abundances, DIB	SI	RVS		MARCS		6	<16
MSC	Spectroscopic, Interstellar	BI	XP	(GSP-Phot)		ExtraTrees	350	<18.3
ESP-ELS	Probabilities, Classification, Hα pEW	S	XP, RVS, ϖ, G	GSP-Phot		Random Forest	0.01 235	<17.65
ESP-HS	Spectroscopic	S	XP, RVS, ϖ, G	ESP-ELS	A/OB		2	<17.6
ESP-CS	Spectroscopic	S	RVS	(GSP-Spec)	MARCS		2	<15
ESP-UCD	Spectroscopic	S	RP, ϖ, G			Gauss.Proc.	0.1	<21
FLAME	Evolution	S	ϖ, G	GSP-Phot	MARCS/BaSTI		280	<18.3
				GSP-Spec	MARCS/BaSTI		5	<13
TGE	Map	I		GSP-Phot			–	<19

Notes. The following notation is used: XP = BP/RP spectra (through sampled mean spectra), RVS = RVS spectra; ϖ = parallax, pm = proper motions, G = *Gaia* photometry which implies G, G_{BP}, and/or G_{RP}. Object Type is S, B, E, I, O for star, binary, extragalactic, interstellar, and outlier, respectively. Under Apsis Input Data, values in "()" mean that APs are used to initialise the analysis or as selection criteria only.

Table 2. CU8 synthetic stellar libraries list of BP/RP spectra.

Library name	N models	T_{eff} [K]	log g [dex]	[Fe/H][dex]	Provider/contact
A (LL models)	2332	6000–16 000 (250)	2.5–4.5 (0.25)	−1.5–+0.5 (0.5)	O. Kochukhov/D. Shulyak
MARCS	27951	2800–8000 (25)	−0.5–5.0 (0.5)	−5–+1 (0.25)	B. Edvardsson
PHOENIX	4651	3000–10 000 (100)	−0.5–5.5 (0.5)	−2.5–+0.5 (0.5)	I. Brott/P. Hauschildt
OB	2162	15 000–55 000 (1000)	1.75–4.75 (0.25)	0.0–0.6 (0.1)	Y. Frémat
BTSettl	2084	400–3000	−0.5–5.5	−4.0–+0.5	F. Allard(†)/L. Sarro
(HotSpot	31957	3000–7000	3–5	–0.5–+1.0	A. Lanzafame)
(WD	186	6000–90 000	7–9	+0.0	D. Koester)

Notes. Minimum grid spacing is listed in parentheses. (†)France Allard (1963–2020).

files can be retrieved from the *Gaia* DR3 auxiliary data web pages.

4.1.2. Synthetic spectra for RVS

For the parametrisation of the infrared RVS spectra within the GSP-Spec module, large grids of synthetic spectra were computed. These spectra were calculated from MARCS atmospheric models for FGKM-type stars using the TURBOSPECTRUM code (Plez 2012) and specific atomic and molecular line lists (Contursi et al. 2021). The covered parameter space of these grids is: 2600 to 8000 K for T_{eff}, −0.5 to 5.5 for log g (g in cm s^{-2}) and −5.0 to 1.0 dex for the mean metallicity, with varying α-element enrichment with respect to iron, as explained above. Individual chemical-abundance variations were also considered to derive abundances of N, Mg, Si, S, Ca, Ti, Cr, Fe, Ni, Zr, Ce, and Nd. The adopted solar abundances are those of Grevesse et al. (2007). The computation of these grids of synthetic spectra is discussed in Recio-Blanco et al. (2023).

For the other modules using RVS data (i.e. ESP-CS and ESP-HS), the same model atmosphere grids used to prepare the synthetic BP/RP spectra (Table 2) were adopted to compute the flux in the 846–870 nm wavelength domain. The library used by ESP-HS was prepared assuming a Solar chemical composition for $T_{eff} > 7000$ K, while for ESP-CS the MARCS models were considered for T_{eff} ranging from 3000 to 7000 K, log g from 3 to 5 dex, and [Fe/H] from −0.5 to +0.75.

4.2. Extinction

Observed spectra are attenuated by the amount of interstellar dust present in the line of sight between the observer and the source. In this sense, extinction can be considered an astrophysical parameter of a given source, and can be inferred

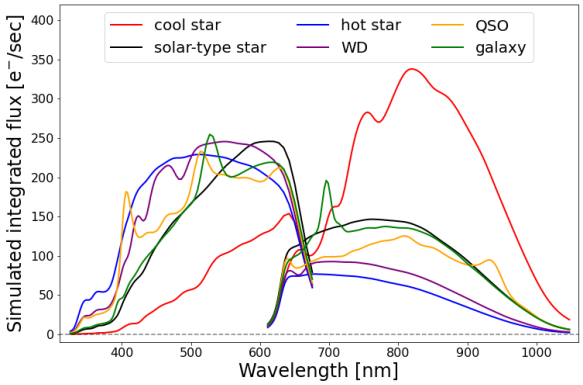


Fig. 9. Examples of BP/RP simulations of different types of sources. All sources are simulated at $G = 15$. Red and black lines show spectra from the MARCS library with $T_{\rm eff} = 3500$ and 6000 K. Blue: OB spectrum with $T_{\rm eff} = 30\,000$ K. Purple: WDA spectrum with $T_{\rm eff} = 15\,000$ K and $\log g = 8.0$. Orange and green: SDSS QSO and an SDSS galaxy, with redshifts of $z = 2.3$ and 0.06 respectively (randomly selected).

from the spectra. To estimate this parameter from the algorithms, we use simulations of the BP/RP spectra that cover a wide range of extinction values.

For Apsis simulations, we adopted the wavelength-dependent extinction law by Fitzpatrick (1999); see Sect. 11.2.3 in the online documentation. We use the parameter A_0, which is the monochromatic extinction at $\lambda_0 = 541.4$ nm[6]. A_0 and $A_{\rm V}$ are often confused in the literature, the latter being the actual extinction computed in the V band, and as such intrinsically dependent on the spectral shape of the emitting source. This dependence is often, and justifiably, neglected in the Johnson V band, but is particularly evident in the very wide *Gaia* bands and therefore should not be neglected.

Simulations are provided covering a semi-regular grid of 56 values of A_0, from 0 to 10 magnitudes, while the parameter R is kept fixed at 3.1 (see Fitzpatrick 1999, their Table 3). For each spectrum and for each A_0 the extinction in a given band (A_G, $A_{\rm BP}$, $A_{\rm RP}$) is computed by comparing the unreddened and the attenuated flux in the given *Gaia* passband. The values of extinction in these bands, and in addition in the V-band, for different APs and A_0 values are made available to the community in the parameter files (Sect. 4.1) on the *Gaia* DR3 auxiliary data web pages.

4.3. Bolometric corrections

In order to derive the bolometric luminosity of stars, specifically in the FLAME module, we complemented the observed photometric G magnitude with a bolometric correction, $\rm BC_G$. The $\rm BC_G$ was derived from the MARCS synthetic stellar spectra as a function of $T_{\rm eff}$, $\log g$, [Fe/H], and [α/Fe]. For this data release, we assumed [α/Fe] = 0.0 when calculating the correction for all stars because [α/Fe] is only estimated for a small fraction of the sources. A tool is made available to the community to calculate the $\rm BC_G$ as a function of $T_{\rm eff}$, $\log g$, [M/H], and [α/Fe] and can be found on the *Gaia* DR3 tools webpages[7].

[6] The wavelength can be derived from Table 3 in Fitzpatrick (1999), which gives $A(\lambda)/E(B-V)$ as a function of wavelength. λ_0 is the wavelength λ for which $A(\lambda)/E(B-V)/3.1$ is equal to 1.

[7] https://www.cosmos.esa.int/web/gaia/dr3-bolometric-correction-tool

Table 3. Tables in the *Gaia* DR3 archive with parameters from CU8.

Table name	Number of CU8 fields	Number of sources [millions]
astrophysical_parameters	224	1 590
astrophysical_parameters_supp	173	470
galaxy_candidates	8	2
qso_candidates	11	6
gaia_source	29	1800
		Dimensions
oa_neuron_information	77	30 × 30
oa_neuron_xp_spectra	6	87 × 900
total_galactic_extinction_map	9	4 × HP
total_galactic_extinction_map_opt	6	1 × HP9
mcmc_samples_gsp_phot	14	100 × 450M
mcmc_samples_msc	10	100 × 349M

Notes. The last six are multi-dimensional data tables. The number of sources is approximate to the nearest million. HP = 3 145 728 (=HP9), 786 432, 196 608, and 49 152 for HEALPix level 9, 8, 7, and 6, respectively. gaia_source contains a subset of parameters from the astrophysical_parameters table. The mcmc_samples_gsp_phot contains 2000 samples for sources with $G < 12$ and for a random 1% of all other sources.

We extended the $T_{\rm eff}$ range to intermediate-temperature stars using the A star models. Their $\rm BC_G$ values show a slight offset relative to the MARCS grid (due at least in part to different opacities used in the two sets of models). We therefore added an offset in magnitude units to achieve continuity at 8000 K. The adopted value for the bolometric correction for the Sun is $\rm BC_{G\odot} = +0.08$ mag, where $M_{\rm bol\odot} = 4.74$[8] which yields an absolute magnitude of the Sun $M_{G,\odot} = 4.66$ mag. We estimate an external accuracy on this zero point of ±0.015 mag from comparison with known solar analogues ($M_G = 4.63$–4.69 mag), stellar models ($M_G = 0.465$ mag), and colour transformations using Riello et al. (2021; $V - G = 0.148 \pm 0.003$ mag where $M_V = 4.817$ mag).

To complement this analysis on the solar reference magnitudes, we estimate the solar colours in Gaia Collaboration (2023c) using a set of solar analogues, although we note that these colours were not used in Apsis processing: $(G_{\rm BP} - G_{\rm RP})_\odot = 0.818 \pm 0.029$ mag, $(G_{\rm BP} - G)_\odot = 0.324 \pm 0.016$ mag, and $(G - G_{\rm RP})_\odot = 0.494 \pm 0.020$ mag.

4.4. Stellar evolution models

Stellar evolution models are used in two of the Apsis modules, GSP-Phot and FLAME. For GSP-Phot, the published APs are astrophysically self-consistent within the PARSEC 1.2S Colibri S37 models (Tang et al. 2014; Chen et al. 2015; Pastorelli et al. 2020, and references therein). Imposing these isochrones ensures that GSP-Phot can simultaneously fit the observed apparent magnitude (using the absolute magnitude) and the amplitude of low-resolution BP/RP spectra (using the radius, see Andrae et al. 2023). Moreover, the isochrones ensure that only astrophysically reasonable parameter combinations are possible.

[8] The IAU 2015 Resolution B2 can be found here: https://www.iau.org/static/resolutions/IAU2015_English.pdf

For FLAME the mass, age, and evolutionary stage are based on the use of the BASTI stellar models (Hidalgo et al. 2018). In FLAME, these models cover the ZAMS until the tip of the RGB, corresponding to evolutionary indices of between 100 and 1300 (main sequence < 390; turn-off = 390; subgiant: 420–490, and giant >490), and masses of between 0.5 and 10 $M_{\odot}$. We furthermore imposed a solar-metallicity prior; see Sect. 6.4.1 for a discussion on this assumption.

4.5. Empirical training

One of the drawbacks of training machine learning algorithms on synthetic data is that good results require (a) adequate source models from which to generate the synthetic data, (b) sufficient coverage of the parameter space by the source models, and (c) a good match between the synthetic data (*Gaia* simulations) and the real *Gaia* data of the corresponding objects. For five Apsis modules, namely DSC, MSC, UGC, ESP-UCD, and ESP-ELS, one or more of these conditions could not be achieved, and so for these we use empirical training. This involves training the algorithm on real *Gaia* data, with classes or astrophysical parameters for the training data obtained from external sources. Typically this involves cross-matching *Gaia* to external catalogues, such as the Sloan Digital Sky Survey (SDSS), and using class labels or APs obtained by others, for example from higher resolution spectra. Details of the empirical training used by the five Apsis modules are given in Appendix A.

4.6. Simulations with MIOG

The Mean Instrument Object Generator (MIOG) simulates low-resolution BP/RP spectra from given model spectral energy distributions (SEDs). This was developed by CU5 and is only available internal to DPAC systems. MIOG implements the instrument model and the dispersion law as derived by CU5 as part of the external calibration process (Montegriffo et al. 2023). This external calibration relies on the flux calibration of the spectrophotometric standard stars (SPSSs) by Pancino et al. (2012), Altavilla et al. (2015), and Marinoni et al. (2016).

All synthetic libraries described in Sect. 4.1 were simulated with MIOG. An example of the simulated spectra for stars of different $T_{\rm eff}$ and A_0 is shown in the left panels of Fig. 4. The corresponding real observed spectra are shown in the right panels. Figure 9 shows simulated stellar spectra at different temperatures (and from different libraries), together with spectra for extragalactic sources and that of a white dwarf.

CU5 provided a simplified version of this tool to the community, `GaiaXPy`, which simulates the low-resolution spectra from model SEDs, which is fully compatible with the internal DPAC MIOG simulator (Montegriffo et al. 2023).

5. Catalogue description

The astrophysical parameters produced by CU8 fall under the following categories: (a) classification products, comprising class probabilities and class labels of objects and ELSs, and stellar spectral types; (b) interstellar medium characterisation and distances, including 2D total Galactic extinction maps; (c) stellar spectroscopic and evolutionary properties, including binary star characterisation; (d) redshifts of extragalactic objects; (e) outlier analysis products; and (f) auxiliary data. Most of these products are individual parameters produced on a source-by-source basis. Multi-dimensional (MD) products are also produced, such as the two 2D total Galactic extinction maps, two dedicated outlier tables, and Markov chain Monte Carlo samples from GSP-Phot and MSC containing stellar and interstellar medium parameters and distances. All of these data products are found in one of ten tables in the *Gaia* DR3 archive, with a subset of these also copied to the main archive table (`gaia_source`); see Sect. 5.2.

5.1. Operations

The operations that were run to produce data for DR3 required a total of about 92 days of continuous processing time (1 021 219 CPU hours). This had been preceded by several month-long testing and validation runs, and allowed for sufficient post-operation validation time. With a strict delivery date for production and validation of these data of 30 June 2021, which would ensure *Gaia* DR3 in the first half of 2022, we had to impose processing limitations in some of the modules that produce stellar parameters. This was done either on an observed G magnitude basis or an RVS S/N basis. The processing limits that were imposed are the following: for GSP-Phot: $G \leq 19$, FLAME: $G \leq 18.25$, MSC: $G \leq 18.25$, ESP-ELS: $G \leq 17.65$, ESP-HS: $G \leq 17.65$, ESP-CS: $G \leq 16.62$, and for ESP-UCD, in addition to all sources with $G \leq 17.65$, we also processed a pre-defined list of around 50 million sources with $G > 17.65$. These limits in magnitude ensured roughly the same number of objects in each magnitude bin (∼130 million) and enabled the schedule to be optimised. For GSP-Spec, we imposed a minimum $S/N = 20$ in the RVS spectra. This information is also provided in Table 1.

No limit was necessary for UGC, QSOC, or OA because they process relatively few sources. DSC also had no limitations imposed because it was designed to run fast in order to process all 2 billion[9] sources in *Gaia*. The TGE module is very quick as it works on a HEALPix basis, but as it processes sources from GSP-Phot, no sources with $G > 19$ were included. In Fig. 10, we show the distribution in observed colour–magnitude $[(G_{\rm BP} - G_{\rm RP}), G]$ space for the 12 modules producing data on individual sources.

5.2. CU8 data tables in *Gaia* DR3

The names and dimensions of the tables with CU8 parameters are summarised in Table 3. The first four tables contain APs for which processing was done on a source-by-source basis, such as $T_{\rm eff}$ or redshifts. Most of the stellar parameters, classifications, individual extinction measurements, and auxiliary data are found in the `astrophysical_parameters` and `astrophysical_parameters_supp` tables, which contain only CU8 products. The former table contains one main result from each of the Apsis modules, while the latter table provides supplementary results in the form of specific libraries (GSP-Phot), methods (GSP-Spec), or input source types (FLAME). Some of the parameters from DSC and GSP-Phot from the `astrophysical_parameters` are copied to `gaia_source` for convenience to the user.

The `galaxy_candidates` and `qso_candidates` tables focus on extragalactic objects and consolidate results from different CUs. In these tables, CU8 was responsible for the galaxy and QSO redshifts produced by UGC and QSOC, respectively, along with the extragalactic class probabilities and labels from DSC.

To supplement the AP estimates from GSP-Phot and MSC, a sample of the MCMC is also provided as a

[9] The word billion implies 1000 million.

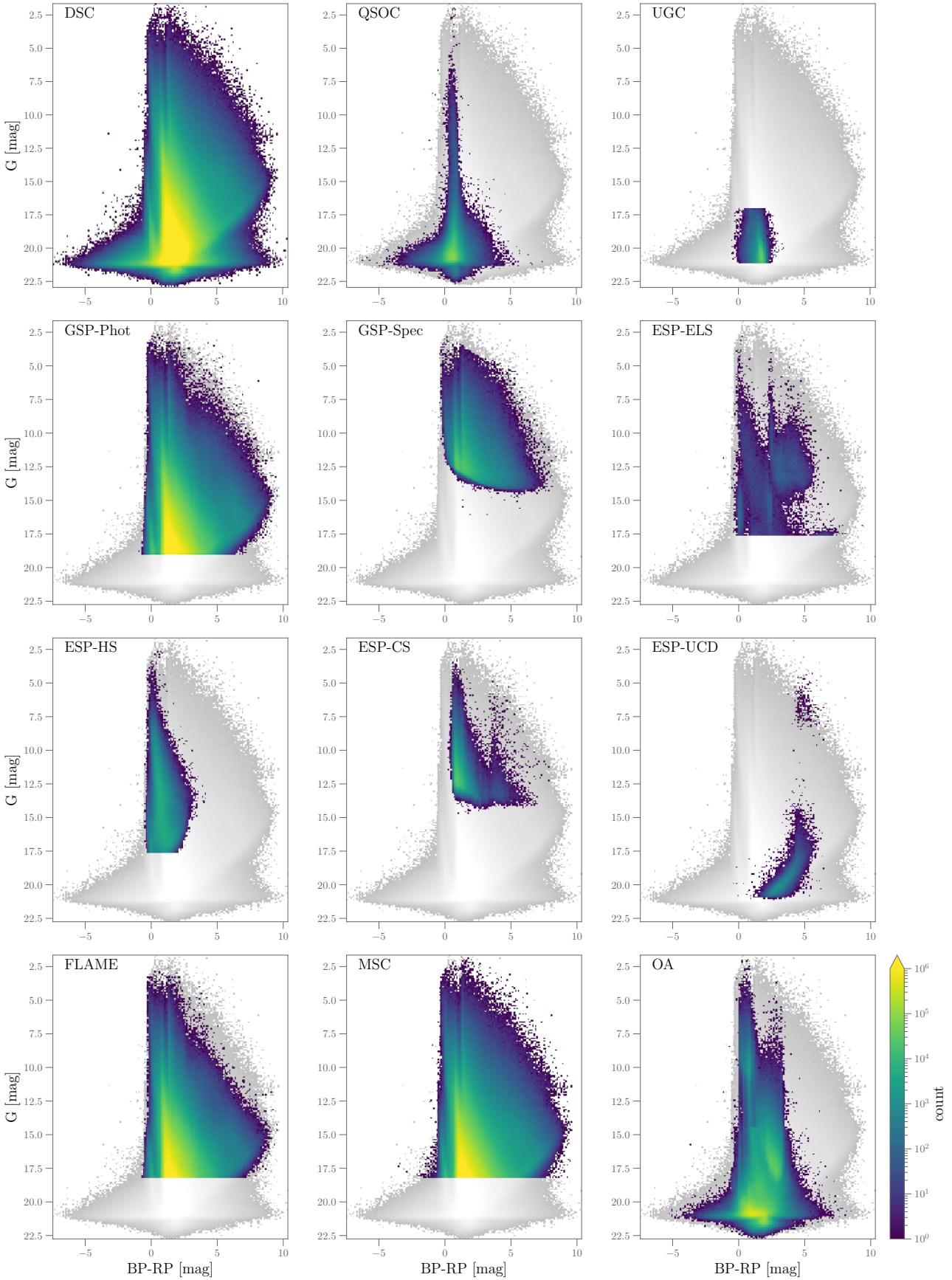


Fig. 10. Distribution in colour–magnitude space of the sources with products from CU8 in *Gaia* DR3, separated by module. The colours represent the results per module, and the colour code represents the density of sources. The distribution shown in grey in all panels indicates the whole *Gaia* DR3 sample for reference. These products are found in the `astrophysical_parameters`, `astrophysical_parameters_supp`, `galaxy_candidates`, and `qso_candidates` tables.

Table 4. Overview of the contents of each table in the *Gaia* archive containing Apsis products, organised by product type.

Product type	*Gaia* DR3 archive tables	Overview content
Classification	astrophysical_parameters	Object class probabilities (quasar, galaxy, star, white dwarf, physical binary star), emission-line class probabilities and label, spectral types
	galaxy_candidates	Galaxy and QSO class probabilities and label, outlier class label
	qso_candidates	Galaxy and QSO class probabilities and label, outlier class label
Interstellar	astrophysical_parameters	Monochromatic extinction and extinction in G_{BP}, G_{RP}, G, colour excess, distances, diffuse interstellar band characteristics
	astrophysical_parameters_supp	Monochromatic extinction and extinction in G_{BP}, G_{RP}, G, colour-excess, distances
	total_galactic_extinction_map [(a)]	Total Galactic extinction 2D map at HEALpix levels 6, 7, 8, 9
	total_galactic_extinction_map_opt [(a)]	Total Galactic extinction 2D map at HEALpix level 9 based on the optimal HEALPix level
Stellar	astrophysical_parameters	Atmospheric parameters for single and binary stars, chemical abundances, equivalent widths, rotation and activity parameters, evolutionary parameters
	astrophysical_parameters_supp	Atmospheric and evolutionary parameters
	mcmc_samples_gsp_phot [(a)]	MCMC samples of stellar parameters from GSP-Phot
	mcmc_samples_msc [(a)]	MCMC samples of stellar parameters from MSC
Extragalactic	qso_candidates	Redshifts of qso candidates
	galaxy_candidates	Redshifts of unresolved galaxy candidates
	astrophysical_parameters	Class probabilities of a source being a galaxy or quasar
Outlier analysis	astrophysical_parameters	Neuron membership and distance between source and neuron prototype BP/RP spectra
	oa_neuron_information [(a)]	Self-organising map (SOM, 30×30) of outlier sources
	oa_neuron_xp_spectra [(a)]	Prototype BP/RP spectra corresponding to each SOM neuron
Auxiliary	astrophysical_parameters	Flags, convergence indicators, bolometric correction, library name
	astrophysical_parameters_supp	Flags, convergence indicators, bolometric correction, library name
	qso_candidates	flags, quality indicators

Notes. A subset of the products from DSC and GSP-Phot also appear in gaia_source. The links in the table point directly to the online data model description. [(a)]Multi-dimensional table.

datalink product. In addition to the sampled APs, the tables mcmcc_samples_gsp_phot and mcmc_samples_msc contain the log posterior and log likelihood, meaning that the user can re-analyse the samples for their own use case; see Sect. 11.3.3 in the online documentation for details. In Appendix B we provide information on retrieving the MCMC data.

The primary result from the OA analysis is a SOM of 30×30 neurons with a statistical description of each neuron, called oa_neuron_information. Additionally, a template spectrum for each neuron is provided in the oa_neuron_xp_spectra table. For the sources identified as outliers, the astrophysical_parameters table contains the neuron membership information. Examples of how to exploit these data are given in Appendix C.

Finally, the results from TGE are given in the total_galactic_extinction_map table in the form of a 2D TGE map at 4 HEALPix levels. The additional total_galactic_extinction_map_opt table contains a HEALPix level 9 map but this is based on the *optimal* HEALPix level.

To help the user navigate to the appropriate table in the archive, Table 4 provides an overview of the contents of each table but organised by the six astrophysical parameter categories mentioned above. For example, if one is interested in classifications, Table 4 provides the link to three relevant tables: astrophysical_parameters, galaxy_candidates, and qso_candidates. In the last column, we give an overview of what type of content is found in each of those tables for that category. As another example, users interested in monochromatic extinction or extinction in the G band should query the astrophysical_parameters and astrophysical_parameters_supp tables.

5.3. Parameters and fields

In the ten archive tables (excluding gaia_source), there are a total of 538 fields produced by CU8 (excluding solution_id and source_id)[10]. Each field has a field name associated with it, along with a data type, unit, and a simple and detailed description. Some of these field names are related. For example, for the chromospheric activity index, there are three related activityindex fields: its value, uncertainty, and information pertaining to the input data. For the stellar mass, there is an upper and lower confidence level associated with the median value (three fields). Also, the mass was derived using two different sets of input data, to give a total of six mass fields. There are also some parameters that are produced by more than one Apsis module, such as class probabilities (classprob), $T_{\rm eff}$ (teff), and [M/H] (mh). We refer to activityindex, mass, and classprob for the fieldname root or parameter. To make it easier for the user to understand the AP content of *Gaia* DR3 and to understand how each of these 538 fields has been derived, the fieldname also includes the name of the Apsis module responsible for deriving that parameter. We adopted a general approach to

[10] This count includes two fields that are reproduced in three tables (classprob_dsc_combmod_quasar, classprob_dsc_combmod_galaxy in astrophysical_parameters, galaxy_candidates, qso_candidates) and three fields that are reproduced in two tables (classlabel_dsc, classlabel_dsc_joint, classlabel_oa in galaxy_candidates, qso_candidates).

Table 5. Summary of CU8 parameters in *Gaia* DR3 source-based tables.

Parameter description	Fieldname root	Number of variants	Total related fields	Maximum sources [million]
Classification				
Class probabilities	`classprob`	4	24	1591
Class labels	`classlabel`	4	7	7
Spectral types	`spectraltype`	1	1	218
Interstellar				
Extinction in G band	`ag`	7	20	470
Extinction at 541 nm	`azero`	7	20	470
Extinction in G_{BP} band	`abp`	5	15	470
Extinction in G_{RP} band	`arp`	5	15	470
Reddening	`ebpminrp`	6	17	470
Diffuse interstellar band	`dib`	1	9	0.5
Distance	`distance`	6	18	470
Stellar-spectroscopic				
Effective temperature	`teff`	9	25	470
Surface gravity	`logg`	8	23	470
Global metallicity	`mh`	8	24	470
α-element over iron abundance	`alphafe`	2	6	5
Iron and elemental abundances	`fem`, `feiim`, `xfe` [(a)]	1	65	3
Activity index	`activityindex`	1	3	3
Projected rotation velocity	`vsini`	1	2	2
Equivalent widths	`ew`	2	8	235
Binary component T_{eff}	`teff_msc1, teff_msc2`	1	6	380
Binary component log g	`logg_msc1, logg_msc2`	1	6	380
Stellar-evolutionary				
Luminosity	`lum`	2	6	280
Absolute G magnitude	`mg`	5	15	470
Radius	`radius`	7	21	470
Mass	`mass`	2	6	140
Age	`age`	2	6	130
Evolutionary stage	`evolstage`	2	2	170
Gravitational redshift	`gravredshift`	2	6	280
Extragalactic and outlier analysis				
redshift	`redshift`	2	6	6
Characteristic of outlier neuron	`neuron`	1	3	56
Auxiliary				
Logarithm of posterior probability	`logposterior`	6	6	
Logarithm of χ^2 from fit	`logchisq`	2	2	
Bolometric correction	`bc`	2	2	
Processing and quality flags	`flags`	10	10	
MCMC acceptance rate	`mcmcaccept`	6	6	
Drift of MSC MCMC chain	`mcmcdrift`	1	1	
Name of GSP-Phot best library	`libname`	1	2	
Redshift CCF ratio	`ccfratio`	1	1	
Redshift quality indicator	`zscore`	1	1	

Notes. See text for the definition of 'variants' and 'related fields' (Sect. 5.3). [(a)]We use `xfe` in this table to indicate abundances with respect to Fe – these are the abundances of 11 different elements: N, Mg, Si, S, Ca, Ti, Cr, Ni, Zr, Ce, Nd, where x in the field name is replaced by the corresponding element symbol in lower case.

naming the individual parameter fields and these mostly take the form of `parameter_module_variant_detail`[11].

Here, `parameter` is one of the 43 main parameters (not counting auxiliary data products) listed in Table 5. Subsequently, `module` is the name of the Apsis module that derived the AP; see Sect. 3. Next, `variant` describes a variant of the method, models, or input data used to derive the AP. This part may be blank if only one method was used, or if the parameter value comes from the 'best' of several methods. Finally, `detail` may be blank if the field contains the value of the AP; otherwise it takes on values such as `upper`, `lower`, or `uncertainty`, where `upper` and `lower` imply upper and lower confidence intervals (generally 68%, but see data model descriptions), or an ELS

[11] Some exceptions are: equivalent width fields from GSP-Spec, T_{eff} and log g from MSC, `classlabel_espels_flag`.

type in case of a class probability field, such as `ttauristar`. As an example, `teff_gspphot_marcs_upper` is the upper confidence level of the $T_{\rm eff}$ value estimated by the module GSP-Phot using the MARCS library of synthetic stellar spectra; `classprob_dsc_combmod_quasar` is the class probability value of being a quasar from the module DSC using the `combmod` method; or `sife_gspspec_nlines` is the number of lines used to estimate the [Si/Fe] abundance from GSP-Spec.

Table 5 describes all of the unique parameters associated with the six categories (classification, interstellar and distances, stellar-spectroscopic/evolutionary, extragalactic, outlier, auxiliary). The description of the unique parameter is given in the first column, and the field-name root (`parameter`) used in the archive field name is given in the second column. The third and fourth columns give the number of variants associated with a unique parameter and the total number of related fields, respectively. Using the above example, for `mass` these numbers are 2 and 6, respectively. As another example, for `classprob` there are four variants (three from DSC and one from ESP-ELS) for a total of 24 related fields[12], and so these numbers are 4 and 24 in the table, respectively. The final column gives the maximum number of sources for which this parameter is available. In the following section, we describe each of the `parameter_module_variant_detail` fields grouped by category.

6. Catalogue results

We describe all of the APs and other data products produced by CU8 and available in the *Gaia* DR3 archive in this section ordered according to their category: classification (Sect. 6.1), ISM and distances (Sect. 6.2), stellar spectroscopic (Sect. 6.3) and evolutionary (Sect. 6.4) parameters, extragalactic redshifts (Sect. 6.5), outliers (Sect. 6.6), and auxiliary parameters (Sect. 6.7).

6.1. Classification

Class probabilities and class labels are provided by three Apsis modules for three categories of objects: DSC provides the probabilities for all sources to belong to the classes quasar, galaxy, star, white dwarf, and physical binary star; OA classifies sources with lower probabilities from DSC; and ESP-ELS provides a spectral type classification and ELS types for stellar sources.

6.1.1. The Discrete Source Classifier

The DSC provides normalised posterior probabilities for five classes from Specmod and Combmod, and for three classes (not white dwarfs or physical binaries) from Allosmod. These are all listed in the `astrophysical_parameters` table. The Combmod probabilities for quasars and galaxies also appear in the `qso_candidates` and `galaxy_candidates` tables, and the Combmod quasar, galaxy, and star probabilities for all objects are duplicated in the `gaia_source` table. Additionally, two class labels derived from these probabilities (defined in Sect. 11.3.2 of the online documentation), `classlabel_dsc` and `classlabel_dsc_joint`, are listed in the `qso_candidates` and `galaxy_candidates` tables.

DSC Combmod and Specmod provide results for 1.59 billion sources. Allosmod has fewer sources, namely 1.37 billion, because some sources have only two-parameter astrometric solutions (i.e. they have positions but lack parallaxes and proper motions). Users can classify sources using the probabilities, either by taking the class with the largest probability or that with the probability above some threshold (in the latter case, multiple classifications or no classification is possible).

[12] Including the two fields from DSC reproduced in three archive tables.

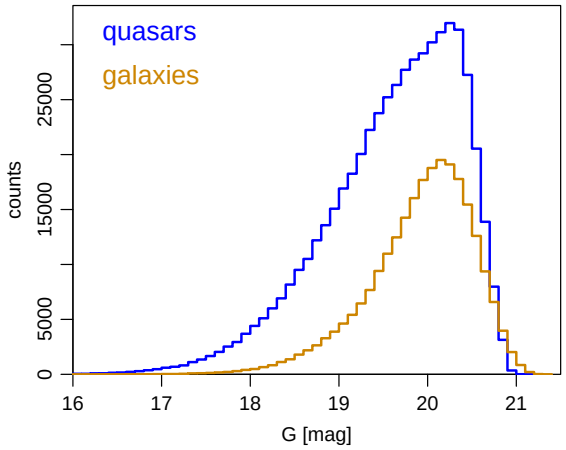


Fig. 11. *G*-band magnitude distribution of the subset of candidates in the `qso_candidates` (blue) and `galaxy_candidates` (orange) tables identified using the `classlabel_dsc_joint` field. These subsets comprise around 547 000 quasars and 251 000 galaxies.

Taking a probability threshold of 0.5 on Combmod, we obtain around 5.2 million quasars and 3.6 million galaxies, although these samples have significant contamination. More complete numbers are given in Table 11.16 in the online documentation. Most objects in *Gaia* are of course stars, and so the star class is of little use in practice. Performance on white dwarfs and physical binaries is poor (the purities are low), and we recommend against using their probabilities for building samples.

The purity and completeness of samples vary with probability threshold, magnitude, and Galactic latitude (and other parameters). Assessments of the purity and completeness are given in Sect. 11.3.2 of the online documentation (summarised in this table), as well as in Gaia Collaboration (2023b), and in more detail in Bailer-Jones (2021). We see there that Specmod, Allosmod, and Combmod show rather different performances, and so users may want to select using one or the other depending on their goals. More advice on the use of the DSC results and the (non-trivial) interpretation of its performance can be found in Delchambre et al. (2023) and Sect. 11.3.2 of the online documentation. The label `classlabel_dsc_joint` in the `qso_candidates` and `galaxy_candidates` tables identifies a set of extragalactic sources with purities of around 63%, increasing to around 83% for the subsets more than 11.5° from the Galactic plane. Their magnitude distributions are show in Fig. 11.

6.1.2. Outlier Analysis

For *Gaia* DR3, OA processed around 56 million objects whose *G* magnitudes peaked around 20.8 mag, which are in general faint stars and extragalactic objects. OA provides an unsupervised classification that complements the one produced by DSC; it does this by analysing the sources with the lowest classification probability from DSC and produces a SOM (see Sect. 6.6) with 900 (30 × 30) neurons; see e.g., Fig. 12. An object belonging to any of the 900 neurons can be found in the `astrophysical_parameters` table by the `neuron_oa_id`. The associated parameters indicate how close the source is to the neuron prototype and its ranking in distance to that prototype, `neuron_oa_dist` and `neuron_oa_dist_percentile_rank`.

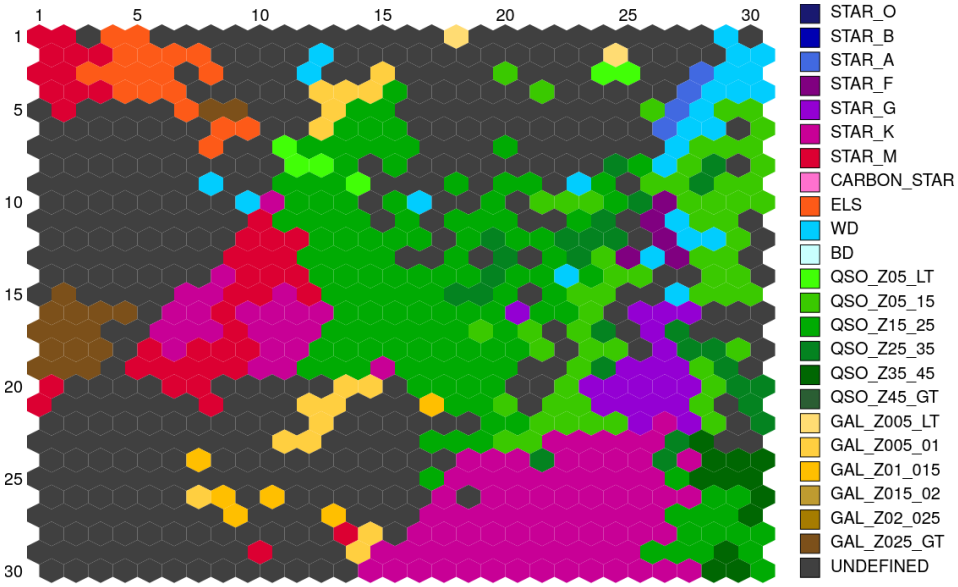


Fig. 12. SOM map lattice visualised using the GUASOM tool (Álvarez et al. 2021) representing the specific class labels assigned to each neuron by the OA module. The OA module analysed the 56 million sources with the poorest classification probabilities from DSC. Those neurons for which such a label cannot be attributed remain 'undefined'.

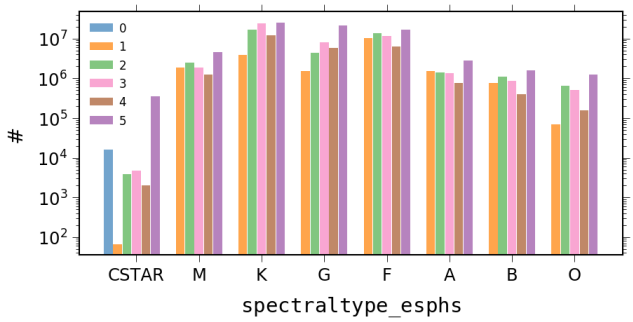


Fig. 13. Histogram of the distribution of `spectraltype_esphs` which processed sources with $G \leq 17.65$. A coloured distinction is made between the different values taken by the quality assessment flag (second digit of `flags_esphs`). Usually, the flag takes values ranging from 1 to 5, with the lower value indicating higher quality. However, for the `CSTAR` tag, this value can also be '0'.

More information on OA and its multi-dimensional data is given in Sect. 6.6. Some examples of exploiting these data are given in Appendix C.

6.1.3. Extended Stellar Parametrizer for emission-line stars

ESP-ELS provides one of the following spectral type tags `spectraltype_esphs`[13] for 218 million targets with $G \leq 17.65$: CSTAR, M, K, G, F, A, B, or O (see Table A.1). An indicator of the spectral tag quality is stored in the second digit (reading from left to right) of `flags_esphs`. In most cases, its value ranges from 1 to 5 (the lower, the better) and is based on the relative value of the first and second highest probabilities. Value 0 was added during the validation to identify those candidate carbon stars (`CSTAR` tag) with BP/RP spectra with significantly stronger C2 and CN molecular bands than in 'normal' stars (Gaia Collaboration 2023c). The distribution of the spectral types according to the quality flag is shown in Fig. 13. As can be seen, only the 'CSTAR' type has a value of 0.

The module also identified 57 511 ELSs, for which it suggests a stellar class (`classlabel_espels`) based on the combined probabilities (e.g., `classprob_espels_wcstar`) provided by two Random Forest classifiers. The ELS classes that were considered, as well as the corresponding `classlabel`, are: Be stars (`beStar`), Herbig Ae/Be stars (`HerbigStar`), T Tauri stars (`TTauri`), active M dwarf stars (`RedDwarfEmStar`), Wolf-Rayet WC (`wC`) and WN (`wN`), and planetary nebula (`PlanetaryNebula`)[14].

6.2. Interstellar medium characterisation and distances

The second category of astrophysical parameters concerns the characterisation of the interstellar medium (ISM) and distances. Source-based ISM characterisation is provided by GSP-Phot, ESP-HS, and MSC as one of the spectroscopic parameters estimated from BP/RP spectra (A_0, A_G, A_{BP}, A_{RP}, $E(G_{BP} - G_{RP})$) and by GSP-Spec based on the analysis of the λ862 nm DIB. The TGE module exploits individual source-based extinction from GSP-Phot to provide a 2D TGE map. Both GSP-Phot and MSC additionally estimate distances. Further details on most of these parameters are found in Fouesneau et al. (2023), while TGE is discussed in Delchambre et al. (2023).

6.2.1. General Stellar Parametrizer from photometry

GSP-Phot estimates the monochromatic extinction A_0, called `azero_gspphot`, for all processed sources by fitting the observed BP/RP spectrum, parallax, and apparent G magnitude. GSP-Phot also estimates the broad-band extinctions A_G, A_{BP}, and A_{RP}. The latter are not free fit parameters but are instead obtained from integrating attenuated model SEDs (see Sect. 11.2.3 of the online documentation). Using these extinction estimates, one can also compute reddenings, for example $E(G_{BP} - G_{RP}) = A_{BP} - A_{RP}$. These extinction and reddening estimates along with upper and lower confidence levels are available in the `astrophysical_parameters` table (A_0, A_G, A_{BP}, A_{RP}, $E(G_{BP} - G_{RP})$) from the best library, that is, the library that produced the highest posterior

[13] ESP-HS and ESP-ELS are related modules, and this field was originally produced by ESP-HS at the time of the archive data model definition, but it was finally produced by the upstream ESP-ELS module.

[14] The `classprob_espels_*` corresponding names are `bestar`, `herbigstar`, `ttauri`, `dmestar`, `wcstar`, `wnstar`, `pne`.

probability for that source; see libname_gspphot. The astrophysical_parameters_supp table contains the five ISM parameters A_0, A_G, A_{BP}, A_{RP}, $E(G_{BP} - G_{RP})$ for the individual library results (MARCS, PHOENIX, A, OB). GSP-Phot additionally derives a distance estimate to be consistent with the inferred parameters. The parameters azero_gspphot, ag_gspphot, ebpminrp_gspphot, and distance_gspphot and their upper and lower confidence levels are copied from the astrophysical_parameters table to the gaia_source table for convenience to the user. A sample of the MCMC from GSP-Phot inference is also made available as a datalink product.

6.2.2. Multiple Star Classifier

Like GSP-Phot, MSC also estimates the A_0 parameter, but by assuming that the BP/RP spectrum is a composite of the two components of an unresolved binary, that is, two stars at the same distance with a common interstellar extinction. These parameters for sources with $G \leq 18.25$ are found in the astrophysical_parameters table: azero_msc and distance_msc along with their upper and lower confidence levels. By assuming that the flux comes from a combined system, the distances are necessarily larger than the GSP-Phot ones; see Sect. 11.4.1 of the online documentation.

6.2.3. Extended Stellar Parametrizer for hot stars

For stars hotter than 7500 K, and using a preliminary classification from ESP-ELS, ESP-HS measures the A_0 interstellar extinction by fitting the observed BP/RP and, where available, also the RVS data, called azero_esphs. While A_0 is the free parameter representing interstellar absorption during the fit, the corresponding extinction in the G band, A_G, and interstellar reddening, $E(G_{BP} - G_{RP})$, along with uncertainties, are derived simultaneously. These results are found in the astrophysical_parameters table for hot stars with $G < 17.65$. We show in Fig. 14 a comparison between the A_0 estimates for the 1 433 932 hot stars ($T_{eff} > 7500$ K) in common between GSP-Phot and ESP-HS. The synthetic spectra adopted by the modules are slightly different, as ESP-HS makes some corrections to account for systematic errors between the observations and simulations, and the wavelength range above 800 nm was not taken into account. The impact of this is mostly seen in the B- and O-type star T_{eff} range where ESP-HS estimates tend to be slightly larger than those obtained by GSP-Phot.

6.2.4. General Stellar Parametrizer from spectroscopy

For the sources where an analysis of the DIB in the RVS spectra is possible (see lower right panel of Fig. 3), we provide a measurement of the DIB λ862 nm equivalent width dibew_gspspec, and the modelled depth dibp0_gspspec and width dibp2_gspspec parameters, together with uncertainties. A quality flag dibqf_gspspec is also available ranging from 0 (highest quality) to 5 (lowest quality). Results for DIB measurements are available for 476 117 stars, and are found in the astrophysical_parameters table for T_{eff} ranging from ~3000 to 50 000 K. A comparison between dibew_gspspec and ebpminrp_gspphot is shown for a high-quality subsample in Fig. 15 for stars with dibqf_gspspec <2, where the median DIB EW increases with $E(G_{BP} - G_{RP})$. A detailed discussion between the correlation of the DIB carrier and the dust extinction can be found in Gaia Collaboration (2023f).

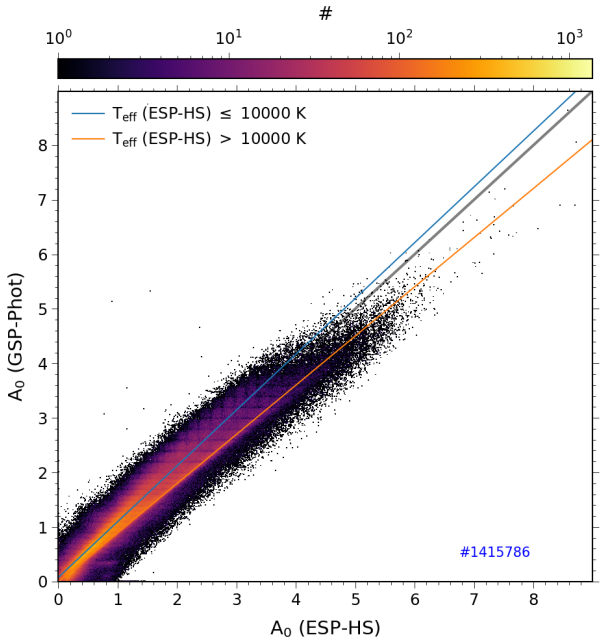


Fig. 14. Distribution of the A_0 derived by GSP-Phot vs. A_0 derived by ESP-HS. The grey diagonal is the identity relation, while the blue and orange lines were fitted through the values obtained for targets cooler and hotter than 10 000 K, respectively.

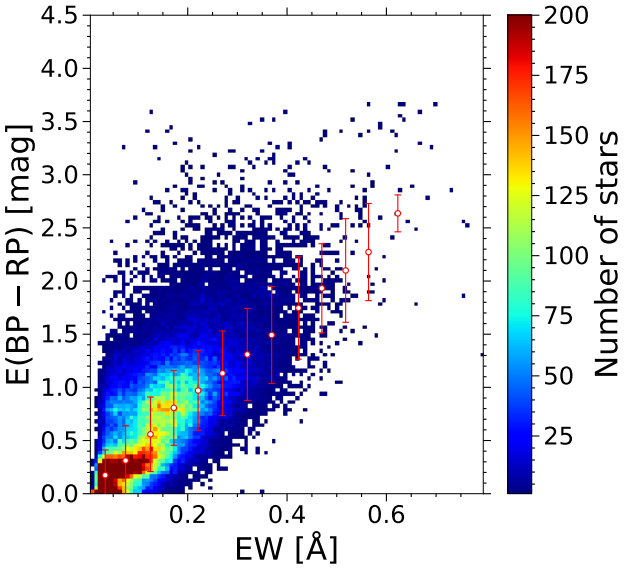


Fig. 15. Comparison between GSP-Spec DIB equivalent width and GSP-Phot $E(G_{BP} - G_{RP})$. The red dots are the median values of $E(G_{BP} - G_{RP})$ taken in EW bins from 0.0 to 0.6 Å with a step of 0.05 Å. The error bars show the standard deviations of $E(G_{BP} - G_{RP})$ for each EW bin.

6.2.5. Total Galactic extinction

All-sky HEALPix maps of the TGE are made available in two separate tables in the *Gaia* DR3 archive at various resolutions (HEALPix levels), namely the tables total_galactic_extinction_map and total_galactic_extinction_map_opt. The estimation of the TGE in each HEALPix is taken as the median A_0 of the extinction tracers, as measured by GSP-Phot, where the tracers are giants outside the ISM layer of the disc of the Milky Way. The first table,

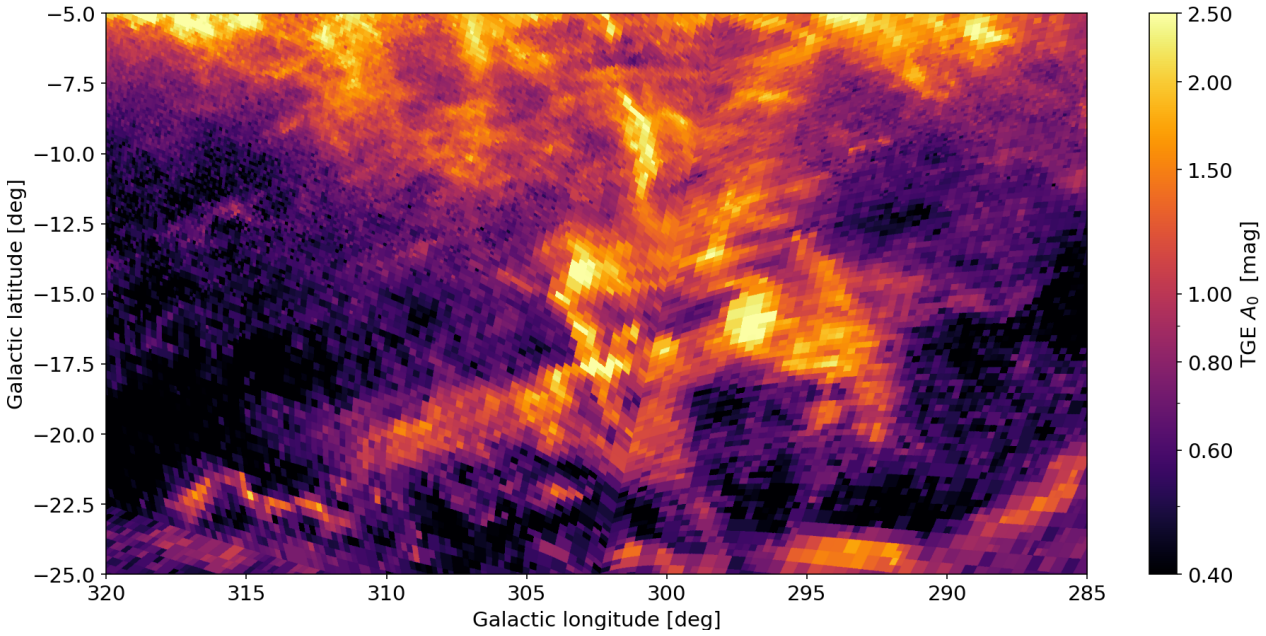


Fig. 16. Total galactic extinction of the Chamaeleon region from the `total_galactic_extinction_map_opt`, showing the extinction at the optimal HEALPix level (between 6 and 9).

`total_galactic_extinction_map`, contains HEALPix maps at levels 6 through 9 (corresponding to pixel sizes of 0.839 to 0.013 deg^2), with extinction estimates for all HEALPixes that have at least three extinction tracers. The second map is a reduced version of this first map, using a subset of the pixels to construct a map at variable resolution, using the highest HEALPix level available (6 through 9) that has at least ten tracers for that HEALPix level. An example of the TGE map for the Chameleon region is shown in Fig. 16.

6.3. Stellar spectroscopic parameters

The BP/RP and RVS spectra contain information about atmospheric parameters of stars: $T_{\rm eff}$, log g, [M/H], along with chemical abundances, an activity index, equivalent widths, and $v \sin i$. The parameters are derived by the two general stellar parametrisers: GSP-Phot and GSP-Spec based on the BP/RP and RVS spectra respectively, assuming a single source. Other estimates of these parameters are produced by modules working in specific stellar regimes, and depending on the scientific case the user may prefer to use these results: ESP-HS, ESP-CS, and ESP-UCD are tailored to analysing hot stars, cool active stars, and ultracool dwarfs, respectively. Finally, MSC provides two $T_{\rm eff}$, two log g, and one [M/H] parameter assuming that the BP/RP spectra are a combination of two components of an unresolved binary. The quality, validation, and use of the stellar spectroscopic and evolutionary parameters are described in the accompanying Paper II.

6.3.1. General Stellar Parametrizer from photometry

GSP-Phot provides estimates of the $T_{\rm eff}$, log g, [M/H], and upper and lower confidence intervals for 470 million sources. These parameters are estimated at the same time as extinction; see Sect. 6.2.1 for details. These parameters are also available in the `mcmc_samples_gsp_phot`. The values for the best library are provided in the `astrophysical_parameters` table, and these are duplicated to `gaia_source` for convenience to the user. The auxiliary parameter `logposterior_gspphot` indicates how well the data fit the model. Results from individual libraries (MARCS, PHOENIX, A, OB) are available in the `astrophysical_parameters_supp` table. A HRD using GSP-Phot $T_{\rm eff}$ and FLAME L is shown in Sect. 6.4.1 (Fig. 24), colour coded according to evolutionary stage.

6.3.2. General Stellar Parametrizer from spectroscopy

The GSP-Spec Matisse-Gauguin method provides 23 independent APs in the `astrophysical_parameters` table for up to 6 million sources derived from the RVS spectra; see the top panels of Fig. 3. These include: $T_{\rm eff}$, log g, [M/H], [α/Fe], goodness-of-fit over the entire spectral range, individual chemical abundances of 12 elements, CN equivalent width and its fitting parameters, and DIB equivalent width and its fitting parameters. For each chemical element abundance, the number of used spectral lines is presented, along with the line-to-line scatter. A histogram with the available chemical abundances and equivalent widths of the CN line and DIB is shown in Fig. 17.

A second method, GSP-Spec-ANN, based on the ANN method (Dafonte et al. 2016; Manteiga et al. 2010), provides four APs in the `astrophysical_parameters_supp` table: `teff_gspspec_ann`, `logg_gspspec_ann`, `mh_gspspec_ann`, `alphafe_gspspec_ann`, and their upper and lower confidence values, along with a goodness-of-fit over the entire spectral range `logchisq_gspspec_ann`.

Finally, following the results of the internal GSP-Spec validation, a long GSP-Spec catalogue flag was implemented during the post-processing and published in both the `astrophysical_parameters` and the `astrophysical_parameters_supp` tables, and the users should therefore check this flag depending on the use case of the parameters; see `flags_gspspec`, and `flags_gspspec_ann` (more details on the use of these flags are provided in Recio-Blanco et al. 2023). A HRD using $T_{\rm eff}$ from GSP-Spec Matisse-Gauguin and the FLAME luminosity is shown in Sect. 6.4.1 (Fig. 24), colour-coded according to stellar age.

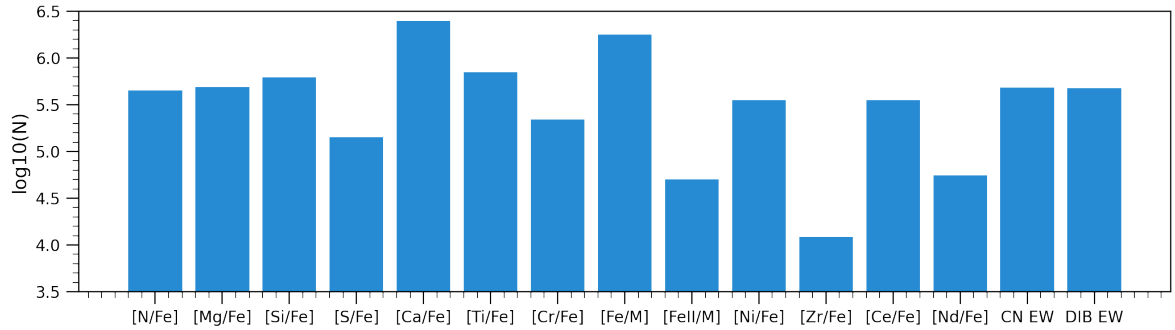


Fig. 17. Histogram showing the number of sources for each chemical species with abundances or equivalent widths in *Gaia* DR3 produced by the GSP-Spec Matisse-Gauguin method, in logarithmic scale. [α/Fe] is derived at the same time as the atmospheric parameters ($T_{\rm eff}$, log g, [M/H]) and is available for approximately 5 million sources. A quality flag `flags_gspspec` is provided for the best use of the elemental abundances.

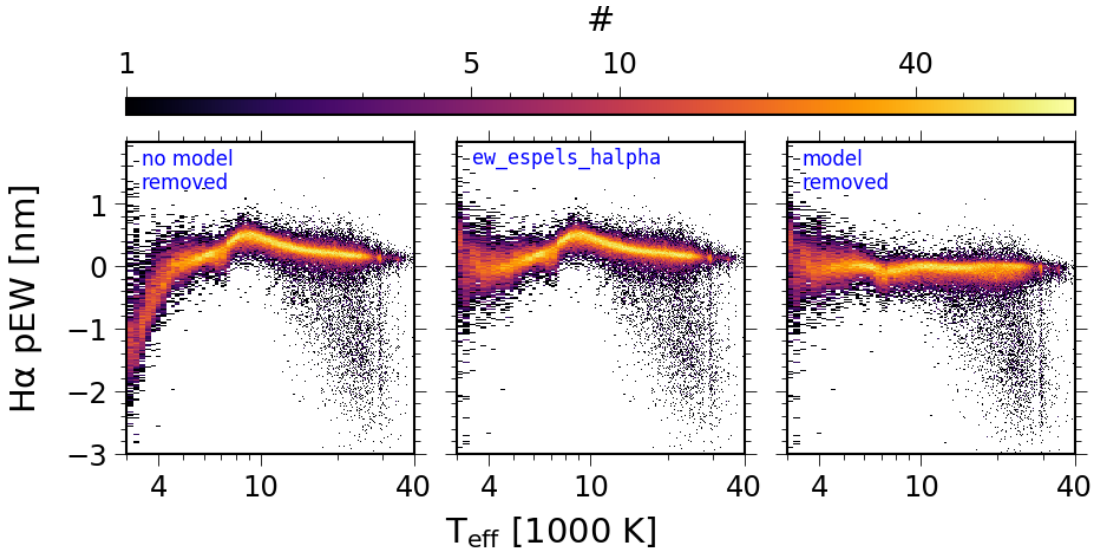


Fig. 18. Distribution of the Hα pseudo-equivalent width (pEW) obtained by ESP-ELS for 135 258 targets chosen in order to homogeneously cover the temperature domain as a function of the effective temperature derived by GSP-Phot and ESP-HS ($T_{\rm eff} > 7500$ K only for the latter). *Left panel*: we report the value obtained before the removal of the model estimate. *Middle panel*: we show the result saved in `ew_espels_halpha` (the model value is removed for sources with $T_{\rm eff} \leq 5000$ K). *Right panel*: we show the result obtained when the model value (`ew_espels_halpha_model`) is also removed for stars hotter than 5000 K.

6.3.3. Extended Stellar Parametrizer for emission-line stars

The ESP-ELS module identifies ELSs in the Hα wavelength domain. An estimate of the Hα pseudo-equivalent width (pEW Hα), `ew_espels_halpha`, for 235 million stars is provided in the catalogue. For stars with `teff_gspphot` ≤5000 K, a correction was applied to mitigate the impact of blends with spectral lines and molecular bands present in the spectra of cooler stars as follows

$$\text{ew_espels_halpha} = \begin{cases} \text{pEWH}\alpha & T_{\rm eff} > 5000\,\text{K} \\ \text{pEWH}\alpha - \text{pEWH}\alpha^{\rm model} & T_{\rm eff} \leq 5000\,\text{K}, \end{cases} \quad (1)$$

where pEWH$\alpha^{\rm model}$ is the pEW Hα value as measured on the simulated and synthetic spectrum that best corresponds to the astrophysical parameters provided by GSP-Phot. The value of the correction is provided by `ew_espels_halpha_model`. When the correction was applied, the value of the Hα quality flag, `ew_espels_halpha_flag`, was set to one; otherwise it was set to zero. Figure 18 shows the temperature distribution of the pseudo-equivalent width (pEW). As expected, when the model estimate is subtracted for the cooler stars (middle panel), the Hα pEW peaks in absorption (i.e. positive values) at temperatures between 8000 and 9000 K. When the model estimate is also applied for the hotter $T_{\rm eff}$ (right panel), the negative estimates are expected to belong to ELSs.

6.3.4. Extended Stellar Parametrizer for hot stars

ESP-HS determines the astrophysical parameters $T_{\rm eff}$ (`teff_esphs`) and log g (`logg_esphs`) for approximately 2 million stars hotter than 7500 K according to the spectral type tag provided by ESP-ELS (`spectraltype_esphs`). These results are found in the `astrophysical_parameters` table. The module assumes a solar chemical composition, and therefore no corresponding metallicity value is saved in the catalogue. The parameters are derived by fitting the BP/RP spectra and, when available, the RVS spectra; see the lower panels of Fig. 3. If RVS data are used, ESP-HS also estimates a line broadening term (i.e. designed to take into account the broadening mechanisms not included when preparing the simulated or synthetic spectra) by assuming that it is only due to the axial rotation of the star ($v \sin i$). We note that an attempt to measure the line broadening term can only be made on the RVS data, when the instrumental broadening does not dominate. Therefore, a value of $v \sin i$ (`vsini_esphs`) is provided along

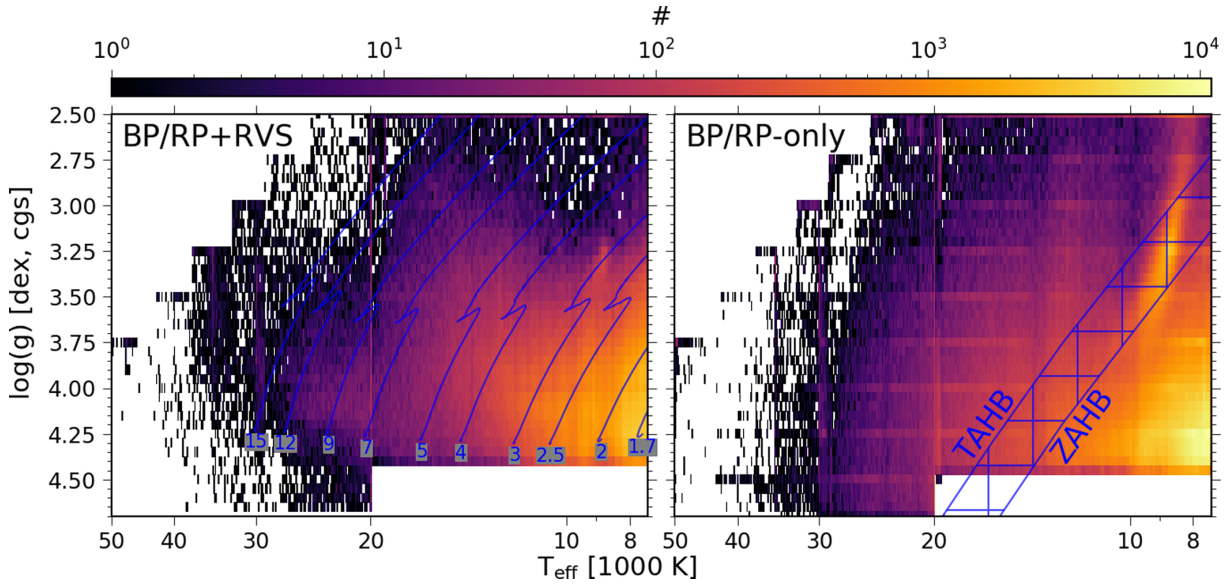


Fig. 19. Kiel diagram of ESP-HS results obtained in BP/RP+RVS (*left panel*) and BP/RP-only (*right panel*) processing modes. *Left panel*: evolutionary tracks of Georgy et al. (2013) for solar metallicity, and $\frac{\omega}{\omega_c} = 0.8$ (rotation at 80% of its critical velocity) are shown in blue. The initial mass in solar masses is indicated at the start of each track. *Right panel*: region occupied by the hot HB stars is delimited by the expected zero age and terminal age HB lines, labelled ZAHB and TAHB, respectively, of which the boundaries are taken from Dorman et al. (1993).

with the spectroscopic parameters for the brighter targets (where RVS spectra were available for processing). The mode adopted to process the data is stored in the first digit reading from the left of the ESP-HS flag `flags_esphs`. Its value is 0 for BP/RP+RVS processing and 1 for BP/RP-only processing. Figure 19 shows the Kiel diagram obtained in both modes. In the fainter magnitude regime (i.e. BP/RP-only mode), the overdensity perpendicular to the main sequence is mainly due to hot horizontal branch stars as was confirmed by a systematic query in the Simbad database (Wenger et al. 2000).

6.3.5. Extended Stellar Parametrizer for ultracool dwarfs

ESP-UCD provides $T_{\rm eff}$ estimates, `teff_espucd`, and uncertainties for ultracool dwarfs (UCDs) for about 94 000 sources in the `astrophysical_parameters` table. An input target list was provided in order to process UCDs; see Sect. 5.1, and these sources were selected according to the following criteria: $\varpi > 1.7$ mas, $G - G_{\rm RP} > 1.0$ mag, $q_{33} > 60$, $q_{50} > 71$, and $q_{67} > 83$, where q_{33}, q_{50}, and q_{67} represent the pixel indices at which the 33.33, 50, and 66.67 percentiles of the total flux in the RP spectrum are attained. These criteria were defined using the *Gaia* UCD sample and include a safety margin to go as far as M6.

In order for the source to appear in the catalogue, we required a $T_{\rm eff}$ estimate in the 500 K to 2700 K range, `phot_rp_n_obs` ≥ 15 and $\log_{10}(\sigma_\varpi) \geq -0.8 + 1.3\log_{10}(\varpi)$, where ϖ is the parallax. We also imposed criteria on the RP flux and distance between the source RP spectrum and its nearest training set template in order to retain the source in the DR3 catalogue[15]. Because the $T_{\rm eff}$ is based on a regression module trained with empirical data, it should be noted by the user that results may be biased for sources with metallicity and gravity departing significantly from the training sample values (solar metallicity and $5.0 \lesssim \log g \lesssim 5.5$).

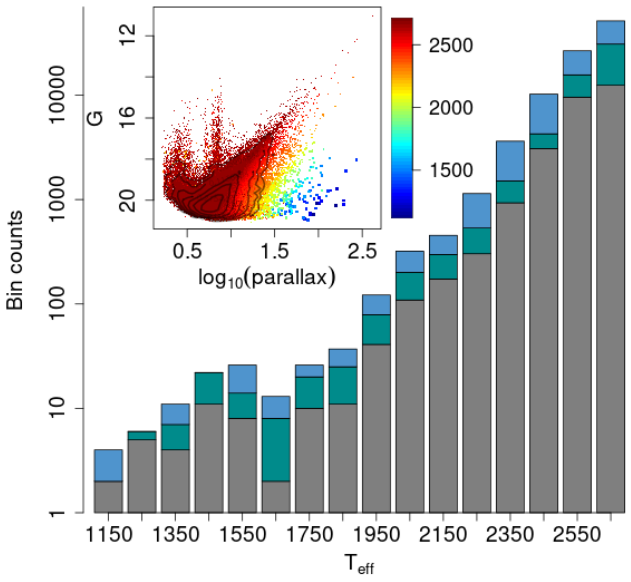


Fig. 20. Distribution of the $T_{\rm eff}$ of UCDs from ESP-UCD according to their quality flag (note the log scale), `flags_espucd` (0 = 40 633, 1 = 26 795, 2 = 26 730 sources). Inset: distribution of these same sources in G and parallax, colour coded according to $T_{\rm eff}$.

The final catalogue of *Gaia* UCDs contains a total of 94 158 sources in three quality categories, `flags_espucd` = 0 (best), 1, 2 (see Sect. 11.3.10 of the online documentation for a more detailed definition). In Fig. 20, we show the distribution of $T_{\rm eff}$ for each of the quality levels across the full $T_{\rm eff}$ range of ultracool dwarfs. The inset shows the distribution of these sources in magnitude–parallax space, colour-coded by `teff_espucd`.

[15] These criteria are the following: the normalised RP spectrum median curvature $\tau < 2.0 \times 10^{-5}$ (see Sect. 11.3.10 of the online documentation for definition of τ), $\log_{10}(d_{\rm TS}) < -2.05$ where $d_{\rm TS}$ is the distance to the template, the sum of negative normalised RP spectrum fluxes ≤ -0.1, and the reddest flux corresponding to the 120th pixel of the (normalised) RP spectrum is less than 0.015.

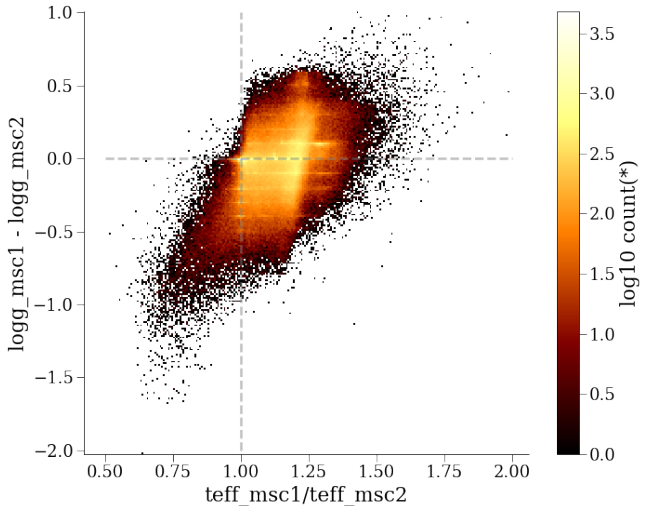


Fig. 21. Distribution of difference in surface gravity ($\log g_1 - \log g_2$ versus effective temperature ratio $T_{\rm eff,1}/T_{\rm eff,2}$ of half a million random sources with results from MSC. MSC assumes that each source is an unresolved binary with the same [M/H], distance, and A_0. The peak uncertainty in $T_{\rm eff,1}/T_{\rm eff,2} \sim 0.2$ and $\log g_1 - \log g_2 \sim 0.7$.

6.3.6. Extended Stellar Parametrizer for cool stars

In *Gaia* DR3, ESP-CS estimates a stellar activity index `activityindex_espcs` and its uncertainties `activityindex_espcs_uncertainty`, in units of nanometres (nm), from the calcium infrared triplet (Ca II IRT, at 849.8, 854.2, and 866.2 nm) in the RVS spectra, see Lanzafame et al. (2023). These parameters and a further parameter, `activityindex_espcs_input`, are found in the `astrophysical_parameters` table for about 2 million sources. The latter parameter indicates whether the source APs used in defining the purely photospheric spectrum to which the RVS spectrum is compared with are from GSP-Spec 'M1' or GSP-Phot 'M2'. During the processing, the default value is to use the parameters from GSP-Spec because the activity index is derived from the same data as the atmospheric parameters, but when they are not available, the ones from GSP-Phot are used.

ESP-CS has processed stars with $G \lesssim 15$, $T_{\rm eff}$ in the range (3000 K, 7000 K), log g in the (3.0, 5.0) range, and [M/H] in the (−0.5, 1.0) range. Only results for sources with the RVS spectrum $S/N \geq 20$ are found in the archive.

Figure 22 shows two examples of the ESP-CS analysis. One is the case of the chromospherically active star *Gaia* DR3 4891212046355683328 (HIP 20737), with `activityindex_espcs` ≈ 0.05 nm. From the analysis of ESO-FEROS archive spectra, Lanzafame et al. (2023) derive a corresponding activity index $R'_{\rm HK} = -3.72$ (Noyes et al. 1984) from the Ca II H&K doublet. The second example is the case of the T Tauri star *Gaia* DR3 6243938170241571 84 with a mass accretion rate of $\log \dot{M} = -10.51\ M_\odot\ {\rm yr}^{-1}$ (Manara et al. 2020).

The ESP-CS activity index is given as an enhancement factor in the core of Ca II IRT lines with respect to a synthetic template representing the spectrum of an inactive star with the same $T_{\rm eff}$, log g, and [M/H]; see Lanzafame et al. (2023) for details. Despite the fact that the method ensures that the photospheric contribution is removed from the activity index parameter, in principle, because of the contrast effect with the underlying continuum, the index derived gives a relative measure of the stellar activity at a given $T_{\rm eff}$. In practice, it can be used to compare stars with similar $T_{\rm eff}$ or the same spectral type, but it is unsuitable for comparing stars with very different $T_{\rm eff}$ or spectral type. Lanzafame et al. (2023) provides a method to derive an index $R'_{\rm IRT}$ from the ESP-CS activity index and $T_{\rm eff}$, which is analogous to the $R'_{\rm HK}$ and largely independent from the contrast effect.

In general, a value of the activity index of around 0.03–0.05 separates the regimes in which the chromospheric activity or mass accretion dominate. The separation in terms of $R'_{\rm IRT}$ is discussed in Lanzafame et al. (2023).

6.3.7. Multiple Star Classifier

The MSC assumes that the BP/RP spectrum is a composition of two unresolved components of a binary system, and estimates $T_{\rm eff}$, `teff_msc1`, and `teff_msc2`, and log g, `logg_msc1`, and `logg_msc2`, for the two components for 349 million sources, with upper and lower confidence intervals. The MSC assumes a solar metallicity prior and estimates one unique metallicity for each source `mh_msc`. These parameters are inferred at the same time as A_0 and distance; see Sect. 6.2.2 In Fig. 21, we show the distribution of the temperature ratio and log g differences of the individual components according to the MSC assumption of an unresolved binary. The grey dashed lines indicate where two sources are of equal mass, that is, where they have the same $T_{\rm eff}$ and log g. Results from MSC are in *Gaia* DR3 if $G < 18.25$. However, users will need to construct a binary sample using external literature sources, or other indicators of binarity, such as `classprob_dsc_combmod_binary` or the other tables in the archive indicating binarity; see Gaia Collaboration (2023a).

6.4. Stellar evolutionary parameters

By stellar evolutionary parameters, we imply the following: mass M[16], luminosity L, absolute magnitude M_G, radius R, radial velocity[17] correction for the stellar gravitational redshift, $rv_{\rm GR}$, the age of a star τ, and the evolutionary stage (`evolstage`). The parameters $rv_{\rm GR}$ and age are in units of kilometers per second and gigayears, respectively. The parameter `evolstage` is an integer from 100 to 1300 indicating the phase of the evolution sequence that the star is in; see Hidalgo et al. (2018). Most of these parameters are derived by FLAME but GSP-Phot also derives M_G and R. These parameters show good agreement between the two modules in most parameter ranges; see Paper II and Sect. 11.4.5 of the online documentation for these comparisons.

6.4.1. Final Luminosity Age Mass Estimator

FLAME produces all of the evolutionary parameters except for M_G, although this can be derived directly using L and BC_G. Two separate results are provided in the archive: the first in the `astrophysical_parameters` table is based on the 'best' library from GSP-Phot for about 280 million sources. A second set of results are based on the Matisse-Gauguin GSP-Spec parameters for approximately 5 million sources, and these are found in the supplementary table `astrophysical_parameters_supp`. Not only are the data found in separate tables, but the field names are also distinguishable with the latter containing `spec`; for example `mass_flame` and `mass_flame_spec`.

[16] Mass is technically not an evolution parameter but is the most important physical property responsible for evolution.
[17] The `radial_velocity` parameter produced by CU6 is found in `gaia_source`. In *Gaia* DR3, it is not corrected for the gravitational redshift nor the convective shift.

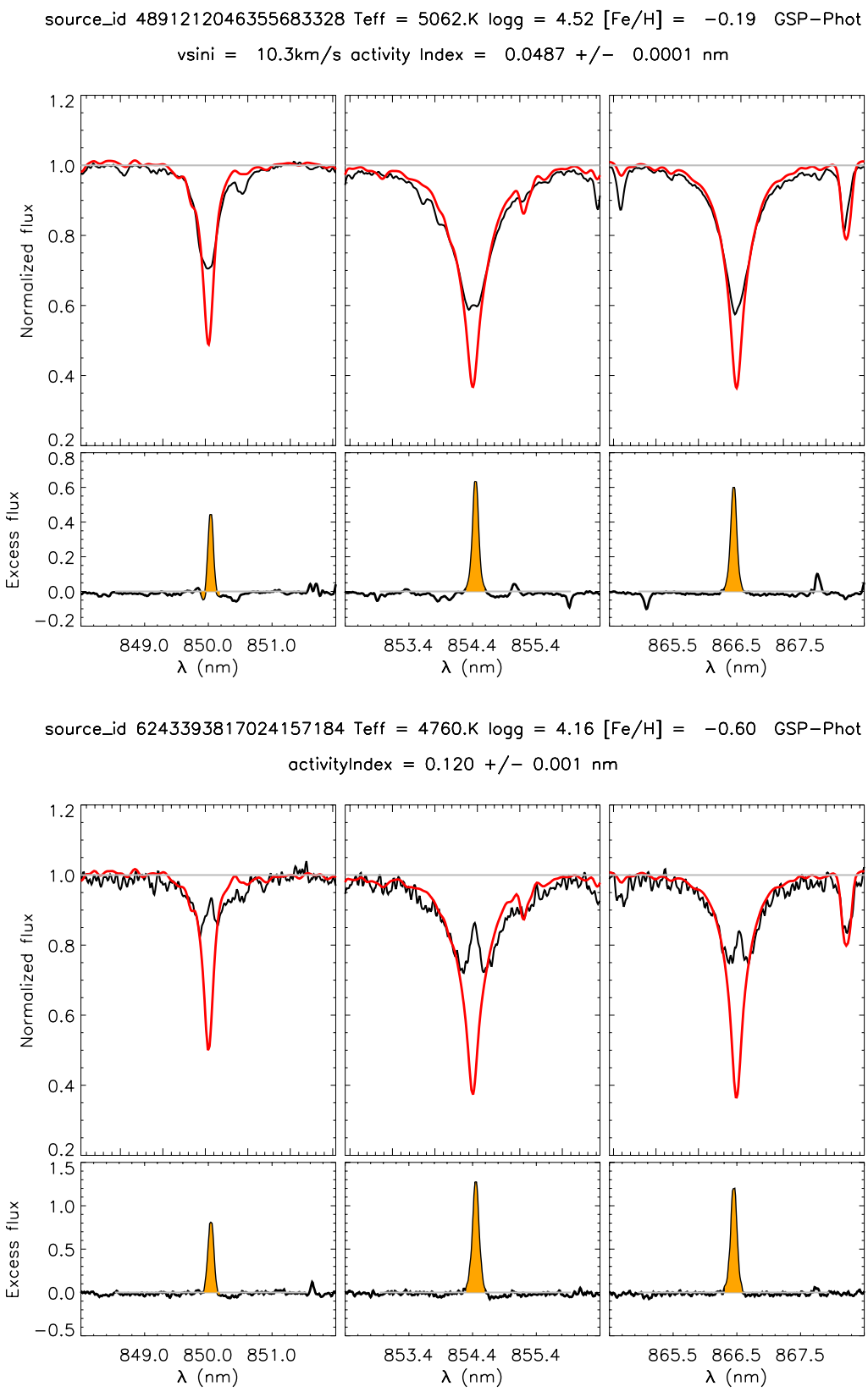


Fig. 22. Examples of the activity index derived by the ESP-CS module. *Top panel*: Ca II IRT RVS spectrum of the chromospherically active star *Gaia* DR3 4891212046355683328 (HIP 20737) with a measured $R'_{\rm HK} = -3.72$ from FEROS spectra using the Ca H&K doublet. *Bottom panel*: RVS spectrum for the T Tauri star *Gaia* DR3 6243393817024157184 with a mass accretion rate of $\log \dot{M} = -10.51\ M_{\odot}\ {\rm yr}^{-1}$. The same method is applied to measure these activity indices for both types of excess flux. Black lines are the observed spectra. Red lines are the purely photospheric spectrum template. The orange filled spectral regions are the area over which the integral of the excess flux is evaluated to produce the activity index.

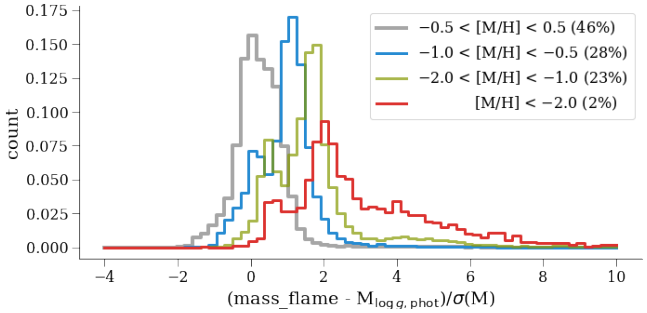


Fig. 23. Comparison of the mass derived from GSP-Phot using log g and R, $M_{\log g,\mathrm{phot}}$, and `mass_flame` for stars of different metallicities. The histograms are normalised for visual purposes and the relative number of stars in each sample is indicated in the label.

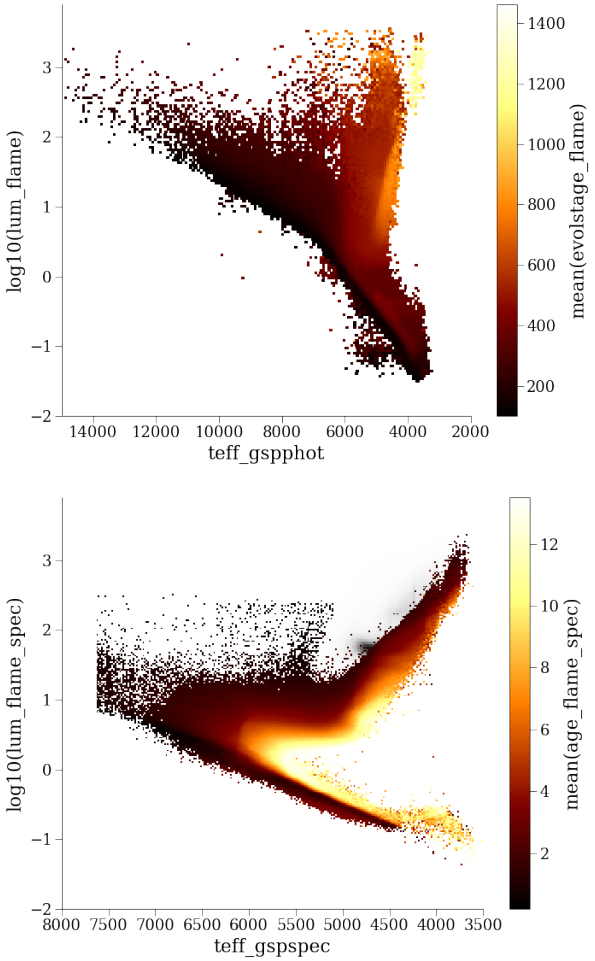


Fig. 24. HRDs colour coded according to FLAME parameters. We note that the colour scale is linear and not a density plot. *Top*: `lum_flame` versus `teff_gspphot` colour coded according to `evolstage_flame` for stars with relative parallax errors of better than 10%. We applied the recommended FLAME filter for giants. *Bottom*: `lum_flame_spec` versus `teff_gspspec` colour coded according to `age_flame_spec` for stars with `flags_gspspec like '0000000000000%'`. In the background, the red clump can be seen in grey. These have no age values associated with them. Even though we made selections on quality on certain parameters, there are still some artefacts that can be seen, such as the high-luminosity, low-T_{eff} giants in the upper panel colour coded in yellow, or the high-luminosity low-mass main sequence stars in the lower panel. These artefacts can be removed by filtering on luminosity uncertainty, or requiring that the T_{eff} from both GSP-Spec and GSP-Phot agree to within 300 K for example, or filtering on spectra S/N.

The values of M, R, L, rv_{GR}, and age are accompanied by an upper and lower confidence level encompassing a confidence interval of 68%; see for example `mass_flame_upper`. To derive L, a bolometric correction for the G band is needed (see Sect. 4.3) and this is provided as an auxiliary parameter `bc_flame(_spec)`[18]. An estimate of the distance is also needed: for the results in the `astrophysical_parameters` table, either the `parallax` or `distance_gspphot` is used. This processing information is provided as the second character of `flags_flame` where '0' implies the use of the parallax, '1' is the use of `distance_gspphot`, while '2' is also parallax but where convergence issues with `distance_gspphot` have been reported; see the online documentation for details.

A solar metallicity prior was assumed for deriving M, age, and `evolstage` in light of the known but unquantified issues with [M/H] at the time of operations; see Sect. 8. This assumption does have an impact on the results for non-solar metallicity stars, in particular for the age, where metal-poor and metal-rich stars will have a biased age towards younger and older ages; see Creevey & Lebreton (2022) for further discussion. One should therefore be cautious when using the age value outside of the $-0.5 < [\mathrm{M/H}] < +0.5$ regime.

As we can derive a mass using `logg_gspphot` and `radius_gspphot`, we can investigate the impact of the solar-metallicity assumption on the masses by comparing these to `mass_flame`. Figure 23 shows the differences in the two mass determinations normalised by their joint uncertainties, for solar-metallicity stars (grey), and then for non-solar metallicity stars (blue, green, red). The histogram is normalised for visual purposes and the percentage of stars of the total in each histogram is indicated in the label. One can see that for low-metallicity stars (<−2.0), the mass from GSP-Phot differs by typically 2σ or more from FLAME. Users may prefer to use such an estimate of mass for the roughly 2% lowest metallicity stars.

Determining the masses and ages of giants is a delicate task compared to the less evolved stars, and our validation shows that the masses should be used with caution for evolved stars. We therefore added quality information in `flags_flame` and `flags_flame_spec` which takes a value of '1' as the first character to indicate that the star is a giant with a published mass (and usually an age) and that these corresponding parameters should be used with caution. Additional validation showed that results for giants with $M > 2\,M_{\odot}$ are misclassified and should not be used. Hertzsprung-Russell diagrams are shown in Fig. 24 using a random subset of 2 million sources from *Gaia* DR3: the top panel shows `lum_flame` versus `teff_gspphot` colour coded according to `evolstage_flame` while the bottom panel shows `lum_flame_spec` versus `teff_gspspec` colour coded according to `age_flame_spec`.

[18] A tool has been provided to calculate this; see `https://gitlab.oca.eu/ordenovic/gaiadr3_bcg`

6.4.2. General Stellar Parametrizer from photometry

Given the use of isochrones by GSP-Phot in a forward model context (see Sect. 6.2.1) GSP-Phot also provides estimates of absolute magnitude M_G and radius R (`mg_gspphot` and `radius_gspphot`) and upper and lower confidence levels. These parameters are found in the `astrophysical_parameters` table for 470 million sources,

and the results for the individual libraries are found in the astrophysical_parameters_supp table. From these GSP-Phot results, the user could also compute GSP-Phot estimates of the (bolometric) luminosity:

$$\frac{L}{L_\odot} = \left(\frac{R}{R_\odot}\right)^2 \left(\frac{T_{\rm eff}}{5772\,{\rm K}}\right)^4. \quad (2)$$

While GSP-Phot does not directly provide absolute magnitudes in the BP or RP bands, GSP-Phot does estimate the distance and extinctions $A_{\rm BP}$ and $A_{\rm RP}$. Therefore, the user can compute absolute magnitudes $M_{\rm XP}$ using the observed apparent magnitudes $G_{\rm XP}$ via

$$M_{\rm XP} = G_{\rm XP} - A_{\rm XP} - 5\log_{10} d + 5, \quad (3)$$

where XP stands for BP or RP, and d is the distance in parsecs. Given those, the user can then compute bolometric corrections in those bands via

$$\mathrm{BC}_{\rm XP} = 4.74 - 2.5\log_{10}(L/L_\odot) - M_{\rm XP}. \quad (4)$$

Uncertainties on those additional quantities can be obtained from the GSP-Phot MCMC samples (see Appendix B) by processing all samples through those equations and then computing their median values and quantiles for example.

6.5. Extragalactic redshifts

The redshifts of extragalactic objects are produced by two modules, QSOC and UGC, which analyse BP/RP spectra of quasars and galaxies, respectively. The selection of the processed sources uses the Combmod class probabilities of DSC, classprob_dsc_combmod_quasar and classprob_dsc_combmod_galaxy. The CU8 extragalactic parameters are found in the qso_candidates and galaxy_candidates tables. More details on the quality and processing of these parameters are given in the accompanying Paper III.

6.5.1. QSO Classifier

QSOC predicts quasar redshifts redshift_qsoc and associated confidence levels redshift_qsoc_lower and redshift_qsoc_upper in the range $0.0826 < z < 6.1295$. Intentionally, the module chose to be complete and produced results on 6.4 million sources, which is three times the expected number of quasars that *Gaia* should theoretically observe. Although this choice may seem questionable, it gives the final user a much higher chance of finding the redshift of the sources they are interested in, at the expense of having many contaminating stars amongst these predictions (see however Gaia Collaboration 2023b, their Sect. 8 for the selection of purer samples).

In order to more easily discriminate between valuable redshift predictions and those where potential processing issues may arise, we defined two quality measurements: ccfratio_qsoc and zscore_qsoc. The ccfratio_qsoc field is associated with the χ^2 resulting from the fit of the BP/RP spectra to the templates at the predicted redshift. Predictions whose redshift is associated with a minimal χ^2 (compared to the χ^2 resulting from alternative redshifts) have ccfratio_qsoc = 1 and less than one otherwise. The zscore_qsoc field is associated with the successful modelling of common quasar emission lines. We have that zscore_qsoc = 1 if all covered quasar emission lines appear in the spectrum, whereas the absence of a single emission line often leads to very low values of zscore_qsoc. These quality measurements are summarised in the flags_qsoc field where boolean flags are set that principally depend on the values of the ccfratio_qsoc and zscore_qsoc fields.

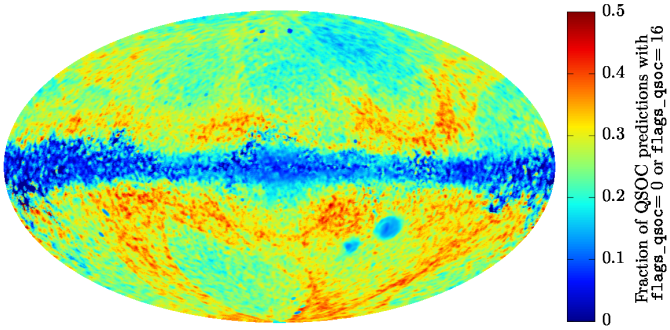


Fig. 25. Galactic sky distribution of the fraction of QSOC predictions that do not raise warning flags (i.e. flags_qsoc = 0), even if they are based on BP/RP spectra of lower quality (i.e. flag Z_BAD_SPEC = 16 can be set). Fractions are computed within radii of 1° over the whole celestial sphere.

To illustrate the potential filtering that can be done using flags_qsoc, Fig. 25 shows the sky distribution of the fraction of QSOC predictions for which all flags other than the Z_BAD_SPEC[19] flag are set to zero. These correspond to sources where no processing error occurs, even though some predictions are based on spectra of lower quality (i.e. predictions with either flags_qsoc = 0 or flags_qsoc = 16). We can clearly see that high-stellar-density regions (Galactic plane, Magellanic Clouds, globular clusters, and nearby galaxies) usually have a lower fraction of predictions with flags_qsoc = {0, 16}. Imprints of the scanning law are also seen. These arise from the higher or lower number of spectral transits that leads to a higher or lower S/N of the spectra and hence more or less confident predictions by QSOC.

6.5.2. Unresolved Galaxy Classifier

The UGC module provides galaxy redshift parameters redshift_ugc, with $0.0 \leq z \leq 0.6$, and associated uncertainties redshift_ugc_lower and redshift_ugc_upper, for 1 367 153 sources in the galaxy_candidates table. We note that these uncertainties are computed from the standard deviation of the SVM predictions of sources with known redshift and should accordingly not be considered as per-source confidence intervals but rather as a measure of the SVM performance.

As the sources in UGC may have a relatively low probability of being galaxies (our selection is classprob_dsc_combmod_galaxy > 0.25), we expect a number of misclassified quasars to contaminate the redshift_ugc results. Potentially, about 1% of these are true quasars with redshifts $z > 0.6$. It is also expected that some true high-redshift galaxies are erroneously processed by UGC. Nevertheless, their number is estimated to be negligibly small. Their predicted redshifts would therefore be underestimated by UGC. Estimated redshifts below 0.02 or larger than 0.40 are not well constrained, and there is a suspicious peak of sources in a very narrow bin at

[19] The Z_BAD_SPEC flag is raised for sources of lower quality, and rely on a combination of the source G magnitude and number of BP/RP spectral transits.

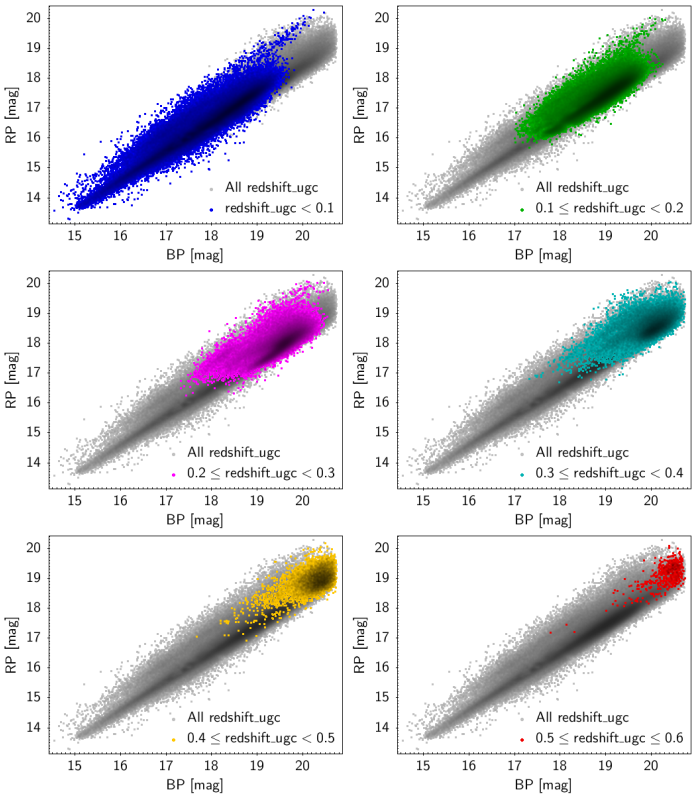


Fig. 26. Magnitude–magnitude diagrams for the 1.4 million galaxies for which redshifts are provided by UGC. Each panel shows the density distribution of a different redshift range in a different colour, while the distribution in grey shows the whole sample.

$0.0707 < z < 0.0709$ (for details see Sect. 11.3.13 of the online documentation). In Fig. 26 we show G_{RP} versus G_{BP} diagrams illustrating the magnitude ranges for which we find galaxies with specific redshifts. Higher redshift galaxies ($z \sim 0.6$) are only found at the very faint end, as expected.

6.6. Outlier Analysis

The results produced by OA can be helpful when performing an extensive analysis of those sources that were assigned a lower classification probability by DSC. Such sources are usually faint stars or extragalactic objects, which can be studied through the parameters generated by OA, but they also contain known objects such as white dwarfs and brown dwarfs. The sources classified as outliers are given in the `astrophysical_parameters` table, as explained in Sect. 6.1.2. Two multi-dimensional tables are also available for further interpretation of these data.

The `oa_neuron_information` table is a SOM arranged in a rectangular lattice composed of 900 neurons, where each neuron groups similar objects that are described by means of different statistical parameters. In order to assess the quality of the clustering, different indices are available in this same table, meaning that high-quality and low-quality neurons can be identified and filtered as required by the user to perform their own analysis or to isolate specific types or groups of objects. To ease such an analysis, an indication of the astronomical type of the sources is also provided for the best-quality neurons. The SOM is shown in Fig. 12 and the coloured neurons are the best-quality ones with class labels.

Each neuron is associated with a synthetic BP/RP spectrum, the so-called prototype, which is representative of the spectra of the sources that are assigned to a certain neuron. In addition, for those neurons where a class label is provided, the BP/RP spectrum of the template is found in the `oa_neuron_xp_spectra` table. Further information is provided in the accompanying Paper III.

6.7. Auxiliary data products

The auxiliary data products comprise quality metrics, convergence indicators, flags, the name of the best library for the GSP-Phot results, and the bolometric correction. Most of these auxiliary data products have been described in one of the above subsections. These are also listed towards the bottom of Table 5.

7. Validation of results

The aim of this paper is to explain the production and overall content of the astrophysical parameters from CU8 in *Gaia* DR3. The accompanying Papers II and III focus on the validation and quality of the results for stellar-based APs (Fouesneau et al. 2023) and the non-stellar content and source classification (Delchambre et al. 2023). Additional validation results are also given in the dedicated online documentation chapter. Validation on GSP-Phot, GSP-Spec, and ESP-CS-specific products are also found in their dedicated papers (Andrae et al. 2023; Lanzafame et al. 2023; Recio-Blanco et al. 2023). Here, we briefly describe our validation procedures.

The validation of the CU8 data products included several steps. At a first level, many Apsis test runs were performed prior to receiving the upstream data during the DR3 development stage (2018–2020). This repeated validation was done on a module-by-module basis for a limited number of mostly random sources (10 million). The teams compiled many validation tests to ensure that the software performed as intended. Such tests comprise checking the astrophysical content of the data, for example HRDs such as that shown in Fig. 24, which helped to point out weaknesses in the codes in certain parameter spaces. Comparisons with external data allowed the teams to check whether their results are consistent with what is already known in the literature. However, it is important to point out that Apsis does not do any calibration of its APs to mimic external catalogues. The external catalogues were merely used as a consistency check. Once the final input data had been received (six months before operations), further test runs were performed to refine the codes and to adapt parameter settings to the final data.

A higher level of validation was performed using a validation database hosted at the ESAC operations centre in Madrid. This allowed the CU8 team to perform many cross-checks between the individual modules (see e.g., Fig. 14), and provided important feedback for necessary modifications to the code before the full operations sequence. It also allowed a statistical check on the full dataset, which then allowed the post-processing codes to be prepared for filtering results and setting archive flags.

A third level of validation was then performed by the Coordination Unit 9 (CU9) archive team once the operational data had been delivered. The CU9 archive team were the first users of the data, and had an external view of the full results in the archive just as a user would; see Babusiaux et al. (2023). Once the CU8 results were final, the validation by CU9 only allowed us to perform minor updates on parameters through

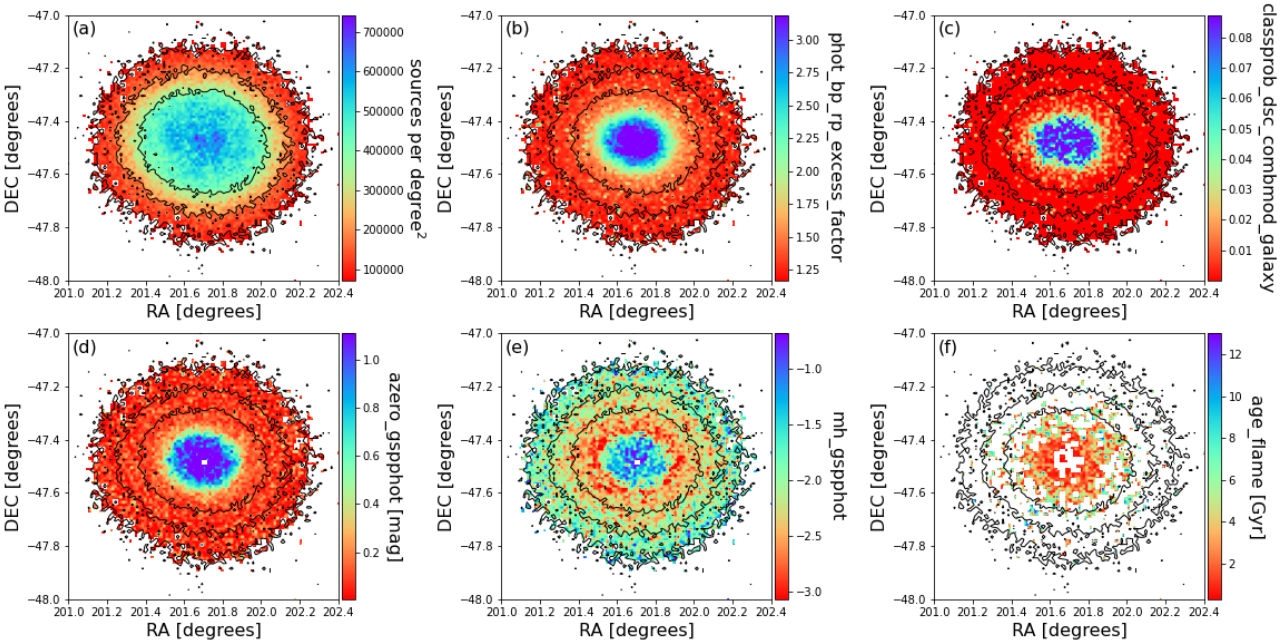


Fig. 27. Crowding effects in the globular cluster Omega Centauri. Black contours are identical in all panels and indicate source density dropping by factors of 2. *Panel a*: source density. *Panel b*: excess factor in photometric flux. *Panel c*: galaxy class probability from DSC-Combmod. *Panel d*: extinction estimate A_0 from GSP-Phot. *Panel e*: metallicity estimate from GSP-Phot. *Panel f*: age estimate from FLAME.

post-processing, for example removing results for some sources or removing fields from the *Gaia* archive. There are, nonetheless, some issues that are now known that could not be corrected, and these along with some caveats are summarised in the following section.

8. Caveats and known issues

There are several caveats that users should be aware of before using the data. Additionally, a number of issues have been found following extensive validation. We list both caveats and the main issues known to us at the time of writing, starting with general comments on variability and crowding, followed by a discussion on a module-by-module basis. The user should consider these issues when using the APs in *Gaia* DR3.

Variability. Apsis processes the mean BP/RP and RVS spectra, astrometry, and mean magnitudes provided by upstream processing systems. Therefore, we advise users to consider the variability of their source before deciding whether the APs from Apsis are adapted to their specific science case. As a concrete example, RR Lyrae stars have large-amplitude variability, and a mean spectrum for these stars will not necessarily represent the mean state of such a star. Additionally, as these spectra vary significantly, the concept of a $T_{\rm eff}$ from one mean spectrum does not make astrophysical sense; see e.g., Clementini et al. (2023).

Crowding. Crowding is a major limitation in dense regions such as stellar clusters. As an example, Fig. 27 shows that CU8 results differ significantly between the dense core and the outer regions of the globular cluster Omega Centauri (Calamida et al. 2020). For low-resolution BP/RP spectra, the allocated CCD window is 3.5 arcsec × 2.1 arcsec (Carrasco et al. 2021), which means that theoretically about 1.76 million windows would fit into one square degree. In practice though, windows on a source over a range of observation epochs will have quasi-random orientations on the sky and Fig. 27a suggests that CCD windows already start to overlap at about 600 000 sources per square degree, thereby producing blended BP/RP spectra (and photometry, see Fig. 27b) that lead to systematically incorrect CU8 results. For RVS spectra, the window size is much larger than for BP/RP spectra (74.2 arcsec × 1.8 arcsec prior to June 2015, 75.3 arcsec × 1.8 arcsec after June 2015, see Cropper et al. 2018) but RVS spectra are deblended (Seabroke et al., in prep.).

DSC. Performance on white dwarfs and physical binaries is poor, and in general the probabilities may not be well calibrated. The purity of quasars and galaxies on the full sample is low, but this improves (to ∼80%) when excluding low Galactic latitudes ($|b| < 10°$) and using `classlabel_dsc_joint`. We note that these purities account for the expected dominance of contaminating stars in a random sample selected from *Gaia*.

GSP-Phot. Distance estimates tend to be underestimated beyond 2–3 kpc because of a harsh extinction prior. Nevertheless, distances remain reliable for high-quality parallax measurements ($\frac{\varpi}{\sigma_\varpi} > 20$) even out to 10 kpc. Metallicities show an offset of about −0.2 dex compared to external literature sources with [M/H] > −1 dex, and additional systematics exist below −1.0 dex. We recommend correcting these metallicities using the empirical correction that has been made available to the community; see Appendix E. Also, the uncertainties are known to be underestimated. This is most probably due to ignoring the off-diagonal elements of the covariance matrix[20]. Another possible explanation could be in the mismatches between model SEDs and observed BP/RP spectra (see Fig. 1 in Andrae et al. 2023). This is the subject of ongoing investigation. Comparisons with external data show median absolute differences on the order of 120 K for FGK-type stars, 340 K for A stars, and 1600 K for B stars; see Table 11.19 in the online documentation.

[20] Ignoring the covariance matrix could also lead to overestimated uncertainties, but tests with GSP-Photshow that the effects are primarily to underestimate them.

GSP-Spec. Quality flags with up to 41 characters have been provided for the best use of the data; users are strongly encouraged to use these for selecting best sample stars. In the case of Matisse-Gauguin (`astrophysical_parameters` table), the parameters log g and [M/H] show some biases with respect to the literature, and these depend on log g. Recio-Blanco et al. (2023) proposed a log g and [M/H]-calibration procedure to be applied only if the user-specific science case requires it. ANN results (`astrophysical_parameters_supp`) also show biases with respect to the literature, as discussed in Recio-Blanco et al. (2023).

MSC uses a solar-metallicity prior for deriving $T_{\rm eff}$ and log g of binary components, and log g values are in general overestimated with respect to external catalogues. Additionally, as MSC treats all stars as binaries, one would need an external catalogue to identify a reliable set of binary stars. We also note that the value reported as the parameter in the `astrophysical_parameters` table is the median value of the last 100 values available in the `mcmc_samples_msc` table, while the 16th and 84th percentiles come from the full MCMC chain from the processing.

FLAME uses a solar-metallicity prior for deriving masses, ages, and evolutionary stage, and therefore the ages of stars with known metallicities < -0.5 should be used with caution. The uncertainties in the FLAME masses and ages are also underestimated, mainly because of the underestimated uncertainties in $T_{\rm eff}$. For the use of the masses of giants, that is, with the first digit of `flags_flame(_spec)` = 1, the published results should only be used within the range 1–2 $M_{\odot}$(approximately 14 million of 27 million sources).

ESP-HS uses a solar-metallicity prior to derive spectroscopic parameters of hot stars. The validity of this assumption should be considered in the context of the user-specific science case.

ESP-UCD uses empirical training data for the prediction of $T_{\rm eff}$, and so sources deviating significantly from the median metallicities or gravities of the training set, solar metallicities, and log $g \approx 5$–5.5 may have biased estimates of $T_{\rm eff}$. Also, the list of UCD candidates in the quality class 2 contains some contaminants due to incorrect astrometry. This is visible in the distribution of UCD candidates on the celestial sphere as overdensities in the Galactic disc plane (see Sect. 11.3.10 of the online documentation). Finally, comparisons with effective temperatures of limited samples in the literature show a systematic difference in the sense that ESP-UCD estimates are ∼65 K lower than the literature values for the hot end of the sample ($T_{\rm eff} > 2300$ K; see Gaia Collaboration 2023c).

QSOC aims for completeness, not purity, and accordingly processed a large fraction of stars. The prediction of the redshifts of $0.9 < z < 1.3$ quasars are complicated by the sole detectable presence of the Mg II emission lines in the BP/RP spectra over this redshift range. QSOC is designed to process Type-I/core-dominated quasars with broad emission lines in the optical and accordingly yields only poor predictions on galaxies, type-II AGN, and BL Lacertae/blazar objects. Use of the flags is also encouraged.

UGC. Some contamination by high-redshift galaxies and high-redshift quasars is expected in the sample. There is also a suspicious peak of sources in the $0.0707 < z < 0.0709$ range.

TGE. While the TGE extinction maps show excellent agreement with comparable extinction maps, there is a small bias at very small extinctions ($A_0 < 0.1$ mag), and possibly at large extinctions ($A_0 > 4$ mag). It is not advisable to use the TGE maps at very low ($|b| < 5°$) Galactic latitudes; see Delchambre et al. (2023) for further details.

9. Conclusions

Gaia DR3 contains one of the most extensive catalogues of astrophysical parameters to be exploited by the community, and is based on *Gaia*-only data. It contains valuable information on stellar and non-stellar sources, and these parameters appear in ten main archive tables. A minor subset of APs also appear in `gaia_source` to simplify querying for a new user of ADQL. There are up to 1.6 billion classifications of objects (star, galaxy, and so on), along with 470 million stellar-based APs and 6 million extragalactic redshifts of quasar (6M) and galaxy (2M) candidates, in addition to a SOM of outliers, total Galactic extinction maps, and MCMC samples. All of these were produced by the Apsis analysis system (Sect. 3).

Stellar-based analyses were performed using general methods by the Apsis modules GSP-Phot and GSP-Spec which derived the spectroscopic parameters ($T_{\rm eff}$, log g, etc.) using the BP/RP and RVS spectra, respectively, assuming each source to be a single star. The MSC module also analysed the BP/RP spectra in a general way but assumed the BP/RP spectrum of each source to be a composition of two unresolved components of a physical binary. The FLAME module derived evolutionary parameters (R, M, etc.) using the spectroscopic parameters from GSP-Phot and GSP-Spec together with distance measures and photometry. Furthermore, specialised modules produced results focussing on specific types of stars: ESP-CS, ESP-HS, ESP-UCD, and ESP-ELS analysed cool active stars, hot stars, ultracool dwarfs, and emission-line stars, respectively, to provide class probabilities and labels of ELSs, activity index, $T_{\rm eff}$, $v \sin i$, H-α equivalent widths, and spectral types. All of the parameters are found in the main `astrophysical_parameters` table, with supplementary results from GSP-Phot, GSP-Spec, and FLAME in the `astrophysical_parameters_supp` table. Samples of the MCMC chains are also available for GSP-Phot and MSC; see Appendix B.

Redshifts of galaxies and quasars are available in the `qso_candidates` and `galaxy_candidates` tables. Two-dimensional total Galactic extinction maps are provided, which are based on the extinction tracers provided by GSP-Phot at four different HEALPix levels and an optimal one. The results of an unsupervised analysis of outliers are found in the `oa_neuron_information` and `oa_neuron_xp_spectra` tables.

As with any extensive catalogue analysing all sources in a homogenous way, there are both high-quality and low-quality results. Caveats and known issues with the catalogue are summarised in Sect. 8. Further information on the quality, validation, and use of the data is provided in the accompanying Papers II and III which target the stellar parameters and the non-stellar content, respectively (Fouesneau et al. 2023; Delchambre et al. 2023). More technical details are available in the online documentation. Dedicated papers focussing on specific Apsis modules are also available for GSP-Phot (Andrae et al. 2023), GSP-Spec (Recio-Blanco et al. 2023), ESP-CS (Lanzafame et al. 2023), and QSOC (Delchambre 2018). In addition to these data, several tools are made available to the community to aid their exploitation; see Appendix E.

Examples of the use and performance of the astrophysical parameters are shown in several papers accompanying *Gaia* DR3. Gaia Collaboration (2023b) explores the extra-galactic content of *Gaia* by combining results from Apsis with other *Gaia* results. Gaia Collaboration (2023d) exploits the Apsis results along with kinematics and known variable sources to map the spiral arms of the Milky Way. Gaia Collaboration (2023c)

focusses on several regions of the HR diagram to provide golden samples of APs, including a focus on carbon stars and solar analogues. Gaia Collaboration (2023e) illustrates the chemo-dynamical analysis of disc and halo populations using the chemical analysis from Apsis, and Gaia Collaboration (2023f) traces the spatial structure of the Galactic ISM using the λ862 nm DIB.

Gaia DR4 (2025) will be based on the analysis of 66 months of *Gaia* data which is almost twice as long as the data in the current data release. This fact along with the application of improved data processing methods will ensure a significant improvement in the quality of the astrometric, photometric, and spectroscopic data. This, in turn, will impact the quality of the astrophysical parameters with improved control of sources of systematic errors, allowing us to deliver an even more extensive catalogue of *Gaia*-based astrophysical parameters to the community.

Acknowledgements. We thank the referee for their careful reading of the paper and for providing constructive feedback. This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* MultiLateral Agreement (MLA). The *Gaia* mission website is https://www.cosmos.esa.int/gaia. The *Gaia* archive website is https://archives.esac.esa.int/gaia. Acknowledgments from the financial institutions are in Appendix F. We thank our DPAC colleague Hector Canovas for providing Python routines to download spectra and MCMC samples. The data analysis made use of Vaex (Breddels & Veljanoski 2018), TOPCAT (Taylor 2005), and R (R Core Team 2013).

References

Abia, C., de Laverny, P., Cristallo, S., Kordopatis, G., & Straniero, O. 2020, A&A, 633, A135
Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, ApJS, 240, 23
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Allard, F., Homeier, D., & Freytag, B. 2013, Mem. Soc. Astron. It., 84, 1053
Altavilla, G., Marinoni, S., Pancino, E., et al. 2015, Astron. Nachr., 336, 515
Álvarez, M. A., Dafonte, C., Manteiga, M., Garabato, D., & Santoveña, R. 2021, Neural Comput. Appl., 34, 1993
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)
Babusiaux, C., Fabricius, C., Khanna, S., et al. 2023, A&A, 674, A32 (*Gaia* DR3 SI)
Bailer-Jones, C. A. L. 2011, MNRAS, 411, 435
Bailer-Jones, C. A. L. 2021, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C8-TN-MPIA-CBJ-094, http://www.cosmos.esa.int/web/gaia/public-dpac-documents
Bailer-Jones, C., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74
Bailer-Jones, C. A. L., Fouesneau, M., & Andrae, R. 2019, MNRAS, 490, 5615
Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28
Breddels, M. A., & Veljanoski, J. 2018, A&A, 618, A13
Calamida, A., Zocchi, A., Bono, G., et al. 2020, ApJ, 891, 167
Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86
Chang, C.-C., & Lin, C.-J. 2011, ACM Trans. Intell. Syst. Technol., 2, 27:1, software available at https://www.csie.ntu.edu.tw/~cjlin/libsvm
Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068
Clementini, G., Ripepi, V., Garofalo, A., et al. 2023, A&A, 674, A18 (*Gaia* DR3 SI)
Coifman, R. R., & Lafon, S. 2006, Appl. Comput. Harmonic Anal., 21, 5
Contursi, G., de Laverny, P., Recio-Blanco, A., & Palicio, P. A. 2021, A&A, 654, A130
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Creevey, O. L., & Lebreton, Y. 2022, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C8-TN-OCA-OLC-035, http://www.cosmos.esa.int/web/gaia/public-dpac-documents
Cropper, M., Katz, D., Sartoretti, P., et al. 2018, A&A, 616, A5
Dafonte, C., Fustes, D., Manteiga, M., et al. 2016, A&A, 594, A68
De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, A&A, 674, A2 (*Gaia* DR3 SI)
Delchambre, L. 2018, MNRAS, 473, 1785
Delchambre, L., Bailer-Jones, C. A. L., Bellas-Velidis, I., et al. 2023, A&A, 674, A31 (*Gaia* DR3 SI)
Dieterich, S. B., Henry, T. J., Jao, W.-C., et al. 2014, AJ, 147, 94
Dorman, B., Rood, R. T., & O'Connell, R. W. 1993, ApJ, 419, 596
Ducourant, C., Krone-Martins, A., Galluccio, L., et al. 2023, A&A, 674, A11 (*Gaia* DR3 SI)
El-Badry, K., & Rix, H.-W. 2018, MNRAS, 480, 4884
Fitzpatrick, E. L. 1999, PASP, 111, 63
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28 (*Gaia* DR3 SI)
Frémat, Y., Zorec, J., Hubert, A. M., & Floquet, M. 2005, A&A, 440, 305
Gaia Collaboration (Brown, A., et al.) 2016, A&A, 595, A2
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2021a, A&A, 650, C3
Gaia Collaboration (Smart, R. L., et al.) 2021b, A&A, 649, A6
Gaia Collaboration (Arenou, F., et al.) 2023a, A&A, 674, A34 (*Gaia* DR3 SI)
Gaia Collaboration (Bailer-Jones, C. A. L., et al.) 2023b, A&A, 674, A41 (*Gaia* DR3 SI)
Gaia Collaboration (Creevey, O. L., et al.) 2023c, A&A, 674, A39 (*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 (Schultheis, M., et al.) 2023f, A&A, 674, A40 (*Gaia* DR3 SI)
Gaia Collaboration (Vallenari, A., et al.) 2023g, A&A, 674, A1 (*Gaia* DR3 SI)
Georgy, C., Ekström, S., Granada, A., et al. 2013, A&A, 553, A24
Geurts, P., Ernst, D., & Wehenkel, L. 2006, Mach. Learn., 63, 3
Grevesse, N., Asplund, M., & Sauval, A. J. 2007, Space Sci. Rev., 130, 105
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hastie, T., & Stuetzle, W. 1989, J. Am. Stat. Assoc., 84, 502
Hidalgo, S. L., Pietrinferni, A., Cassisi, S., et al. 2018, ApJ, 856, 125
Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, AJ, 150, 148
Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, AJ, 160, 120
Kohonen, T. 2001, in Self-Organizing Maps, 3rd edn. (Berlin Heidelberg: Springer-Verlag)
Lanzafame, A. C., Brugaletta, E., Frémat, Y., et al. 2023, A&A, 674, A30 (*Gaia* DR3 SI)
Lindegren, L., Bastian, U., Biermann, M., et al. 2021, A&A, 649, A4
Liu, C., Bailer-Jones, C. A. L., Sordo, R., et al. 2012, MNRAS, 426, 2463
Manara, C. F., Natta, A., Rosotti, G. P., et al. 2020, A&A, 639, A58
Manteiga, M., Ordóñez, D., Dafonte, C., & Arcay, B. 2010, PASP, 122, 608
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
Marinoni, S., Pancino, E., Altavilla, G., et al. 2016, MNRAS, 462, 3616
Montegriffo, P., De Angeli, F., Andrae, R., et al. 2023, A&A, 674, A3 (*Gaia* DR3 SI)
Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. 1984, ApJ, 279, 763
Pancino, E., Altavilla, G., Marinoni, S., et al. 2012, MNRAS, 426, 1767
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Passegger, V. M., Reiners, A., Jeffers, S. V., et al. 2018, A&A, 615, A6
Pastorelli, G., Marigo, P., Girardi, L., et al. 2020, MNRAS, 498, 3283
Plez, B. 2012, Astrophysics Source Code Library [record ascl:1205.004]
Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, A&A, 638, A76
R Core Team 2013, R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing)
Rabus, M., Lachaume, R., Jordán, A., et al. 2019, MNRAS, 484, 2674
Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning (Cambridge, MA, USA: MIT Press), 248
Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, MNRAS, 370, 141
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)
Reiners, A., Homeier, D., Hauschildt, P. H., & Allard, F. 2007, A&A, 473, 245
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Rybizki, J., Demleitner, M., Bailer-Jones, C., et al. 2020, PASP, 132, 074501
Sartoretti, P., Katz, D., Cropper, M., et al. 2018, A&A, 616, A6
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
Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154
Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287
Tanga, P., Pauwels, T., Mignard, F., et al. 2023, A&A, 674, A12 (*Gaia* DR3 SI)
Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Zhao, H., Schultheis, M., Recio-Blanco, A., et al. 2021, A&A, 645, A14
Zwitter, T., Castelli, F., & Munari, U. 2004, A&A, 417, 1055

[1] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[2] INAF – Osservatorio astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
[3] CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[4] Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
[5] Observational Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden
[6] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[7] CIGUS CITIC – Department of Computer Science and Information Technologies, University of A Coruña, Campus de Elviña s/n, A Coruña 15071, Spain
[8] National Observatory of Athens, I. Metaxa and Vas. Pavlou, Palaia Penteli, 15236 Athens, Greece
[9] INAF – Osservatorio Astrofisico di Catania, via S. Sofia 78, 95123 Catania, Italy
[10] Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[11] Dpto. de Inteligencia Artificial, UNED, c/ Juan del Rosal 16, 28040 Madrid, Spain
[12] Institut d'Astrophysique et de Géophysique, Université de Liège, 19c, Allée du 6 Août, 4000 Liège, Belgium
[13] INAF – Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese (TO), Italy
[14] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[15] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, 40129 Bologna, Italy
[16] ATG Europe for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
[17] 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
[18] Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[19] Telespazio for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[20] 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
[21] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[22] Centre for Astrophysics Research, University of Hertfordshire, College Lane, AL10 9AB Hatfield, UK
[23] APAVE SUDEUROPE SAS for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[24] Theoretical Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden
[25] 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
[26] Data Science and Big Data Lab, Pablo de Olavide University, 41013 Seville, Spain
[27] Department of Astrophysics, Astronomy and Mechanics, National and Kapodistrian University of Athens, Panepistimiopolis, Zografos, 15783 Athens, Greece
[28] Dipartimento di Fisica e Astronomia "Ettore Majorana", Università di Catania, Via S. Sofia 64, 95123 Catania, Italy
[29] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 Place Jules Janssen, 92190 Meudon, France
[30] Université Rennes, CNRS, IPR (Institut de Physique de Rennes) – UMR 6251, 35000 Rennes, France
[31] Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen Ø, Denmark
[32] DXC Technology, Retortvej 8, 2500 Valby, Denmark
[33] CIGUS CITIC, Department of Nautical Sciences and Marine Engineering, University of A Coruña, Paseo de Ronda 51, 15071 A Coruña, Spain
[34] IPAC, Mail Code 100-22, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[35] IRAP, Université de Toulouse, CNRS, UPS, CNES, 9 Av. colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France
[36] Thales Services for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[37] Institute of Global Health, University of Geneva, Geneva, Switzerland
[38] Applied Physics Department, Universidade de Vigo, 36310 Vigo, Spain
[39] Sorbonne Université, CNRS, UMR7095, Institut d'Astrophysique de Paris, 98bis bd. Arago, 75014 Paris, France
[40] Departamento de Astrofísica, Centro de Astrobiología (CSIC-INTA), ESA-ESAC, Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain
[41] Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
[42] Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany
[43] INAF-Osservatorio Astronomico d'Abruzzo, via M. Maggini, s/n, 64100 Teramo, Italy
[44] INFN – Sezione di Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy
[45] European Organisation for Astronomical Research in the Southern Hemisphere, Alonso de Córdova 3107, Vitacura 19001, Casilla, Santiago de Chile, Chile
[46] Institut für Astrophysik, Georg-August-Universität, 37077 Göttingen, Germany
[47] Instituto de Astrofísica de Andalucía (Consejo Superior de Investigaciones Científicas), 18008 Granada, Spain

Appendix A: Empirical training in Apsis modules

Here we provide details of algorithm training using empirical data for the modules that take this approach, namely DSC, UGC, ESP-UCD, ESP-ELS, and MSC.

A.1. DSC

DSC classifies sources into one of five classes: quasar, galaxy, star, white dwarf, or physical binary star. DSC is trained empirically, meaning it is trained on a labelled subset of the *Gaia* data that it is later applied to. We select an external catalogue for each class of objects (e.g. quasars), assumed to have high purity, and cross-match this to *Gaia* sources. The BP/RP spectra, photometry, and astrometry of those matched sources (of order 10^5) are then used as the training data for that class. Once trained, the machine learning models – ExtraTrees and Gaussian Mixture Models – map the *Gaia* data to the class probability for each source.

For the quasar and galaxy classes, sources were selected from spectroscopically confirmed objects in SDSS catalogues (Pâris et al. 2018; Aguado et al. 2019). White dwarfs were taken from the Montreal White Dwarf Database[21]. Binaries were constructed artificially by combining real BP/RP spectra of single stars; hence this class (only) is defined by objects that do not appear explicitly in the *Gaia* data. The final class in DSC, 'star', is simply a sample of objects drawn at random from *Gaia* DR3 that are not in the other training sets. This class is therefore strictly an anonymous class of objects. But as the vast majority of sources in *Gaia* are stars (well over 99%), and most of these appear single in *Gaia* data, this class is essentially equivalent to apparently single stars.

As usual in empirical training, the selection of the training objects defines the DSC classes. We therefore note that the galaxy class does not include all conceivable galaxies, but only that subset of galaxies that both *Gaia* and SDSS observe. We further note that not all types of galaxies have complete data in *Gaia* (many lack parallaxes and proper motions, meaning, for example, that they cannot be classified by DSC-Allosmod). DSC is a posterior probabilistic classifier, meaning that its published probabilities depend both on a likelihood model, and on a prior for each class. This prior is set explicitly to reflect their expected global occurrence in *Gaia* DR3, in particular the rareness of extragalactic objects. Such a prior must be used in any classification project to avoid misleading performance, as explained in more detail in Bailer-Jones et al. (2019). More details of the training sets and the DSC models can be found in the online documentation for *Gaia* DR3.

A.2. UGC

UGCuses support vector machines (SVM) through the LIBSVM (Chang & Lin 2011) package for the SVM model development, configuration, tuning, training, and testing. The SVM-models for redshift prediction are trained empirically, using a set of BP/RP spectra of unresolved galaxies observed by *Gaia* and sampled by SMSGen. The selection of sources suitable for this was taken from the SDSS DR16 archive (Ahumada et al. 2020; Blanton et al. 2017). There are 2 714 637 sources with SDSS class 'GALAXY' that have reliable photometry as well as spectroscopic redshifts. Of these sources, 1 189 812 have been observed by *Gaia* while BP/RP spectra are available for 711 600 of them. These spectra, along with the corresponding SDSS redshifts, were used for training and testing the SVM models. Because of the very small number of galaxies with redshifts lower than 0.01 and higher than 0.6, the redshift range for which the SVMs were trained was limited to $0.01 \leq z \leq 0.60$, leaving 709 449 galaxies, of which 6000 were selected for the SVM models training set, while the rest were used for performance testing.

The suitability of the galaxies selected for training has been established via a number of rules and conditions involving various source parameters. The allowed range of magnitudes was limited to $13.0 \leq G \leq 21.0$, and the galaxy class probability as estimated by DSC-Combmod had to be larger than 0.25. Moreover, the galaxy image size, as defined in SDSS by the Petrosian radius at 50% of the total flux in the r-band, was restricted to $0.5'' \leq petroRad50_r \leq 5.0''$ so as to avoid both suspiciously compact sources and very extended ones. An interstellar extinction (as defined by the SDSS r-band) upper limit of 0.5 was also applied, to exclude significantly reddened sources. Finally, it was required that the mean BP/RP spectrum was constructed from at least six epoch spectra, and that the mean fluxes lie within limits $0.3 \leq$ `bp_flux_mean` ≤ 100 e^-s^{-1} and $0.5 \leq$ `rp_flux_mean` ≤ 200 e^-s^{-1}, thus removing spectra with low S/Ns or with suspiciously high signals. Details on the SVM model preparation can be found in Delchambre et al. (2023).

A.3. ESP-UCD

The effective temperatures inferred by the ESP-UCD module come from a Gaussian Process regression model (Rasmussen & Williams 2006) trained on the set of *Gaia* RP spectra described below. Empirical training was adopted for ESP-UCD because systematic differences were found between the simulated RP spectra (see Sect. 4.6 above) obtained for the synthetic library of BT Settl spectra (Allard et al. 2013) and the observed RP spectra of ultracool dwarfs in the *Gaia* UltraCool Dwarf Sample (Smart et al. 2017, 2019, hereafter GUCDS). The empirical training set is composed of a total of 995 observed *Gaia* RP spectra. The first 36 are GUCDS sources defined as spectral standards and their effective temperatures are derived from the calibration by Stephens et al. (2009). We then added 282 sources with temperature determinations from high resolution spectra (Passegger et al. 2018) or derived using interferometric radii (see e.g. Rabus et al. 2019; Dieterich et al. 2014), most (but not all) of them are hotter than the UCD regime but included to allow for extrapolation. These 36+282 examples with effective temperature estimates were not enough and did not cover homogeneously the range of effective temperatures of the UCD regime. In order to complete the training set, we added 679 sources selected from the *Gaia* Catalogue of Nearby Stars (GCNS Gaia Collaboration 2021b), lying outside the high source density regions of the sky (Galactic bulge and disc, and the Magellanic Clouds), with $G < 19$ mag and $G - G_{RP} < 1.8$ mag. Their RP spectra were visually inspected to confirm their UCD nature. Since these 679 sources did not have effective temperature estimates, we had to infer them – as described below – in order to use them as part of the training set.

The data set then consists of 995 RP spectra with 120 flux values each. The features of these RP spectra were assumed to vary smoothly as a function of temperature so that we could use the labelled spectra (those of the 36+282 sources with effective temperatures in the literature) to calibrate the relation and assign

[21] http://www.montrealwhitedwarfdatabase.org/tables-and-charts.html

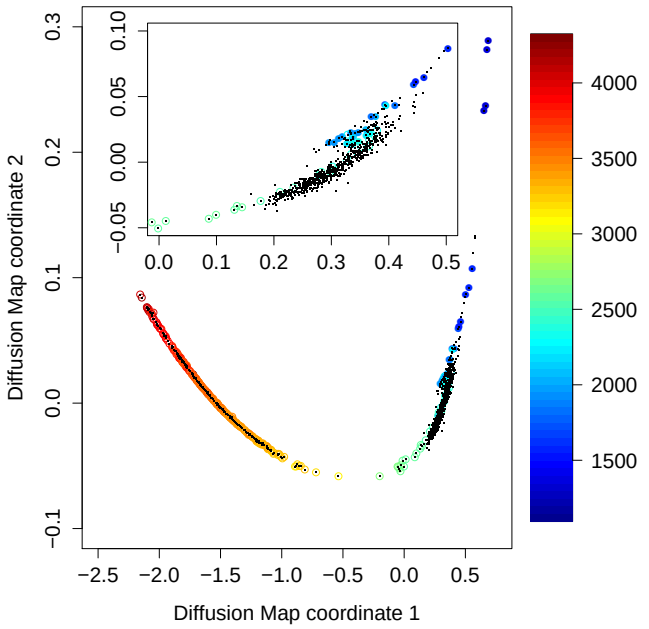


Fig. A.1. Projection onto the first two diffusion map coordinates of the set of 995 spectra defined in Sect. A.3. The colour code reflects the temperatures assigned according to their spectral types (filled circles) or literature values (empty circles). Black dots represent the entire set of sources used to construct the diffusion map. The inset plot is a zoom of the problematic zone around 2000 K where temperature variations causes only very subtle changes in the RP spectrum.

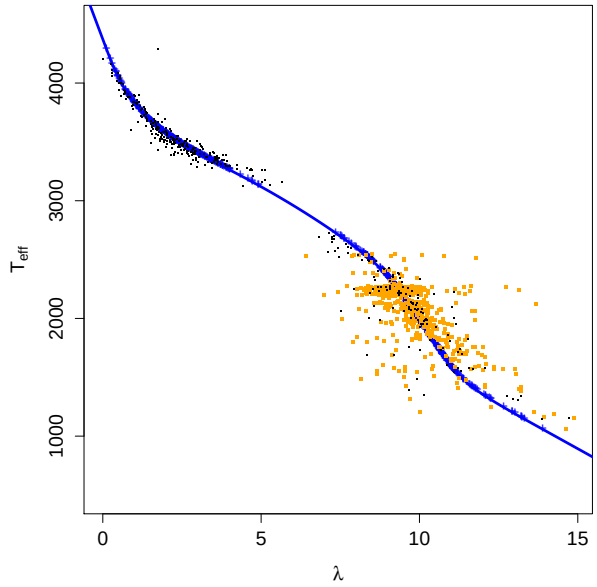


Fig. A.2. Calibration in temperature of the position of sources along the principal curve fit in the 3D space comprised of the two first diffusion map coordinates and the absolute magnitude, M_G. The position along the curve is parametrised with λ and artificial scatter is added to help visualize the number of sources used. Black points represent sources with effective temperatures derived from high-resolution spectra and GUCDS standards while orange points represent the rest of GUCDS sources.

temperatures to the unlabelled sources. If this assumption holds, it is justified to add the 679 sources to the training set. Quantifying the relation between the 120 fluxes of each spectrum and the effective temperatures is however a high-dimensional problem. In order to simplify the task, we constructed a diffusion map (Coifman & Lafon 2006) to reduce the dimensionality of the data set and found that, as hypothesised, the 995 RP spectra trace a curve in the first two diffusion map coordinates with a very small scatter around it. The two-dimensional curve shown in Figure A.1 represents in practice an ordering of the 995 RP spectra where the position along the curve (non-linearly) parametrises the temperature as shown by the coloured circles.

We used the position of the sources with effective temperature estimates from the literature to calibrate the relation between the diffusion map coordinates and temperature. This was done by fitting a principal curve (Hastie & Stuetzle 1989) to the first two diffusion map coordinates and the $G + 5\log_{10}(\varpi) + 5$ ($=M_G$) values of the 995 sources. The third coordinate was introduced to avoid the non-monotonicity of the curve in the diffusion map coordinates. The principal curve represents the minimum scatter maximum-likelihood fit and implicitly defines a parameter λ along it. We then calibrated the relation between the curve parameter λ and the effective temperature using the labelled examples and a spline regression model. The resulting calibration is shown in Figure A.2. Finally, we used this regression model to infer effective temperatures for the set of 679 sources which were then included in the training set.

A.4. ESP-ELS

Three Random Forest classifiers were used during the processing of ESP-ELS: ELSRFC1, ELSRFC2, and ELSRFC3.

Table A.1. Adopted effective temperature range covered by the spectral type tag assigned to each synthetic spectrum during the training of ELSRFC1.

tag	$T_{\rm eff}$ range [K]		
O		>	30 000
B	9 790	-	30 000
A	7 300	-	9 790
F	5 940	-	7 300
G	5 150	-	5 940
K	3 840	-	5 150
M		≤	3 840

ELSRFC1 provides a spectral type tag to each star in order to avoid processing carbon stars and to allow ESP-HS to preselect O-, B- and A-type stars. It was trained on the MARCS, OB, A, and BTSettl synthetic spectra, see Sect. 4.1, with spectral type tags: 'O', 'B', 'A', 'F', 'G', 'K', 'M' (see Table A.1) and on the observed BP/RP spectra of the Galactic carbon N stars compiled by Abia et al. (2020) with spectral type tag 'CSTAR'. The wavelength domain considered during the training varied from 340 to 600 nm in BP, and from 640 to 850 nm in RP. Both passbands were normalised individually to their respective integrated flux, while their colour indices ($G_{\rm BP}-G$) and ($G-G_{\rm RP}$) were added to the flux arrays.

ELSRFC2 identifies those BP/RP spectra belonging to the Wolf-Rayet stars, WC and WN, and planetary nebulae, PNe. To these three classes we added the tag 'unknown' to be given to all targets not showing features expected in Wolf Rayet stars or planetary nebulae. The classification is based on the flux

measured in various wavelength domains extracted from the BP/RP spectra and normalised at their edges, as well as on the astrophysical parameters (i.e. $T_{\rm eff}$, log g, and A_0). These wavelength ranges were selected following the line features generally expected to be seen in emission in various classes of ELS (see Fig. 8). Before the extraction and normalisation of the features, the spectra are divided by the instrument response provided by the MIOG simulator (Sect. 4.6). For the training, we adopted the available MARCS, OB, A, and BTSettl synthetic libraries as representative of the spectra of non-ELSs. We further added the observed BP/RP spectra of Be, Herbig Ae/Be, T Tauri, and dMe stars (i.e. as a representation of targets that are not Wolf-Rayet nor planetary nebulae), and observed BP/RP spectra of WC, WN, and PNe. The observed data of targets with known stellar classification were carefully inspected to only keep those spectra with striking and unambiguous emission features. The number of targets finally considered is given in the online documentation.

ELSRFC3 only processed targets with previously detected Hα emission. It was trained on the same features as ELSRFC2, but this time only extracted and normalised for the reference Be, Herbig Ae/Be, T Tauri, and active M dwarf stars. To the spectroscopic information, we further also added the astrophysical parameters derived by GSP-Phot during the processing (i.e. no filters and correction applied).

A.5. MSC

MSC uses an ExtraTrees algorithm (Geurts et al. 2006) to initialise its MCMC chain. The algorithm was trained on stellar parameters from a wide binary sample (El-Badry & Rix 2018) for which we artificially summed the BP/RP spectra[22]. The forward model used in the MCMC inference is based on an empirical BP/RP spectra grid (next paragraph). The distance and extinction prior use data from Rybizki et al. (2020) and the flux ratio and HR-diagram prior are based on the wide binary sample's parameter distribution.

MSC uses a model grid of empirical BP/RP spectra instead of simulated spectra. We thereby circumvent the problem of instrument modelling and the unavoidable mismatch between simulated and real spectra. The grid is a function of $T_{\rm eff}$, log g, [M/H], and A_0in the space of absolute BP/RP spectra (i.e. flux at 10 pc distance). We used the ExtraTrees machine learning algorithm (Geurts et al. 2006) on data from a sample of 80 000 APOGEE (Holtzman et al. 2015) stars for which distance and extinction estimates are available from StarHorse (Queiroz et al. 2020) and $T_{\rm eff}$, log g, and [M/H] estimates from the ASPCAP pipeline (Jönsson et al. 2020) crossmatched with *Gaia* BP/RP spectra.

Appendix B: Accessing MCMC chains for GSP-Phot and MSC

MCMC chains are provided for GSP-Phot and MSC through the Gaia Archive DataLink. There are two methods to access them.

[22] The stellar parameters were inferred from *Gaia* parallaxes and photometry using https://github.com/jan-rybizki/isochrone_fitting_example with PARSEC isochrones (Marigo et al. 2017) under the assumption of equal age, extinction, metallicity and distance. The extinction was fixed using a 3D extinction map from Rybizki et al. (2020).

First, for any ADQL query results, the user can click on the rightmost symbol showing two interlocked links. This will open a pop-up window where the user needs to select 'Gaia DR3' as data release and 'RAW' as data structure, and the MCMC data available for download will be listed (together with, e.g., BP/RP spectra also available for download). However, this is limited to a maximum of 5000 MCMC chains.

Second, MCMC chains (and also BP/RP or RVS spectra) can be downloaded without the 5000 limit using Python. A tutorial with example Python scripts for downloads can be found here: https://www.cosmos.esa.int/web/gaia-users/archive/datalink-products#datalink_jntb_get_above_lim

Appendix C: Accessing outlier results

The OAproducts are essentially contained in three different tables of the archive:

- oa_neuron_information table contains non-multidimensional parameters related to the neurons, such as statistical descriptions for different Gaia products (G, $G_{\rm BP}$, $G_{\rm RP}$, ϖ, $G_{\rm BP}-G_{\rm RP}$, etc.), and some quality measurements about the clustering itself. Moreover, a class label is also provided for the best quality neurons.
- oa_neuron_xp_spectra table contains multidimensional data related to BP/RP spectrophotometry of the neurons, so that the preprocessed spectra for both prototypes (xp_spectrum_prototype_flux) and templates (if available, xp_spectrum_template_flux) for each neuron can be retrieved.
- astrophysical_parameters table contains information about the correspondence between the sources and the neurons, among other results produced by many different modules from Apsis. Regarding OAparameters, these are the identification of the neuron within which a source lies (if it was processed by OA, neuron_oa_id), distance between the source BP/RP spectra and the neuron prototype (neuron_oa_dist, and neuron_oa_dist_percentile_rank). Additionally, it also provides a processing flag (flags_oa).

The following examples can guide the interested users to access OAresults in the Gaia Archive:

1. Retrieve the identifications of all the sources that belong to a certain neuron. In this example the first neuron (located at (0, 0), and identified by neuron_id = 201052812054400000) is used.
   Query:
```
SELECT a.source_id
FROM gaiadr3.oa_neuron_information n
JOIN gaiadr3.astrophysical_parameters a ON
n.neuron_id = a.neuron_oa_id
WHERE n.neuron_id = 201052812054400000;
```

2. Retrieve the identifications of all the sources that belong to a SOM neuron that were assigned a specific class label. For this example all galaxy labels were considered. Please, note that this type of query could potentially lead to a huge amount of data being accessed.
   Query:
```
SELECT a.source_id
FROM gaiadr3.oa_neuron_information n
JOIN gaiadr3.astrophysical_parameters a ON
n.neuron_id = a.neuron_oa_id
```

```
WHERE n.class_label ILIKE '%GAL%';
```

3. Obtain all the neurons that achieve a certain quality category threshold. In this example, just high quality neurons are retrieved.
Query:
```
SELECT *
FROM gaiadr3.oa_neuron_information
WHERE quality_category < 4;
```

Appendix D: Selection function

The processing limits and filtering performed during post-processing can lead in some cases to a complicated selection function, see also Fig. 10. We summarise the filters that were imposed on the results for each module in the Apsis pipeline in Section 11.4 of the online documentation.

Appendix E: CU8 tools for the community

CU8 has made available a number of tools and datasets for the community to accompany *Gaia* DR3 APs. The datasets are made available on the Gaia cosmos webpage at https://www.cosmos.esa.int/web/gaia/dr3-auxiliary-data. The first of these is a set of parameter simulation files[23] used in Apsis. These comprise the full set of parameters that defines the BP/RP spectral simulations of stellar sources, along with extinction in G, $G_{\rm BP}$, and $G_{\rm RP}$ passbands, and the corresponding V and $E(B-V)$, see is 4.1 and 4.2 for details. We also provide the wavelength sampling scheme[24] that was used in Apsis processing.

Information on the tools provided by CU8 can be found on the Gaia DR3 tools pages https://www.cosmos.esa.int/web/gaia/dr3-software-tools. These are the following:
– A visualisation tool[25] *GUASOM flavour DR3*[26] has been developed to analyse the outputs produced by OA (Álvarez et al. 2021), and it allows the user to interactively explore the SOM maps through several visualisations.
– The results from TGE have been implemented into the Python dustmaps package[27], and allows one to retrive the total Galactic extinction based on coordinates[28].
– A G-band bolometric correction function[29] has been provided in Python[30] based on the method that was used in FLAME.
– A GSP-Phot metallicity calibration[31] tool has been made available in Python[32] based on LAMOST DR6[33]
– A set of tools to compute photometric extinction coefficients[34] has been made available through the dustapprox[35] python package.

Appendix F: DPAC Acknowledgements

The *Gaia* mission and data processing have financially been supported by, in alphabetical order by country:
– the Algerian Centre de Recherche en Astronomie, Astrophysique et Géophysique of Bouzareah Observatory;
– the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) Hertha Firnberg Programme through grants T359, P20046, and P23737;
– the BELgian federal Science Policy Office (BELSPO) through various PROgramme de Développement d'Expériences scientifiques (PRODEX) grants, the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek) through grant VS.091.16N, the Fonds de la Recherche Scientifique (FNRS), and the Research Council of Katholieke Universiteit (KU) Leuven through grant C16/18/005 (Pushing AsteRoseismology to the next level with TESS, GaiA, and the Sloan DIgital Sky SurvEy – PARADISE);
– the Brazil-France exchange programmes Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Coordenação de Aperfeicoamento de Pessoal de Nível Superior (CAPES) - Comité Français d'Evaluation de la Coopération Universitaire et Scientifique avec le Brésil (COFECUB);
– the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) Regular Project 1210992 (L. Chemin);
– the National Natural Science Foundation of China (NSFC) through grants 11573054, 11703065, and 12173069, the China Scholarship Council through grant 201806040200, and the Natural Science Foundation of Shanghai through grant 21ZR1474100;
– the Tenure Track Pilot Programme of the Croatian Science Foundation and the École Polytechnique Fédérale de Lausanne and the project TTP-2018-07-1171 'Mining the Variable Sky', with the funds of the Croatian-Swiss Research Programme;
– the Czech-Republic Ministry of Education, Youth, and Sports through grant LG 15010 and INTER-EXCELLENCE grant LTAUSA18093, and the Czech Space Office through ESA PECS contract 98058;
– the Danish Ministry of Science;
– the Estonian Ministry of Education and Research through grant IUT40-1;
– the European Commission's Sixth Framework Programme through the European Leadership in Space Astrometry (ELSA) Marie Curie Research Training Network (MRTN-CT-2006-033481), through Marie Curie project PIOF-GA-2009-255267 (Space AsteroSeismology & RR Lyrae stars, SAS-RRL), and through a Marie Curie Transfer-of-Knowledge (ToK) fellowship (MTKD-CT-2004-014188); the European Commission's Seventh Framework Programme through grant FP7-606740 (FP7-SPACE-2013-1) for the *Gaia* European Network for Improved data User Services (GENIUS) and through grant 264895 for the *Gaia* Research for European Astronomy Training (GREAT-ITN) network;
– the European Cooperation in Science and Technology (COST) through COST Action CA18104 'Revealing the Milky Way

[23] https://www.cosmos.esa.int/web/gaia/dr3-astrophysical-parameter-inference
[24] https://www.cosmos.esa.int/web/gaia/dr3-aps-wavelength-sampling
[25] https://www.cosmos.esa.int/web/gaia/dr3-oa-self-organising-map-tool
[26] https://guasom.citic.udc.es
[27] https://dustmaps.readthedocs.io
[28] https://www.cosmos.esa.int/web/gaia/dr3-extinction-as-function-of-l-b
[29] https://www.cosmos.esa.int/web/gaia/dr3-bolometric-correction-tool
[30] https://gitlab.oca.eu/ordenovic/gaiadr3_bcg
[31] https://www.cosmos.esa.int/web/gaia/dr3-gspphot-metallicity-calibration
[32] https://github.com/mpi-astronomy/gdr3apcal
[33] http://dr6.lamost.org/v2/catalogue
[34] https://www.cosmos.esa.int/web/gaia/dr3-extinction-coefficients-in-various-passbands
[35] https://mfouesneau.github.io/dustapprox/

with *Gaia*(MW-Gaia)';
– the European Research Council (ERC) through grants 320360, 647208, and 834148 and through the European Union's Horizon 2020 research and innovation and excellent science programmes through Marie Skłodowska-Curie grant 745617 (Our Galaxy at full HD – Gal-HD) and 895174 (The build-up and fate of self-gravitating systems in the Universe) as well as grants 687378 (Small Bodies: Near and Far), 682115 (Using the Magellanic Clouds to Understand the Interaction of Galaxies), 695099 (A sub-percent distance scale from binaries and Cepheids – CepBin), 716155 (Structured ACCREtion Disks – SACCRED), 951549 (Sub-percent calibration of the extragalactic distance scale in the era of big surveys – UniverScale), and 101004214 (Innovative Scientific Data Exploration and Exploitation Applications for Space Sciences – EXPLORE);
– the European Science Foundation (ESF), in the framework of the *Gaia* Research for European Astronomy Training Research Network Programme (GREAT-ESF);
– the European Space Agency (ESA) in the framework of the *Gaia* project, through the Plan for European Cooperating States (PECS) programme through contracts C98090 and 4000106398/12/NL/KML for Hungary, through contract 4000115263/15/NL/IB for Germany, and through PROgramme de Développement d'Expériences scientifiques (PRODEX) grant 4000127986 for Slovenia;
– the Academy of Finland through grants 299543, 307157, 325805, 328654, 336546, and 345115 and the Magnus Ehrnrooth Foundation;
– the French Centre National d'Études Spatiales (CNES), the Agence Nationale de la Recherche (ANR) through grant ANR-10-IDEX-0001-02 for the 'Investissements d'avenir' programme, through grant ANR-15-CE31-0007 for project 'Modelling the Milky Way in the *Gaia*era' (MOD4Gaia), through grant ANR-14-CE33-0014-01 for project 'The Milky Way disc formation in the *Gaia*era' (ARCHEOGAL), through grant ANR-15-CE31-0012-01 for project 'Unlocking the potential of Cepheids as primary distance calibrators' (UnlockCepheids), through grant ANR-19-CE31-0017 for project 'Secular evolution of galxies' (SEGAL), and through grant ANR-18-CE31-0006 for project 'Galactic Dark Matter' (GaDaMa), the Centre National de la Recherche Scientifique (CNRS) and its SNO *Gaia*of the Institut des Sciences de l'Univers (INSU), its Programmes Nationaux: Cosmologie et Galaxies (PNCG), Gravitation Références Astronomie Métrologie (PNGRAM), Planétologie (PNP), Physique et Chimie du Milieu Interstellaire (PCMI), and Physique Stellaire (PNPS), the 'Action Fédératrice *Gaia*' of the Observatoire de Paris, the Région de Franche-Comté, the Institut National Polytechnique (INP) and the Institut National de Physique nucléaire et de Physique des Particules (IN2P3) co-funded by CNES;
– the German Aerospace Agency (Deutsches Zentrum für Luft- und Raumfahrt e.V., DLR) through grants 50QG0501, 50QG0601, 50QG0602, 50QG0701, 50QG0901, 50QG1001, 50QG1101, 50QG1401, 50QG1402, 50QG1403, 50QG1404, 50QG1904, 50QG2101, 50QG2102, and 50QG2202, and the Centre for Information Services and High Performance Computing (ZIH) at the Technische Universität Dresden for generous allocations of computer time;
– the Hungarian Academy of Sciences through the Lendület Programme grants LP2014-17 and LP2018-7 and the Hungarian National Research, Development, and Innovation Office (NKFIH) through grant KKP-137523 ('SeismoLab');
– the Science Foundation Ireland (SFI) through a Royal Society - SFI University Research Fellowship (M. Fraser);
– the Israel Ministry of Science and Technology through grant 3-18143 and the Tel Aviv University Center for Artificial Intelligence and Data Science (TAD) through a grant;
– the Agenzia Spaziale Italiana (ASI) through contracts I/037/08/0, I/058/10/0, 2014-025-R.0, 2014-025-R.1.2015, and 2018-24-HH.0 to the Italian Istituto Nazionale di Astrofisica (INAF), contract 2014-049-R.0/1/2 to INAF for the Space Science Data Centre (SSDC, formerly known as the ASI Science Data Center, ASDC), contracts I/008/10/0, 2013/030/I.0, 2013-030-I.0.1-2015, and 2016-17-I.0 to the Aerospace Logistics Technology Engineering Company (ALTEC S.p.A.), INAF, and the Italian Ministry of Education, University, and Research (Ministero dell'Istruzione, dell'Università e della Ricerca) through the Premiale project 'MIning The Cosmos Big Data and Innovative Italian Technology for Frontier Astrophysics and Cosmology' (MITiC);
– the Netherlands Organisation for Scientific Research (NWO) through grant NWO-M-614.061.414, through a VICI grant (A. Helmi), and through a Spinoza prize (A. Helmi), and the Netherlands Research School for Astronomy (NOVA);
– the Polish National Science Centre through HARMONIA grant 2018/30/M/ST9/00311 and DAINA grant 2017/27/L/ST9/03221 and the Ministry of Science and Higher Education (MNiSW) through grant DIR/WK/2018/12;
– the Portuguese Fundação para a Ciência e a Tecnologia (FCT) through national funds, grants SFRH/BD/128840/2017 and PTDC/FIS-AST/30389/2017, and work contract DL 57/2016/CP1364/CT0006, the Fundo Europeu de Desenvolvimento Regional (FEDER) through grant POCI-01-0145-FEDER-030389 and its Programa Operacional Competitividade e Internacionalização (COMPETE2020) through grants UIDB/04434/2020 and UIDP/04434/2020, and the Strategic Programme UIDB/00099/2020 for the Centro de Astrofísica e Gravitação (CENTRA);
– the Slovenian Research Agency through grant P1-0188;
– the Spanish Ministry of Economy (MINECO/FEDER, UE), the Spanish Ministry of Science and Innovation (MICIN), the Spanish Ministry of Education, Culture, and Sports, and the Spanish Government through grants BES-2016-078499, BES-2017-083126, BES-C-2017-0085, ESP2016-80079-C2-1-R, ESP2016-80079-C2-2-R, FPU16/03827, PDC2021-121059-C22, RTI2018-095076-B-C22, and TIN2015-65316-P ('Computación de Altas Prestaciones VII'), the Juan de la Cierva Incorporación Programme (FJCI-2015-2671 and IJC2019-04862-I for F. Anders), the Severo Ochoa Centre of Excellence Programme (SEV2015-0493), and MICIN/AEI/10.13039/501100011033 (and the European Union through European Regional Development Fund 'A way of making Europe') through grant RTI2018-095076-B-C21, the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia 'María de Maeztu') through grant CEX2019-000918-M, the University of Barcelona's official doctoral programme for the development of an R+D+i project through an Ajuts de Personal Investigador en Formació (APIF) grant, the Spanish Virtual Observatory through project AyA2017-84089, the Galician Regional Government, Xunta de Galicia, through grants ED431B-2021/36, ED481A-2019/155, and ED481A-2021/296, the Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC), funded by the Xunta de Galicia and the European Union (European Regional Development Fund – Galicia 2014-2020 Programme), through grant ED431G-2019/01, the Red Española de Supercomputación (RES) computer resources at MareNostrum, the Barcelona Supercomputing Centre - Centro Nacional de Supercomputación (BSC-CNS)

through activities AECT-2017-2-0002, AECT-2017-3-0006, AECT-2018-1-0017, AECT-2018-2-0013, AECT-2018-3-0011, AECT-2019-1-0010, AECT-2019-2-0014, AECT-2019-3-0003, AECT-2020-1-0004, and DATA-2020-1-0010, the Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya through grant 2014-SGR-1051 for project 'Models de Programació i Entorns d'Execució Parallels' (MPEXPAR), and Ramon y Cajal Fellowship RYC2018-025968-I funded by MICIN/AEI/10.13039/501100011033 and the European Science Foundation ('Investing in your future');
– the Swedish National Space Agency (SNSA/Rymdstyrelsen);
– the Swiss State Secretariat for Education, Research, and Innovation through the Swiss Activités Nationales Complémentaires and the Swiss National Science Foundation through an Eccellenza Professorial Fellowship (award PCEFP2_194638 for R. Anderson);
– the United Kingdom Particle Physics and Astronomy Research Council (PPARC), the United Kingdom Science and Technology Facilities Council (STFC), and the United Kingdom Space Agency (UKSA) through the following grants to the University of Bristol, the University of Cambridge, the University of Edinburgh, the University of Leicester, the Mullard Space Sciences Laboratory of University College London, and the United Kingdom Rutherford Appleton Laboratory (RAL): PP/D006511/1, PP/D006546/1, PP/D006570/1, ST/I000852/1, ST/J005045/1, ST/K00056X/1, ST/K000209/1, ST/K000756/1, ST/L006561/1, ST/N000595/1, ST/N000641/1, ST/N000978/1, ST/N001117/1, ST/S000089/1, ST/S000976/1, ST/S000984/1, ST/S001123/1, ST/S001948/1, ST/S001980/1, ST/S002103/1, ST/V000969/1, ST/W002469/1, ST/W002493/1, ST/W002671/1, ST/W002809/1, and EP/V520342/1.

The GBOT programme uses observations collected at (i) the European Organisation for Astronomical Research in the Southern Hemisphere (ESO) with the VLT Survey Telescope (VST), under ESO programmes 092.B-0165, 093.B-0236, 094.B-0181, 095.B-0046, 096.B-0162, 097.B-0304, 098.B-0030, 099.B-0034, 0100.B-0131, 0101.B-0156, 0102.B-0174, and 0103.B-0165; and (ii) the Liverpool Telescope, which is operated on the island of La Palma by Liverpool John Moores University in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias with financial support from the United Kingdom Science and Technology Facilities Council, and (iii) telescopes of the Las Cumbres Observatory Global Telescope Network.