Ultracool dwarfs in Gaia DR3

L. Sarro, A. Berihuete, R. Smart, C. Reylé, D. Barrado, M. García-Torres, W. Cooper, H. Jones, F. Marocco, O. Creevey, et al.

A&A 669, A139 (2023)
https://doi.org/10.1051/0004-6361/202244507

Astronomy & Astrophysics

Ultracool dwarfs in *Gaia* DR3⋆

L. M. Sarro[1], A. Berihuete[28], R. L. Smart[2], C. Reylé[27], D. Barrado[26], M. García-Torres[19], W. J. Cooper[17,2], H. R. A. Jones[17], F. Marocco[24], O. L. Creevey[3], R. Sordo[4], C. A. L. Bailer-Jones[8], P. Montegriffo[13], R. Carballo[15], R. Andrae[8], M. Fouesneau[8], A. C. Lanzafame[9,21], F. Pailler[5], F. Thévenin[3], A. Lobel[6], L. Delchambre[11], A. J. Korn[7], A. Recio-Blanco[3], M. S. Schultheis[3], F. De Angeli[12], N. Brouillet[14], L. Casamiquela[14,16], G. Contursi[3], P. de Laverny[3], P. García-Lario[18], G. Kordopatis[3], Y. Lebreton[22,23], E. Livanou[20], A. Lorca[10], P. A. Palicio[3], I. Slezak-Oreshina[3], C. Soubiran[14], A. Ulla[25], and H. Zhao[3]

(Affiliations can be found after the references)

Received 15 July 2022 / Accepted 9 November 2022

ABSTRACT

Context. Previous *Gaia* data releases offered the opportunity to uncover ultracool dwarfs (UCDs) through astrometric, rather than purely photometric, selection. The most recent, the third data release (DR3), offers in addition the opportunity to use low-resolution spectra to refine and widen the selection.

Aims. In this work we use the *Gaia* DR3 set of UCD candidates and complement the *Gaia* spectrophotometry with additional photometry in order to characterise the global properties of the set. This includes the inference of the distances, their locus in the *Gaia* colour-absolute magnitude diagram, and the (biased through selection) luminosity function at the faint end of the main sequence. We study the overall changes in the *Gaia* RP spectra as a function of spectral type. We study the UCDs in binary systems, we attempt to identify low-mass members of nearby young associations, star-forming regions, and clusters, and we analyse their variability properties.

Methods. We used a forward model and the Bayesian inference framework to produce posterior probabilities for the distribution parameters and a calibration of the colour index as a function of the absolute magnitude in the form of a Gaussian process. Additionally, we applied the hierarchical mode association clustering (HMAC) unsupervised classification algorithm for the detection and characterisation of overdensities in the space of celestial coordinates, projected velocities, and parallaxes.

Results. We detect 57 young, kinematically homogeneous groups, some of which are identified as well-known star-forming regions, associations, and clusters of different ages. We find that the primary members of the 880 binary systems with a UCD belong to the thin and thick disc components of the Milky Way. We identify 1109 variable UCDs using the variability tables in the *Gaia* archive, 728 of which belong to the star-forming regions defined by HMAC. We define two groups of variable UCDs with extreme bright or faint outliers.

Conclusions. The set of sources identified as UCDs in the *Gaia* archive contains a wealth of information that will require focused follow-up studies and observations. It will help advance our understanding of the nature of the faint end of the main sequence and the stellar–substellar transition.

Key words. brown dwarfs – stars: low-mass – stars: late-type – methods: statistical – Hertzsprung-Russell and C-M diagrams – stars: pre-main sequence

1. Introduction

Since its launch in December 2013, the European Space Agency astrometric mission *Gaia* (Gaia Collaboration 2016b) has produced three data releases (Gaia Collaboration 2016a, 2018b, 2021a) and is already changing our perception of the Galaxy and the properties of different types of astronomical objects, including substellar objects. Brown dwarfs (BDs) are characterised by masses below ~0.072 $M_\odot$ (depending on the specific models and the metallicity; see, for instance, Baraffe et al. 2015) and very low temperatures and luminosities, which depend strongly on the age (BDs are significantly brighter and warmer at young ages). Although they are very numerous, these properties make them hard to find and characterise. Within the *Gaia* framework, ultracool dwarfs (UCDs) are defined as objects (both stellar and substellar) with spectral types M7 or later, and they include the spectral types L, T, and Y, which are characterised by strong and wide potassium lines in the optical and water, methane, and ammonia absorption bands in the near-IR (see Kirkpatrick et al. 1999; Martín et al. 1997; Burgasser et al. 2002; Burningham et al. 2008, amongst others).

The stellar–substellar transition is still poorly understood, and new processes come into play at these low temperatures, such as the condensation of solids inside clouds in the atmospheres. In general, the BD properties are bracketed between those of low-mass stars and massive hot planets. However, their formation mechanism is still under debate, since they can be formed in star-like scenarios – via turbulent fragmentation (Padoan & Nordlund 2004 or Hennebelle & Chabrier 2008), gravitational collapse and fragmentation (Riaz et al. 2018), or ejection from multiple protostellar systems as in Bate (2012) – or in a planet-like scenario – via disc fragmentation (Whitworth & Stamatellos 2006) followed by ejection by dynamical interactions (Veras & Raymond 2012). Additional mechanisms are possible, such as aborted stellar embryos (Reipurth & Clarke 2001),

⋆ Full Table 4 is also available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/669/A139

photo-evaporation of cores near massive stars (Whitworth & Zinnecker 2004), gravitational fragmentation of dense filaments formed in young stellar associations (Bonnell et al. 2008), or compression by turbulent flows in molecular clouds (Stamer & Inutsuka 2019). An overview can be found in Chabrier et al. (2014). In any case, the photometric and spectroscopic properties of BDs fully bridge the gap between those of stars and planets and, in fact, can be used to improve our models of exoplanetary atmospheres (grain scattering and absorption, pressure-temperature profiles, chemistry, molecular opacities, condensation, cloud formation, and rainout) since they are easier to observe and generally lack the effects provoked by the irradiation from a host star. Among the most interesting characteristics are the formation of cloud decks (iron and silicates for L dwarfs; chlorides and sulphides for T and early-type Y dwarfs; and water clouds in the coolest Y dwarfs) and non-equilibrium chemistry (Saumon et al. 2000; Zahnle & Marley 2014).

Sarro et al. (2013) estimated the expected end-of-mission number of UCDs in the *Gaia* archive: 600 objects between L0 and L5, 30 objects between L5 and T0, and 10 objects between T0 and T8. Later on, in Smart et al. (2017), we cross-matched the known UCDs with the first *Gaia* data release (DR1) and identified 321 dwarfs with spectral types between L and T, creating the *Gaia* Ultracool Dwarf Sample (GUCDS). This sample was then used as a starting point by the Data Processing and Analysis Consortium (DPAC) pipeline for parameter estimation purposes based on the *Gaia* RP spectra (resolution 50–30 between 630 and 1090 nm; see Carrasco et al. 2021; De Angeli et al. 2023). Gaia Collaboration (2018a) identified 601 UCDs by comparing very restrictive subsets of the second *Gaia* data release (DR2) with several all-sky catalogues, and Reylé (2018) extended this work and identified ≈13 000 sources in *Gaia* DR2 with spectral types ≥M7 (631 with spectral type L).

Here we discuss the UCD candidate content in the *Gaia* catalogue. This UCD content is defined and characterised using the RP spectra (De Angeli et al. 2023) in addition to all other *Gaia* measurements used in previous data releases. We define this catalogue of UCDs as composed of sources in the *Gaia* Data Release 3 (DR3) archive with estimates of the effective temperature, $T_{\rm eff}$. This implies that they were identified as UCDs by the software module ESP-UCD, which is in charge of selecting and characterising UCD candidates in the context of the *Gaia* DPAC. In the following, we refer to the set of *Gaia* sources processed by the ESP-UCD module, and therefore with $T_{\rm eff}$ estimates from it, as the *Gaia* UCD catalogue, or UCD catalogue for the sake of conciseness.

In Sect. 2 we characterise the distribution of sources in several diagrams, including external photometry, and compare it to previous compilations based on *Gaia* data. In Sect. 3 we discuss the RP spectra of the UCD candidates and compare them with the ground-based high-resolution spectra of a few examples. In Sect. 4 we describe a probabilistic model for inferring distances and luminosities (absolute magnitudes) under very simple prior assumptions. These results are affected by the selection function and can only be taken as useful first approximations of the unbiased distributions. The characterisation of the selection function and the inference of de-biased distributions will be the subject of a subsequent article. In Sect. 5 we examine potential wide binary pairs in our sample with UCD components. In Sect. 6 we study overdensities of UCDs in the celestial sphere and associate them with young stellar associations and clusters. This allows us to study the changes in RP spectra as a function of age that can potentially serve as a tool for identifying youth indicators in future data releases. In Sect. 7 we check the variability properties (produced as part of *Gaia* DR3) of the UCD candidates. Finally, in Sect. 8 we summarise the results.

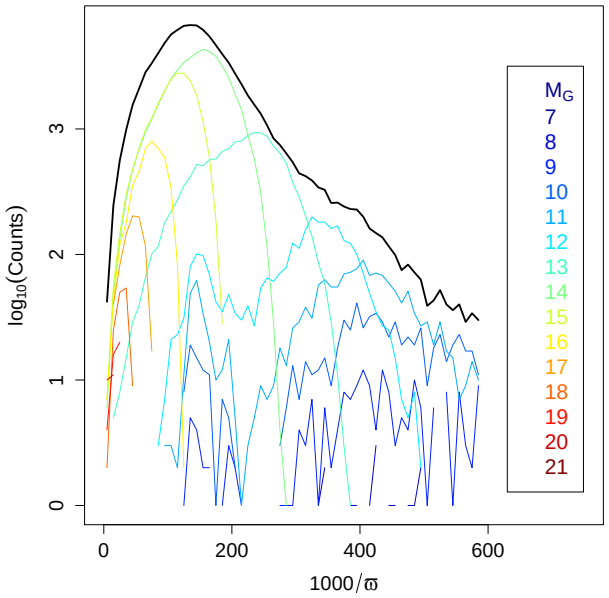

Fig. 1. Histogram of $1000/\varpi$ (with the parallax expressed in milliarcseconds) in several bins of $G + 5 \cdot \log_{10}(\varpi/1000) + 5$. The y-axis represents the decadic logarithm of the counts in each bin. The black line represents the histogram for all sources in the *Gaia* UCD catalogue regardless of their brightness.

2. The *Gaia* catalogue of UCD candidates

The *Gaia* DR3 archive contains entries for ∼1.8 billion sources including astrophysical parameters produced by the Coordination Unit 8 (CU8). This is accomplished with Apsis, the CU8 software chain composed of thirteen modules, which includes the Extended Stellar Parametrizer for UCDs ESP-UCD. The ESP-UCD software module produced $T_{\rm eff}$ estimates for 94 158 *Gaia* sources that constitute the subject of analysis of this work. The data processing in CU8 and the results included in *Gaia* DR3 are described in the CU8 chapter of the official documentation and in Creevey et al. (2023), Fouesneau et al. (2022), and Delchambre et al. (2023).

2.1. Characterisation of the Gaia UCD catalogue

Figure 1 shows the histogram (the decadic logarithm of the counts in each bin) of $1000/\varpi$ (with ϖ in units of milliarcseconds) in different absolute magnitude bins for all sources in the *Gaia* UCD catalogue. In the following we use the term absolute G magnitude (M_G) to refer to $G + 5 \cdot \log(\varpi/1000) + 5$ and likewise for other photometric bands. We are aware that this represents an oversimplification that neglects the potential effects of extinction and reddening and the naive derivation of distances (in parsecs) as the reciprocal of the parallax (again, assumed in units of milliarcseconds; see Luri et al. 2018; Bailer-Jones 2015, for a discussion of the proper inference of distances from measured parallaxes). A proper treatment of the inference of distances from observed parallaxes is included in Sect. 4. For absolute magnitudes fainter than 13 mag the histogram shows a coherent picture of decreasing typical distances for fainter sources. For sources brighter than $M_G = 13$ mag, the counts are dominated by a few stellar associations and star-forming regions; this is discussed in Sect. 6.

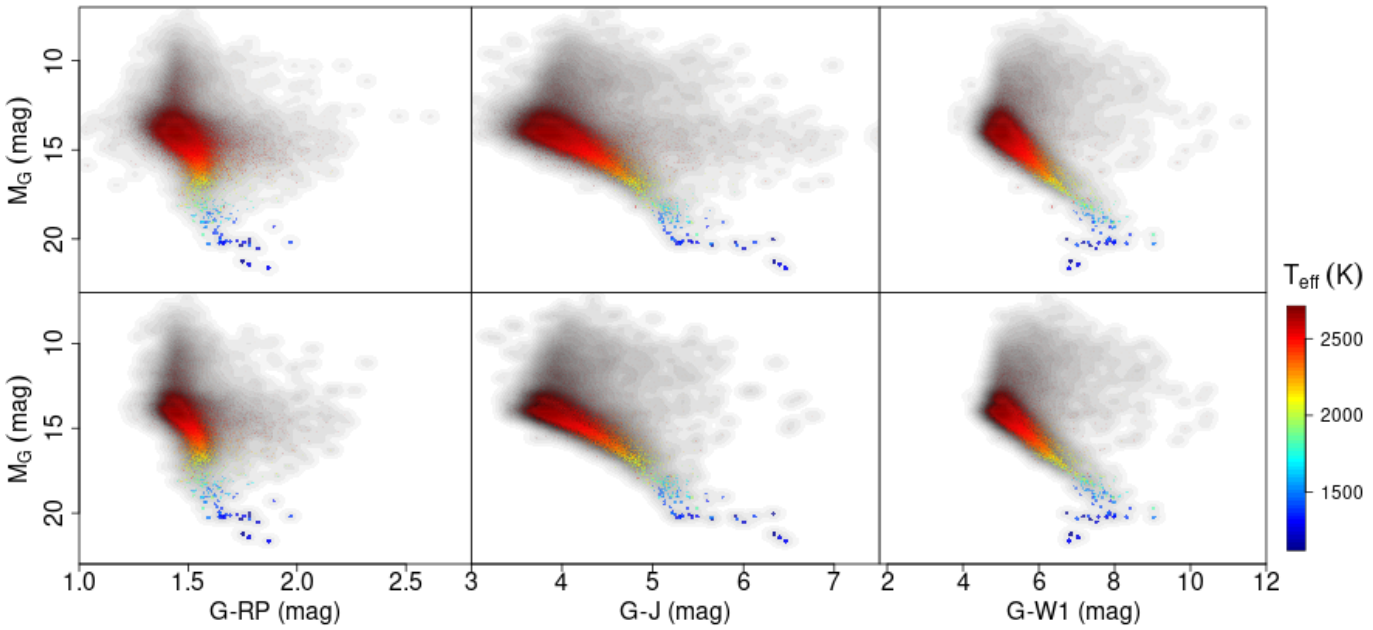

Fig. 2. CAMDs combining several *Gaia*, 2MASS, and WISE magnitudes. The top row represents all sources with available measurements, and the bottom row shows the subset with quality flags A or B in the corresponding 2MASS and AllWISE catalogues. A kernel density estimate is shown using a grey scale. The transparency and the symbol sizes were chosen to enhance the visibility of the main densities. The colour code reflects the estimated $T_{\rm eff}$ as indicated by the colour bar at the lower-right edge of the figure.

In the following we discuss the distribution of the *Gaia* UCD catalogue in the space of multi-band photometry built by adding 2MASS (Skrutskie et al. 2006) and AllWISE (Wright et al. 2010; Mainzer et al. 2011) measurements. For this purpose we used the pre-computed cross-matches allwise_best_neighbour and tmass_psc_xsc_best_neighbour in *Gaia* DR3.

Figure 2 shows the distribution in three colour-absolute magnitude diagrams (CAMDs) using the *Gaia*, 2MASS J, and AllWISE $W1$ passbands. The symbol sizes and transparencies have been chosen to enhance visibility. We plot the cube root of a kernel density estimate of the distribution of sources using a grey scale. The top row represents all sources with available measurements and the bottom row shows the subset with quality flags A or B in the corresponding 2MASS and AllWISE catalogues. The axes span the range of values of the colours and absolute magnitude. Appendix A includes other two-dimensional projections of the multi-band space.

The various CAMDs show two parallel sequences at the hotter end: the main sequence and the equal mass binary sequence (EMBS), 0.75 mag brighter. However, the EMBS seems to vanish for temperatures around 2000 K. We interpret it as the result of the drop in the number of sources fainter than $M_G \approx 13$ mag detected by *Gaia*, due to the intrinsic faintness and the shift of the peak emission towards longer wavelengths. It affects both the much denser main sequence and the EMBS, but, given that the latter is only a fraction of the former, the drop in density results in the vanishing of the EMBS, as previously observed in *Gaia* data (Gaia Collaboration 2018a). The overdensity above the locus of these parallel sequences is mainly due to the stellar associations and star-forming regions discussed in Sect. 6.

The leftmost panels of Fig. 2 show a significant scatter to the right of the main sequence. This is a well-known problem with the *Gaia* photometry of some sources, as discussed in Gaia Collaboration (2021b) and Riello et al. (2021). It is explained by the presence of more than one source in the BP and RP windows, which adds contaminating flux from the secondary source(s). This results in more flux in BP or RP, a decrease in the $G_{\rm BP}$ and $G_{\rm RP}$ magnitudes and the subsequent increase in colour indices that include the (unaffected) G magnitude.

2.2. Comparison with other catalogues

As mentioned above, the ESP-UCD module produced 94 158 UCD candidates with $T_{\rm eff}$ estimates below 2700 K. Given the luminosity function (see for example Bardalez Gagliuffi et al. 2019, and references therein, for a recent estimate), the actual number of UCD candidates at the faint end represents only a small fraction of these 94 158 sources, which are dominated by the brighter regime. In order to further assess the selection function applied to the full *Gaia* catalogue we compare here the *Gaia* UCD catalogue with previous compilations of UCD candidates based on *Gaia* data. We note that the main difference resides in the fact that this catalogue is based on the observed *Gaia* RP spectra and hence can be expected to be cleaner but also potentially more incomplete due to the quality criteria imposed on the spectra.

2.2.1. The *Gaia* Ultracool Dwarf Sample (GUCDS)

The GUCDS (Smart et al. 2017) is a continuous effort to match existing UCD catalogues in the literature with the *Gaia* catalogue. We use the version produced in May 2022 that contains 20108 entries corresponding to UCDs and any companions identified and spectroscopically confirmed in the literature. This list of objects is cross-matched using a cone search with a large 5″ radius to the *Gaia* DR3. This is a relaxed cross-match that can lead to many mismatches, so for each entry we estimate its *Gaia* G magnitude from the 2MASS J magnitude and published spectral type, if this is more than 2 magnitudes away from the *Gaia* DR3 value we reject the match. We then check any outliers in various colour-magnitude diagrams and we manually remove the misidentifications. There are 5856 entries in the GUCDS

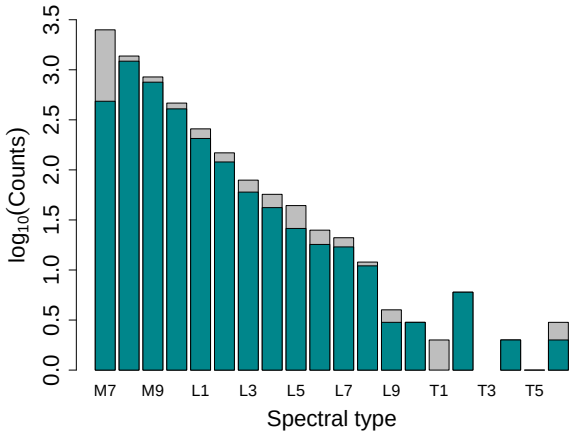

Fig. 3. Distribution of the spectral types of GUCDS sources with *Gaia* cross-matches (grey) and of those included in the *Gaia* UCD catalogue (turquoise). The vertical axis is the decadic logarithm of the counts.

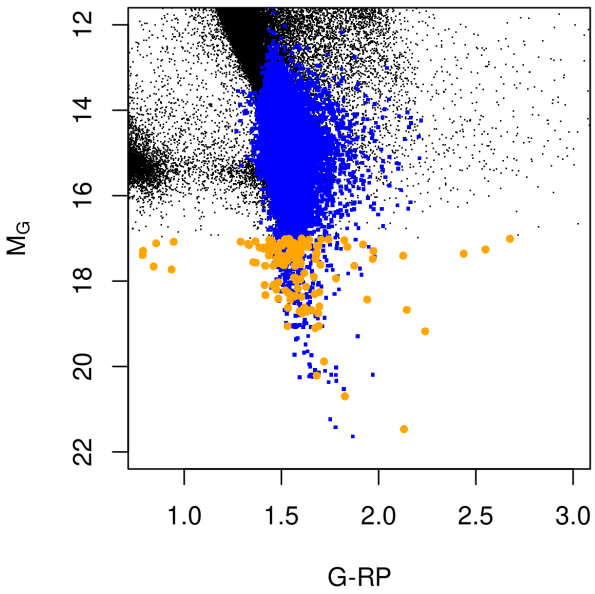

Fig. 4. CAMD of the GCNS sources (black) in the *Gaia* DR3 UCD catalogue (blue). Orange circles highlight the GCNS entries with $M_G >$ 17 mag missing from the UCD catalogue.

with spectral types later than or equal to M7 (corresponding to a temperature of 2656 K according to Stephens et al. 2009) and a *Gaia* DR3 identifier. Of these, only 5201 have the three *Gaia* measurements that are required to be included in the ESP-UCD input list (ϖ, G and $G_{\rm RP}$), and 5180 fulfil the input list selection criteria ($\varpi > 1.7$ mas and $(G - G_{\rm RP}) > 1.0$ mag). Finally, 4206 of the 5180 are included in the *Gaia* DR3 UCD catalogue. Hence, there are 974 sources identified in the literature as UCDs but excluded from the *Gaia* UCD catalogue because they did not fulfil one or several of the selection criteria established for publication. For illustration purposes, 144 of them did not fulfil the RP flux percentiles criteria[1], 362 have temperature estimates $T_{\rm eff}$ > 2700 K and 384 fail the criterion on the Euclidean distance to the set of training set templates. The rest are excluded due to the astrometric quality criterion ($\log_{10}(\sigma_\varpi) < -0.8 + 1.3 \cdot \log_{10}(\varpi)$) or a number of RP transits less than 15. Figure 3 shows the spectral type distribution of the GUCDS sources and of those in the UCD catalogue. The apparent incompleteness of the leftmost bin is due to the sharp cut in estimated $T_{\rm eff}$ that only covers part of the bin temperature range.

Hence, incompleteness of our catalogue is due mostly to unavailability of the measurements required for selection by ESP-UCD (11%), indications of temperatures above the UCD limit (6%), or quality criteria (8%).

2.2.2. The DR2 catalogue of UCD candidates by Reylé (2018)

We used the *Gaia* archive mapping between DR2 and DR3 source identifiers to track the UCD candidate list by Reylé (2018) in our catalogue. There are 14 915 sources in the original list based on DR2 data. Of these, 12 656 are included in the DR3 catalogue of UCDs. The remaining 2259 were rejected due to the RP flux percentile filters (210) or because the estimated $T_{\rm eff}$ > 2700 K (1928). The remaining 121 fail the criteria for the number of transits or distance to the templates.

The number of UCDs retrieved in this work is much larger. The main reason is that Reylé (2018) applied very strict filters on the data based on astrometric and photometric features to define a list of robust candidates, with a simple selection from their locus in the CAMD. At that point and without the possibility of confirmation from RP spectra, the strict quality filters imposed were the only possibility to avoid the many potential contaminants. Another reason for the larger number of candidates in DR3 is that the $T_{\rm eff}$ selection we used includes objects earlier than M7 (the *Gaia* DR3 UCD limit is 2700 K while, according to Stephens et al. 2009 M7 corresponds to 2656 K). Nevertheless, if we focus on the region of the CAMD used by Reylé (2018) to select M7 and later dwarfs, the availability of the RP spectra combined with the use of the Gaussian process regression module allows us to retrieve about 60 000 objects rejected by Reylé (2018) including sources in young associations as described in Sect. 6.

2.2.3. The *Gaia* Catalogue of Nearby Stars

Finally, we compare the UCD content of the *Gaia* Catalogue of Nearby Stars (GCNS; Gaia Collaboration 2021b). In this case, a direct comparison of the samples is difficult because the GCNS is not restricted to UCDs. For illustration purposes, we consider the subsample of sources with $G + 5 \cdot \log(\varpi/1000) + 5 > 17$ mag. There are 155 GCNS sources in that subsample missing from the *Gaia* DR3 UCD catalogue. Of these, 122 fulfil the ESP-UCD input list criteria (again, $\varpi > 1.7$ mas and $(G - G_{\rm RP}) > 1.0$ mag) but are rejected on the basis of the quality of the RP spectra[2] (103 cases). Also, 19 sources are missing because the mean RP spectrum was derived from fewer than 15 transits. Figure 4 shows the two catalogues in the *Gaia* CAMD, with the GCNS represented with black dots, those also included in the *Gaia* DR3 UCD catalogue as blue small circles and the missing sources

[1] In the first stage, ESP-UCD candidates are retained only if they have a sufficient fraction of the total RP flux at very red wavelengths. Let q_N denote the pixel position where the Nth percentile of the RP spectrum flux distribution (accumulated from low to high wavelengths) is attained. Then, the selection sub-module requires that $q_{33} > 60$, $q_{50} >$ 71 and $q_{67} > 83$, which reduces the 47 million input sources to 8.3 millions. For reference, pixel positions 60, 71 and 83 correspond to central wavelengths of approximately 776.9, 818.6, and 858.6 nm, respectively.

[2] $T_{\rm eff}$ estimates for sources with total negative fluxes in the normalised RP spectrum greater than −0.1, with a median RP curvature $\tau \geq 2.0 \times 10^{-5}$, or with fluxes at the reddest bin greater than 0.015 were not selected for publication.

fainter than $M_G = 17$ mag as orange circles. The cool end of the white dwarf (WD) sequence is visible as black dots with $G - G_{RP}$ colour indices bluer that approximately 1.

3. RP Spectra

The *Gaia* DR3 includes for the first time the BP and RP low-resolution spectra described in De Angeli et al. (2023). These are internally calibrated spectra affected by the instrumental response, the line spread function and the wavelength-dependent dispersion relation. Only differential effects are tackled by the internal calibration: variations of the instrument across the different observing conditions (time, CCD, field of view, window class, and gate configurations). In fact, the *Gaia* DR3 does not include the internally calibrated spectra themselves but the coefficients of their representations in a set of basis functions as described in De Angeli et al. (2023) and Montegriffo et al. (2023). In this work we analyse and concentrate on the *Gaia* internally calibrated spectra published as part of the *Gaia* DR3 in the form of coefficients. Since UCDs are intrinsically faint and very red objects, their BP spectra only contain noise in all but the brightest and hottest candidates and even there, only at the reddest wavelengths. Hence, we only discuss here the internally calibrated RP spectra, abbreviated hereafter as RP spectra. In this section we briefly mention the externally calibrated spectra. These are available from the *Gaia* archive only for sources with G magnitudes brighter than 15 mag (48 UCDs) but can be generated from the coefficients used to represent the internally calibrated spectra using the *GaiaXPy* software[3]. The externally calibrated spectra of UCDs present problems discussed in Montegriffo et al. (2023) and illustrated in Fig. 7 below.

Figure 5 shows in turquoise, simulations of BT-Settl synthetic spectra obtained using the mean instrument object generator (MIOG; briefly described in Creevey et al. 2023). The spectral types were assigned from the BT-Settl $T_{\rm eff}$ using the Stephens et al. (2009) calibration. The black lines correspond to the median RP spectrum in each spectral type (assigned again using the same calibration and the ESP-UCD temperatures) and they aim to exemplify the appearance of a 'typical' object; the red lines show comparison objects with the same spectral types. The comparison objects were selected from the spectral classifications in Kirkpatrick et al. (2019), which had RP spectra in our validation subset and were visually similar. Three of these objects (VB 10, LP 271-25, and SIPS J1058-1548 with spectral types M8, M9, and L3, respectively) are listed in Kirkpatrick et al. (1999) too, whilst two objects (2MASS J05591914-1404488 and 2MASS J15031961+2525196 with spectral types T5 and T6, respectively) also appear listed in Burgasser et al. (2003a). All of these works represent long accepted optical spectroscopic standards of UCDs. An example of an object that was visually rejected is Kelu-1, which – probably because of its binarity – appeared too red. This comparison sample is given in Table 1. The differences between the BT-Settl models and the literature standards for each spectral type or the median RP spectra between M9 and L8 are evident and they do not simply correspond to effective temperature offsets that could be explained by a different spectral type-temperature calibration.

In Sect. 6 we study UCDs in the catalogue that we identify as young and therefore potentially useful in defining low gravity diagnostics based on their RP spectra. The detection of subdwarfs in the catalogue will be addressed in a subsequent paper (Cooper et al., in prep.).

[3] Available from https://gaia-dpci.github.io/GaiaXPy-website/

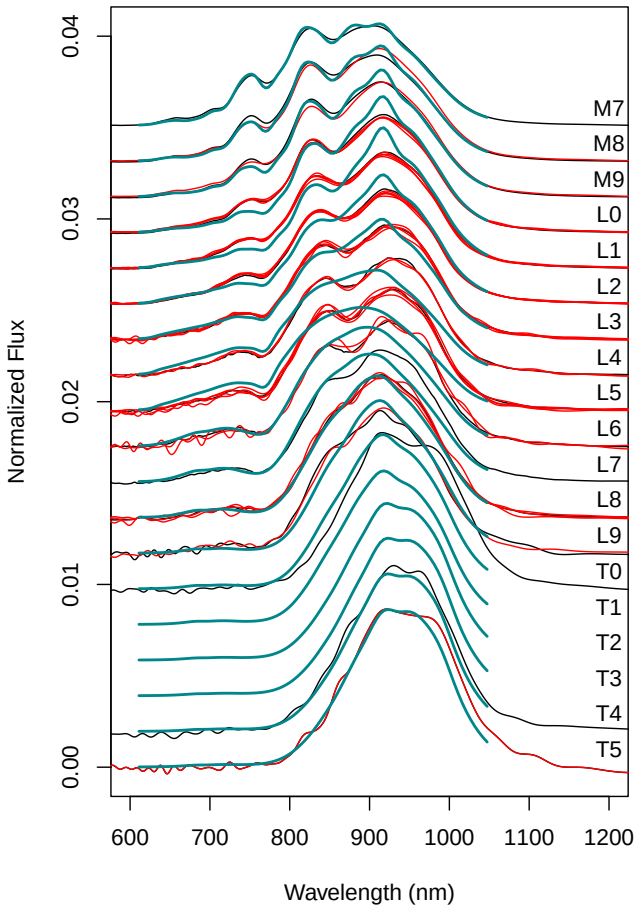

Fig. 5. RP spectra for spectral types in the UCD regime, from M7 to T6. The red lines are spectral type comparison objects, black lines represent the median RP spectrum of all UCD candidates in each spectral type, and the turquoise lines represent the MIOG simulations based on BT-Settl synthetic spectra. The conversion between effective temperatures and spectral types for the black and turquoise lines is done using the calibrations of Stephens et al. (2009).

At the low resolution typical of the RP spectra (50–30 in $\lambda/\Delta\lambda$ Montegriffo et al. 2023), individual features cannot be discerned since multiple nearby spectral features, both lines and bands, are blended and merged. The systematic changes and dependences of the RP spectra with astrophysical parameters such as $T_{\rm eff}$, log g or metallicity are not immediately evident due to this blending of spectral features. Also, because the effects in different features appear as opposing factors that can cancel each other out or partially compensate for each other. Figure 5 shows how these merged absorption features differ with spectral type. For example, the majority of RP spectra of L dwarfs have a peak near 800 nm, the strength of which (and its redward trough) is affected by the pressure broadening on the K I resonance doublet (which strengthens with spectral type, Tinney & Reid 1998); by the weakening of Na I (again with spectral type) and by a weakening of TiO (vanishing in the early L types but still present in late M).

We use ground-based optical spectra to better understand the morphological features seen in *Gaia* RP spectra. This is demonstrated in Fig. 6, which shows the *Gaia* RP spectra of six UCDs and the MIOG simulated RP spectra based on mid-resolution spectra from the GTC/OSIRIS instrument for the same set of

Table 1. List of comparison UCDs used to calibrate the ESP-UCD module empirical training set in effective temperature.

Gaia DR3 Source ID	α (hms)	δ (dms)	ϖ (mas)	Object name	Spectral type	Teff (K)
4293315765165489536	19 16 57	+5 08 39.7	169.0 ± 0.1	VB 10 [1]	M8 [2]	2404 ± 8
1287312100751643776	14 28 43	+33 10 27.9	91.2 ± 0.1	LP 271-25 [3]	M9 [4]	2238 ± 9
5761985432616501376	8 53 36	–3 29 35.4	115.5 ± 0.1	LP 666-9 [3]	M9 [4]	2272 ± 26
4595127343251508992	17 31 30	+27 21 19.2	83.7 ± 0.1	LSPM J1731+2721 [5]	L0 [6]	2233 ± 24
31235033696866688	3 14 03	+16 03 04.6	72.6 ± 0.2	2MASS J03140344+1603056 [7]	L0 [6]	2201 ± 40
3701479918946381184	12 21 28	+2 57 19.1	53.8 ± 0.2	2MASS J12212770+0257198 [7]	L0 [8]	2210 ± 41
3457493517036545280	6 02 31	+39 10 50.5	85.8 ± 0.1	LSR J0602+3910 [9]	L1 [10]	2044 ± 24
3802665122192531712	10 45 23	–1 49 57.9	58.8 ± 0.2	2MASS J10452400-0149576 [11]	L1 [12]	2073 ± 71
3808159454810609280	10 48 42	+1 11 54.5	66.6 ± 0.2	LSPM J1048+0111 [11]	L1 [8]	2077 ± 38
1649407285800074240	16 58 03	+70 26 56.7	54.1 ± 0.1	LSPM J1658+7027 [13]	L1 [13]	2069 ± 39
3460806448649173504	11 55 40	–37 27 48.2	84.7 ± 0.1	2MASS J11553952-3727350 [12]	L2 [12]	1978 ± 31
4878035808244168832	4 45 54	–30 48 27.4	61.9 ± 0.1	2MASS J04455387-3048204 [14]	L2 [15]	2017 ± 47
5723739672264914176	8 28 34	–13 09 19.4	85.6 ± 0.1	SSSPM J0829-1309 [16]	L2 [16]	1981 ± 51
851053031037729408	10 51 19	+56 13 03.6	63.9 ± 0.1	2MASS J10511900+5613086 [7]	L2 [6]	2025 ± 81
5733429157137237760	8 47 29	–15 32 40.6	57.5 ± 0.2	SIPS J0847-1532 [14]	L2 [8]	2040 ± 50
4910850870213836928	1 28 26	–55 45 32.5	53.9 ± 0.2	SIPS J0128-5545 [17]	L2 [18]	1993 ± 65
1182574753387703680	15 06 53	+13 21 05.9	85.4 ± 0.2	2MASSW J1506544+132106 [19]	L3 [13]	1787 ± 56
167202325215063168	4 01 37	+28 49 51.1	80.4 ± 0.2	2MASS J04013766+2849529 [20]	L3 [20]	1872 ± 59
1329942262499164544	16 15 44	+35 58 51.1	50.2 ± 0.3	2MASSW J1615441+355900 [21]	L3 [21]	1791 ± 216
3238449635184620672	5 00 21	+3 30 44.5	75.6 ± 0.3	2MASS J05002100+0330501 [7]	L3 [22]	1735 ± 119
3562717226488303360	10 58 48	–15 48 16.8	55.1 ± 0.3	SIPS J1058-1548 [23]	L3 [24]	1834 ± 109
6118581861234228352	14 25 28	–36 50 30.8	84.4 ± 0.3	2MASS J14252798-3650229 [25]	L4 [22]	1819 ± 52
5908794218026022144	17 53 45	–66 00 01.1	63.6 ± 0.3	SIPS J1753-6559 [7]	L4 [18]	1703 ± 147
6306068659857135232	15 07 48	–16 27 54.5	134.9 ± 0.3	2MASSW J1507476-162738 [19]	L5 [21]	1552 ± 102
2467182154313027712	1 44 36	–7 16 17.5	78.5 ± 0.5	2MASS J01443536-0716142 [26]	L5 [8]	1603 ± 90
3698979462002285824	12 03 57	+0 15 45.6	66.3 ± 0.5	2MASS J12035812+0015500 [27]	L5 [28]	1642 ± 219
3597096309389074816	12 13 03	–4 32 44.3	59.1 ± 0.6	2MASS J12130336-0432437 [14]	L5 [28]	1580 ± 152
4220379661283166720	20 02 51	–5 21 54.4	56.7 ± 1.4	2MASSI J2002507-052152 [29]	L6 [8]	1547 ± 187
4371611781971072768	17 50 24	–0 16 11.8	108.6 ± 0.2	2MASS J17502484-0016151 [30]	L6 [31]	1542 ± 71
1954170404122975232	21 48 17	+40 04 06.7	123.7 ± 0.4	2MASSW J2148162+400359 [32]	L6 [8]	1511 ± 160
4752399493622045696	2 55 05	–47 01 00.2	205.4 ± 0.2	DENIS J025503.3-470049 [33]	L8 [8]	1365 ± 38
5052876333365036928	2 57 27	–31 05 46.7	102.7 ± 0.5	2MASS J02572581-3105523 [15]	L8 [8]	1354 ± 71
1037131492704550656	8 57 58	+57 08 45.2	72.7 ± 0.7	2MASS J08575849+5708514 [11]	L8 [11]	1361 ± 316
3426333598021539840	6 07 38	+24 29 51.7	138.1 ± 0.5	2MASS J06073908+2429574 [34]	L9 [20]	1355 ± 105
2997171394834174976	5 59 20	–14 04 54.6	95.3 ± 0.7	2MASS J05591914-1404488 [35]	T5 [36]	1147 ± 86
1267906854386665088	15 03 20	+25 25 28.7	155.8 ± 0.8	2MASS J15031961+2525196 [37]	T6 [37]	1132 ± 102

Notes. Astrometry is from *Gaia* DR3 and the $T_{\rm eff}$ values are those produced by ESP-UCD and published as part of the Data Release.

References. [1]Luyten (1955), [2]Kirkpatrick et al. (1991), [3]Luyten (1979), [4]Reid & Gizis (2005), [5]Lépine & Shara (2005), [6]Reid et al. (2008), [7]Reid et al. (2006), [8]Schneider et al. (2014), [9]Lépine et al. (2002), [10]Salim et al. (2003), [11]Hawley et al. (2002), [12]Gizis (2002), [13]Gizis et al. (2000), [14]Cruz et al. (2003), [15]Schmidt et al. (2007), [16]Scholz & Meusinger (2002), [17]Deacon & Hambly (2007), [18]Marocco et al. (2013), [19]Reid et al. (2000), [20]Castro et al. (2013), [21]Kirkpatrick et al. (2000), [22]Gagné et al. (2015), [23]Delfosse et al. (1997), [24]Kirkpatrick et al. (1999), [25]Kendall et al. (2004), [26]Hall (2002), [27]Fan et al. (2000), [28]Bardalez Gagliuffi et al. (2014), [29]Cruz et al. (2007), [30]Kendall et al. (2007), [31]Burgasser et al. (2010), [32]Looper et al. (2008), [33]Martín et al. (1999), [34]Castro & Gizis (2012), [35]Burgasser et al. (2000), [36]Geballe et al. (2002), [37]Burgasser et al. (2003b).

sources. This sequence is also illustrated in Fig. 7, which shows the externally calibrated (c.f. `gaiaxpy.calibrate`) RP spectra (Montegriffo et al. 2023). It shows spurious oscillations and significant discrepancies with respect to the ground-based spectra (particularly evident in the L5 case). The appearance of these oscillations is discussed in Montegriffo et al. (2023) and is not yet fully understood. However, the apparent amplification of these wiggles at longer wavelengths is due to the fact that the externally calibrated spectral energy distributions are normalised by the inverse of the response model, which in the RP case drops quickly to very small values beyond 900 nm. These externally calibrated spectra were not used as input to Apsis or ESP-UCD. ESP-UCD and the rest of Apsis modules used only internally calibrated spectra for the prediction of astrophysical parameters.

4. Bayesian distances and the luminosity function

The *Gaia* DR3 catalogue of UCDs is truncated as a result of the various selection filters described in previous sections. It reaches out to maximum barycentric distances that depend on the intrinsic brightness of the UCD and the quality criteria applied result in further incompletenesses that are difficult to formulate

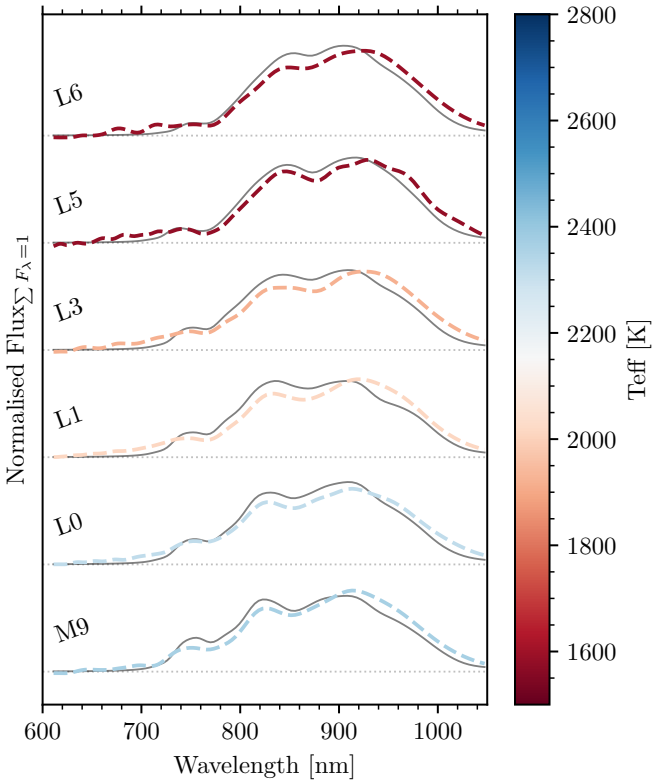

Fig. 6. Ground-based optical spectra from the GTC (in dark grey; Cooper et al., in prep.). These spectra have been simulated from the original spectra by passing through MIOG. The object short names are: J1717+6526 – L6, J1213-0432 – L5, J0453-1751 – L3, J1745-1640 – L1, J0935-2934 – L0, and J0938+0443 – M9. Over-plotted are the corresponding observed RP spectra of the same objects coloured by effective temperature and labelled by spectral type. All fluxes are normalised by the area and linearly offset.

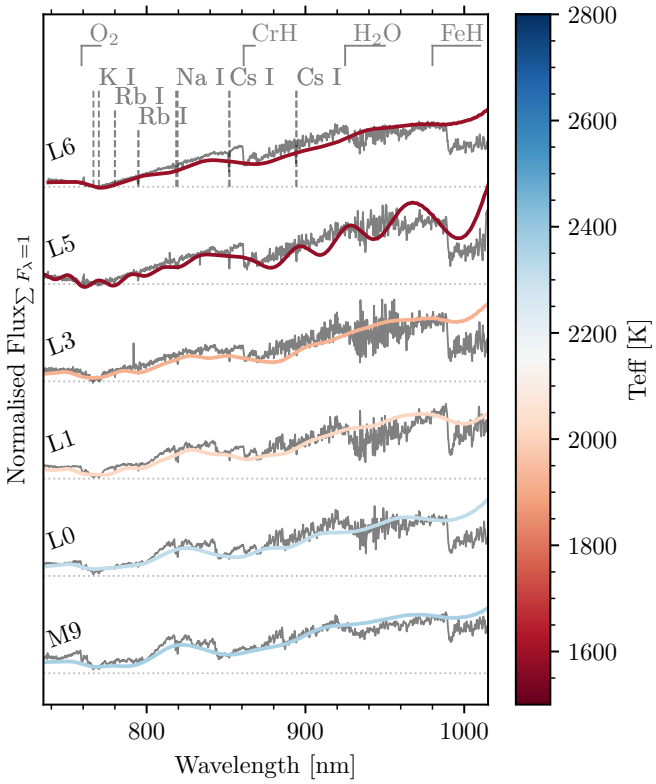

Fig. 7. Same as Fig. 6 but for externally calibrated RP spectra. The calibrated spectra were constructed using the `gaiaxpy.calibrate` function. The GTC spectra as shown here have not been passed through MIOG and represent the actual spectra with resolution ≈2500. A selection of features typical of late-M to mid-L dwarfs are shown above.

in terms of the UCD properties. The analysis of the selection function is out of the scope of this work; it will be addressed in a subsequent paper and the lessons learnt will be incorporated to the ESP-UCD processing for *Gaia* DR4. In this Sect. we elaborate on a simple model that aims to infer UCD population properties as a first step towards a full probabilistic treatment that takes into account the selection biases inherent to the production of the catalogue. In particular, we attempt to infer true distances from the *Gaia* observations, the relationship between absolute G magnitudes and the $(G - G_{\rm RP})$ at the faint end of the main sequence, and the empirical (that is, affected by truncation) luminosity distribution.

We have defined a hierarchical Bayesian model that attempts to capture the probabilistic relationship between unobserved physical quantities (distances, absolute magnitudes, etc.) and the observations $\mathcal{D}$. In Bayesian inference, the posterior probability of the model parameters, $\boldsymbol{\theta}$ (the object of our inference), is related to the prior probability and the likelihood according to Bayes' theorem:

$$p(\boldsymbol{\theta}|D,\mathcal{H}) \propto p(D|\boldsymbol{\theta},\mathcal{H}) \cdot p(\boldsymbol{\theta}|\mathcal{H}), \quad (1)$$

where $p(\boldsymbol{\theta}|D,\mathcal{H})$ is the posterior probability of the model parameters $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\mathcal{H})$ the prior probability distribution, and $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{H})$ the likelihood. The notation $\mathcal{H}$ refers to all the physical assumptions that define the hierarchical model. In our case, $\boldsymbol{\theta}$ is a high-dimensional parameter vector, and, for the sake of clarity, we distinguish between two components of $\boldsymbol{\theta}$ involved in different parts of the model's hierarchy. We subdivide the set of all parameters as $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathcal{GP}}, \boldsymbol{\theta}_{\rm star})$, where $\boldsymbol{\theta}_{\rm star} = (\mathbf{d}, \mathbf{M}_G, \mathbf{G}_{\rm RP})$ represents the vectors of true values of the distance ($\mathbf{d}$), absolute magnitude ($\mathbf{M}_G$) and $\mathbf{G}_{\rm RP}$ magnitude for the set of N stars used for inference. The relation between the true absolute magnitude M_G and the true colour index $G - G_{\rm RP}$ is modelled with an approximation of a Gaussian process (Higdon 2002; Rasmussen & Williams 2006, see Fig. 8 and the explanation below) and $\boldsymbol{\theta}_{\mathcal{GP}} = (\mu_{G-G_{\rm RP}}, \boldsymbol{\beta}, \sigma_\beta, \gamma, \delta)$ is the subset of parameters that defines that approximation. Taking into account this notation, the prior distribution in Eq. (1) is such that

$$P(\boldsymbol{\theta}|\mathcal{H}) = P((\boldsymbol{\theta}_{\mathcal{GP}}|\mathcal{H}) \cdot P(\boldsymbol{\theta}_{\rm star}|\boldsymbol{\theta}_{\mathcal{GP}}, \mathcal{H}).$$

Figure 8 shows graphically the probabilistic hierarchical model and Table 2 lists the (hyper-)parameters prior probability distributions specified in the following paragraphs where we also explain the conditional relations underlying the model.

At the top level of the graph, the prior for the absolute magnitude M_G is a PERT(13.7, 14.49, 21.5, 21.73) distribution (see Appendix B.1 for a definition of the PERT distribution and parameters). The PERT 'peak' and 'temperature' parameters (14.49 and 21.73, respectively) are the maximum likelihood estimates obtained after fixing the 'low' and 'high' parameters to 13.7 and 21.5 mag, respectively (values that include and extend the range of absolute magnitudes obtained by naive inversion of the parallaxes). The parameter name temperature can be misleading in this context but we maintain it for consistency with the literature. This distribution peaks at the same absolute magnitude as the kernel density estimate of the distribution of observed absolute magnitudes but has non-vanishing prior probability

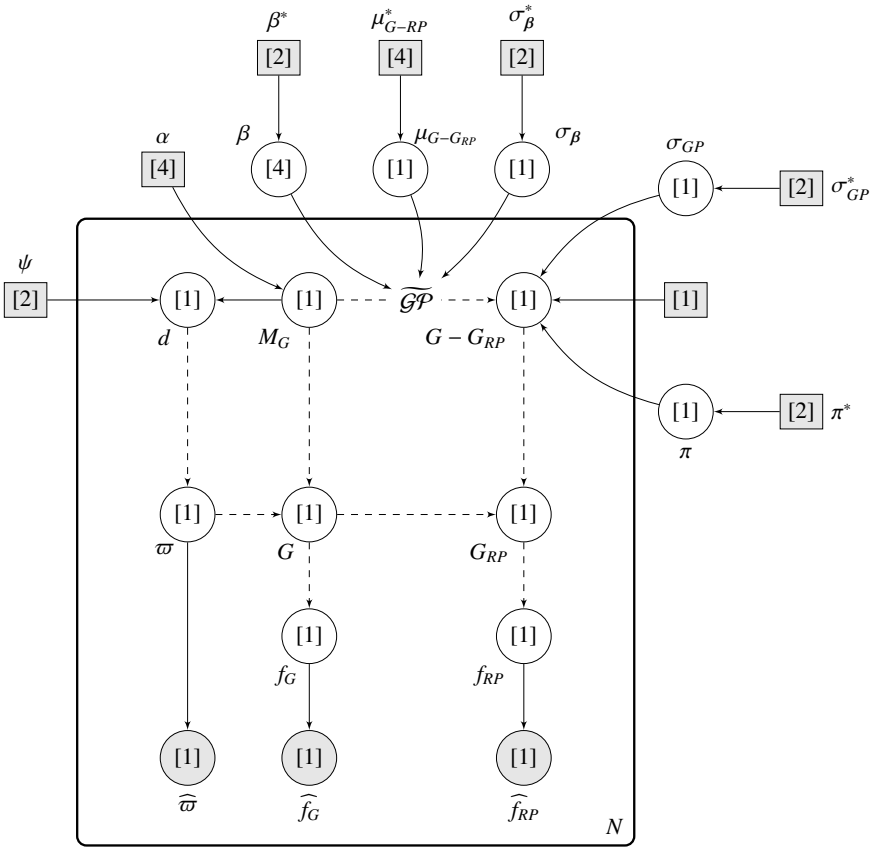

Fig. 8. Hierarchical model in plate notation. Squares indicate fixed hyperparameters, and circles indicate random variables. Filled-in shapes indicate fixed values that are not subject to inference. Dashed lines indicate a functional non-probabilistic relationship. The dimension of the parameter vectors are indicated inside the shapes in brackets, and the fixed hyperparameters are listed in Table 2. For instance, $\sigma^*_\beta = (0, 0.5)$ are the parameters of the Lognormal prior for σ_β.

Table 2. Prior distributions for the parameters in the inference model.

Parameter	Prior	Units
$\mu_{G-\mathrm{RP}}$	PERT$(1, 1.5, 2, 5)$	mag
σ_β	Lognormal$(0, 0.5)$	mag
$\boldsymbol{\beta}$	$\mathcal{N}_4(0, \sigma_\beta)$	mag
$M_{G,i}$	PERT$(13.7, 14.49, 21.5, 21.73)$	mag
d_i	CPD$(M_{G,i}, d_{\max}, d_{\mathrm{extra}})$	pc
σ_{GP}	Lognormal$(-3, 0.1)$	mag
π	$\mathcal{U}(0, 0.15)$	–
$G_i - G_{\mathrm{RP},i}$	$(1-\pi)\cdot\mathcal{N}(\widetilde{\mu}, \sigma_{\mathrm{GP}}) + \pi\cdot\mathcal{N}(\widetilde{\mu}, 1.5)$	mag

Notes. Due to its hierarchical nature, some distributions depend on others. Specific details and definition of the PERT distribution are given in the Appendix B.1.

densities in the range $(13.7, 21.5)$ mag. From the true value of M_G we obtain the probability distribution for true distance d using a custom probability distribution (CPD) defined as

$$f(x) = \begin{cases} \dfrac{x^2}{C} & x < d_{\max} \\ \dfrac{d_{\max}^2 \cdot \exp\left(\dfrac{d_{\max} - x}{H}\right)}{C} & x \geq d_{\max} \end{cases}, \tag{2}$$

where $d_{\max}$ and C are as follows

$$d_{\max} = 10^{-(M_G - 20.48 - 5)/5},$$

$$d_{\mathrm{extra}} = 10^{-(M_G - 20.7 - 5)/5},$$

$$H = \frac{d_{\max} - d_{\mathrm{extra}}}{\log(0.01)},$$

$$C = \frac{d_{\max}^3}{3} - d_{\max}^2 \cdot H \cdot \left(\exp\left(\frac{d_{\max} - d_{\mathrm{extra}}}{H}\right) - 1\right).$$

Equation (2) represents the expected distribution of distances for a uniform volume density of UCDs with an exponential decay at the maximum distance $d_{\max}$ defined by the absolute magnitude M_G and the *Gaia* limiting magnitude for completeness assumed here to be 20.48 mag (Gaia Collaboration 2021b). The exponential decay is introduced to incorporate the fact that there is no hard cut in magnitudes that can be detected by *Gaia* and is parametrised by a length scale H that defines the range of distances d_{extra} beyond $d_{\max}$ over which the decay takes place. The value of d_{extra} is defined for convenience based on a nominal limiting magnitude of 20.7 mag (Gaia Collaboration 2016a). Several simplifications are adopted here that will be lifted in subsequent investigations that will properly incorporate the selection function. For example, the vertical stratification of the Milky Way disc is a more realistic prescription of the volume density than the uniform distribution (see for example Gaia Collaboration 2021b). Additionally, neither 20.48 nor 20.7 mag are the *Gaia* limiting magnitudes as these depend on the celestial coordinates under consideration via the scanning law.

The true G_{RP} magnitude of every star given the absolute magnitude follows a probability distribution given by a Gaussian process. This Gaussian process models the faint end of

the main sequence in the CAMD constructed with M_G and $(G - G_{\rm RP})$. For the sake of computational efficiency, we use an approximate Gaussian process, $\widetilde{\mathcal{GP}}$ (see Appendix B.2), that depends on a set of parameters $\mu_{G-G_{\rm RP}}$, $\boldsymbol{\beta}$ and σ_{β}. Their prior probability densities are defined as PERT$(1, 1.5, 2, 5)$, $\mathcal{N}_4(0, 1)$ and Lognormal$(0, 0.5)$, respectively. Finally, we assume that the approximate Gaussian process describes the faint end of the main sequence, but also that for a given absolute magnitude outliers can exist that result in a skewed distribution of $G - G_{\rm RP}$ colour indices with respect to the GP prescription. We model the presence of outliers using a Gaussian mixture distribution:

$$(G_i - G_{{\rm RP},i}) \sim (1 - \pi) \cdot \mathcal{N}(\widetilde{\mu}, \sigma_{\rm GP}) + \pi \cdot \mathcal{N}(\widetilde{\mu}, 1.5), \tag{3}$$

where $\widetilde{\mu_i} = \widetilde{\mathcal{GP}}(M_{G,i}, \mu_{G-{\rm RP}}, \boldsymbol{\beta})$. Hence, the distribution of colour indices has its mode given by the Gaussian process with an added scatter term parameterised by $\sigma_{\rm GP}$. Outliers are modelled using a wide second Gaussian distribution with the same mean and a standard deviation fixed to 1.5 mag. The prior probability density for $\sigma_{\rm GP}$ is defined as a Lognormal$(-3, 0.1)$. The prior distribution for the mixture proportion follows a Uniform distribution between 0 and 0.15.

The model parameters d, M_G and $G_{\rm RP}$, all represent true values. From them, we can derive true values of the parallax ϖ and of the fluxes f_G and $f_{\rm RP}$ as

$$\begin{aligned} \varpi &= 1000/d, \\ G &= M_G - 5 \cdot \log_{10}(\varpi/1000) - 5, \\ f_G &= 10^{(G-25.6874)/-2.5}, \\ f_{\rm RP} &= 10^{(G_{\rm RP}-24.7479)/-2.5}. \end{aligned}$$

In the likelihood, we assume that the *Gaia* observations for each UCD are independent and $P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{H})$ can be factorised as

$$P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{H}) = \prod_{i=1}^{N} p(\mathcal{D}_i|\boldsymbol{\theta}, \mathcal{H}),$$

where the likelihood for each UCD, $p(\mathcal{D}_i|\boldsymbol{\theta}, \mathcal{H})$, is defined as the product of three normal distributions centred at the true values and with standard deviations given by the catalogue uncertainties:

$$p(\mathcal{D}_i|\boldsymbol{\theta}, \mathcal{H}) = \mathcal{N}(\widehat{\varpi} \,|\, \varpi, \widehat{\sigma}_{\varpi}) \cdot \mathcal{N}(\widehat{f_G} \,|\, f_G, \widehat{\sigma}_{f_G}) \cdot \mathcal{N}(\widehat{f_{\rm RP}} \,|\, f_{\rm RP}, \widehat{\sigma}_{f_{\rm RP}}). \tag{4}$$

The data comprise the measured parallaxes and fluxes in the G and RP bands of all sources in the *Gaia* UCD catalogue with quality categories 0 or 1 and their associated uncertainties:

$$\mathcal{D} = \{\widehat{\varpi}_i, \widehat{\sigma}_{\varpi,i}, \widehat{f}_{G,i}, \widehat{\sigma}_{f_G,i}, \widehat{f}_{{\rm RP},i}, \widehat{\sigma}_{f_{\rm RP},i}\}_{i=1}^{N},$$

where $\widehat{f}_{G,i}$ and $\widehat{f}_{{\rm RP},i}$ are the measured fluxes in the G and RP bands, respectively, and the index i runs between 1 and the total number of stars in the sample. The notation uses the circumflex symbol to distinguish observations (assumed affected by measurement noise) from true values. The observations amounts to a total number of stars $N = 67\,428$. From this set we select only sources with $G + 5 \cdot \log_{10}(1000/\varpi) + 5 > 13.8$ to avoid including the very young sources discussed in Sect. 6. A fraction of very young sources will still be present in the data but these will be dealt with under the category of outliers (as described later in the section). The total number of sources used in the inference is then 43 795.

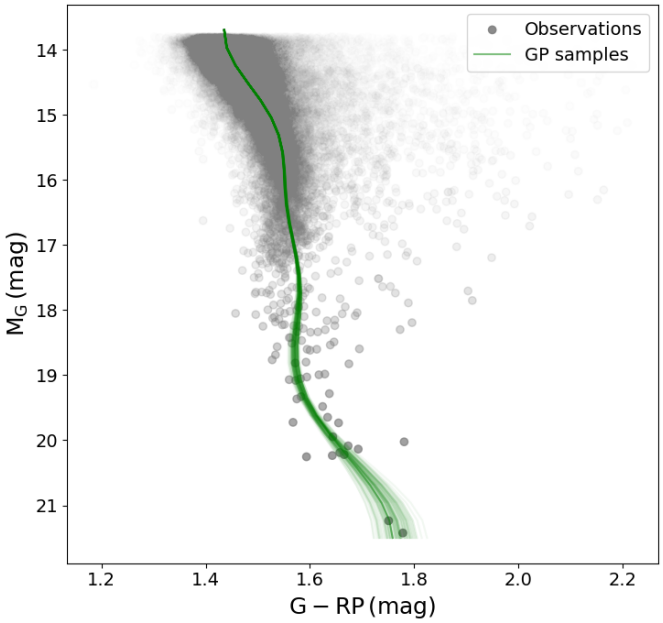

Fig. 9. CAMD with posterior samples of the Gaussian process (green) and the observations (black). The transparency and the symbol sizes were chosen to enhance the visibility of the main densities.

Table 3. Weighted average of the temperatures estimated by the ESP-UCD module for several absolute magnitudes along the mean of the Gaussian process posterior.

M_G (mag)	$T_{\rm eff}$ (K)
14	2650.5
15	2477
16	2317
17	2050
18	1836
19	1553
20	1374
21	1145

As usual in these kinds of complex problems, it is not possible to obtain a closed-form expression for the posterior probability distribution from the prior and the likelihood, and we describe the posterior distribution of the parameters using samples obtained with Markov chain Monte Carlo (MCMC) sampling techniques. Figure 9 shows the data used for the inference and 100 posterior samples of the Gaussian process (green). We used the ESP-UCD effective temperature estimates to define a relationship between M_G and an average $T_{\rm eff}$ using the Gaussian process mean, and weighting the contribution of each source to the effective temperature. Table 3 lists the average $T_{\rm eff}$ values for several M_G values along the Gaussian process sequence.

Figure 10 shows in logarithmic scale the empirical luminosity function derived from the absolute magnitudes inferred by the model (top panel). It has been derived using the 600 samples from the posterior and then dividing the counts in each bin of 0.5 mag by 600. Since we cut the initial sample at M_G = 13.8 mag, the first bin of the histogram is only partially observed

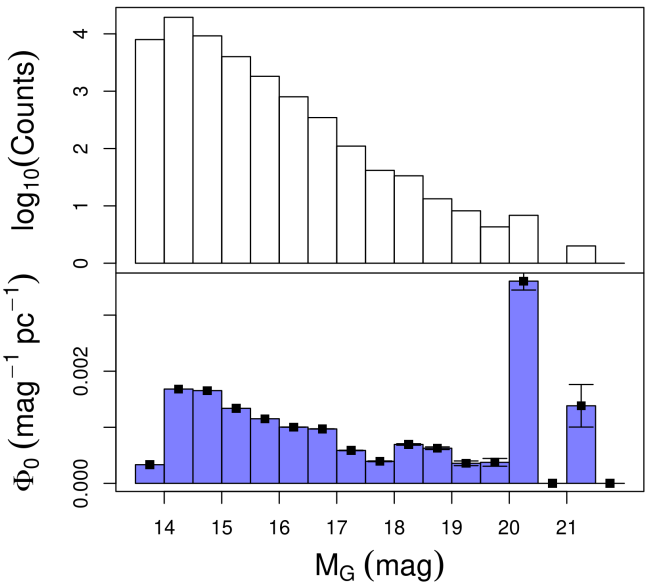

Fig. 10. Distributions inferred by the hierarchical model. Top panel: histogram of the absolute magnitudes inferred by the hierarchical Bayesian model. Bottom panel: luminosity function (in units of mag^{-1} pc^{-3}) derived as described in the text.

and can be ignored. Obviously this empirical distribution function is affected by all the selection biases derived from the filters described in Sect. 2 including the decreasing three-dimensional volume explored as the absolute magnitude bins get fainter.

The true luminosity function can in principle be recovered if we can formulate all the filters applied as selection functions. These selection filters (listed below) are described in the ESP-UCD section of the official documentation[4], in Creevey et al. (2023), and in Fouesneau et al. (2022). Rix et al. (2021) provide a simplified example of this type of reconstruction using *Gaia* data and focusing on the domain of WDs. In our case, the first factor of the selection function comes from the filters that define the sources to which ESP-UCD is applied. These are the ≈47 million sources with $G - G_{\mathrm{RP}} > 1$ and $\varpi > 1.7$ mas. While these filters may remove very young and bright sources that could be detected beyond this limit or subdwarfs redder than 1.0 mag in $G - G_{\mathrm{RP}}$, we assume that their numbers (if not zero) are small enough to neglect their effect in the reconstruction of the luminosity function. However, we assume that only 89% of the UCDs have ϖ, G, and G_{RP} measurements available (see the GUCDS analysis in Sect. 2.2).

The next selection filter is applied by the ESP-UCD module based on the RP spectrum pixels where the 33, 50 and 67 flux percentiles are attained. Since the filter definitions are inclusive definitions of the hot boundary of the UCD regime, in principle we only expect selection effects at that boundary. The dataset used for inference with the Bayesian model is cut at M_G = 13.8 mag and we do not expect significant selection effects. This is illustrated in Figs. 2 and 17 that show that M_G = 13.8 leaves out all the very young sources discussed in Sect. 6 and the brightest end of the parallel sequence of equal mass binaries.

Finally, the main selection filters applied are related to the quality of the RP data. This includes the removal of sources with fewer than 15 RP transits, high RP median curvature values (above 2.0×10^{-5}), large Euclidean distances to the set of templates d_{TS} and/or poor astrometric measurements (the ESP-UCD module removed sources with $\log_{10}(\sigma_{\varpi}) > -0.8 + 1.3 \cdot \log_{10}(\varpi)$). While the requirement on the number of RP transits depends on the celestial coordinates of each source in a way that is complex but quantifiable, the selection function related to the RP quality requirements are difficult to estimate. The quality classes used in this section (0 and 1) were defined in terms of the decadic logarithm of the Euclidean distances d_{TS} to the spectral type standards and of the relative RP flux uncertainties $\sigma_{\mathrm{RP}}/f_{\mathrm{RP}}$. As a first order approximation to the selection function, we fit mixtures of Gaussian components to the distributions of sources in the two-dimensional spaces $\log_{10}(d_{\mathrm{TS}})$ –$\sigma_{\mathrm{RP}}/f_{\mathrm{RP}}$ and $\log_{10}(\varpi)$–$\log_{10}(\sigma_{\varpi})$. We use the Gaussian components to estimate the number of sources lost in the application of the filters. We further assume that the distribution of M_G of these sources is the same as that of the UCDs with quality class 2 (the worst quality class of the three). From them, we estimate a M_G-dependent selection function. The lower panel of Fig. 10 shows the resulting point estimates of the luminosity function derived as in Rix et al. (2021) assuming an exponentially decaying spatial density with scale height H = 365 (Gaia Collaboration 2021b), Poisson distributions, and a flat prior for the number density $\Phi_0(M_G)$. The last bin of the luminosity function is affected by the low number of sources but the Poisson uncertainty makes it compatible with the trend that can be deduced from the rest of the function. The bin from M_G = 20 to 20.5 mag, however, seems inconsistent with the rest of the bins and is due to the sequence of UCDs in the CAMD running horizontally at those absolute magnitudes (see Fig. 2). This figure is in broad agreement with other distributions in the literature (for example, Kirkpatrick et al. 2021; Bardalez Gagliuffi et al. 2019) in the range of absolute magnitudes from 14 to 20 mag.

As mentioned above, this is the result of a simplified treatment of the selection function applied to the raw data and a detailed comparison with existing luminosity functions in the literature is postponed until a more accurate specification of the selection function is available.

5. UCDs in binary systems

In this section we explore the existence of co-moving pairs consisting of sources in the UCD catalogue and a primary component. We search the *Gaia* archive for potential primaries that fulfil the following criteria (where the sub-indices p and UCD denote the primary and the UCD components of the pair): (1) the projected distance (in the tangential plane of the sky and calculated using the naive inversion of the parallax and the sine of the angular separation) from the UCD candidate is less than 0.1 pc; (2) its parallax (ϖ_p) is greater than 1 mas; (3) ϖ_p is in the interval defined by $\varpi_{\mathrm{UCD}} \pm 3 \cdot (\sigma_{\varpi_p} + \sigma_{\varpi_{\mathrm{UCD}}})$; (4) $\mu_{\mathrm{RA};p}$ is in the interval defined by $\mu_{\mathrm{RA;UCD}} \pm 3 \cdot (\sigma_{\mu_{\mathrm{RA}};p} + \sigma_{\mu_{\mathrm{RA}};\mathrm{UCD}})$; (5) $\mu_{\mathrm{Dec};p}$ is in the interval defined by $\mu_{\mathrm{Dec;UCD}} \pm 3 \cdot (\sigma_{\mu_{\mathrm{Dec}};p} + \sigma_{\mu_{\mathrm{Dec}};\mathrm{UCD}})$; and (6) the absolute magnitude of the primary (assuming negligible extinction and inferred by inverting the parallax) is brighter than that of the UCD.

Adding the uncertainties in quadrature would imply a more restrictive threshold, which at this stage was not necessary. This leads to a list of 28 704 candidates primaries (that is, 28 704 potential pairs containing at least one UCD). Figure 11 and subsequent figures in this section concentrate on the 880 systems where both components have signal-to-noise ratios $\varpi/\sigma_{\varpi} > 15$. This selection removes the large fraction of unlikely primary candidates with large parallax uncertainties. These large uncertainties make them pass the selection criteria

[4] https://geapre.esac.esa.int/archive/documentation/GDR3/Data_analysis/chap_cu8par/sec_cu8par_apsis/ssec_cu8par_apsis_espucd.html

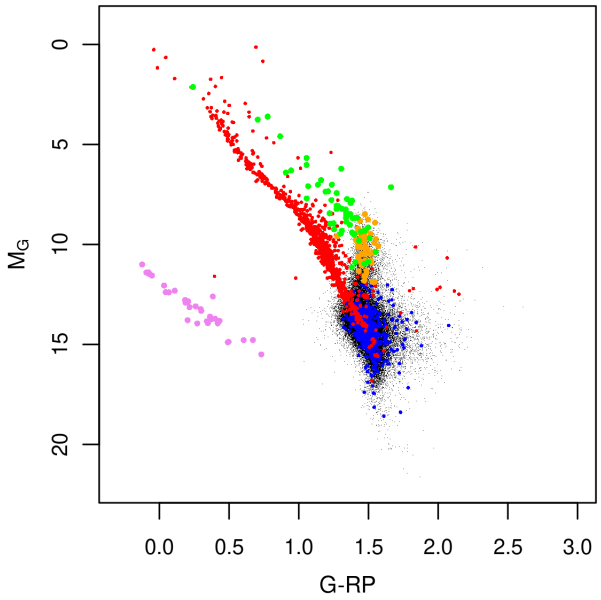

Fig. 11. CAMD for binary candidates. Black dots represent all UCD candidates, and blue and red circles represent UCD secondaries and candidate primaries, respectively, with $\varpi/\sigma_\varpi > 15$. Magenta circles correspond to sources identified as WDs using the linear boundary $M_G > 10 + 6 \cdot (G - G_{\rm RP})$. Orange circles tag UCD candidates in stellar associations and green circles their corresponding primaries.

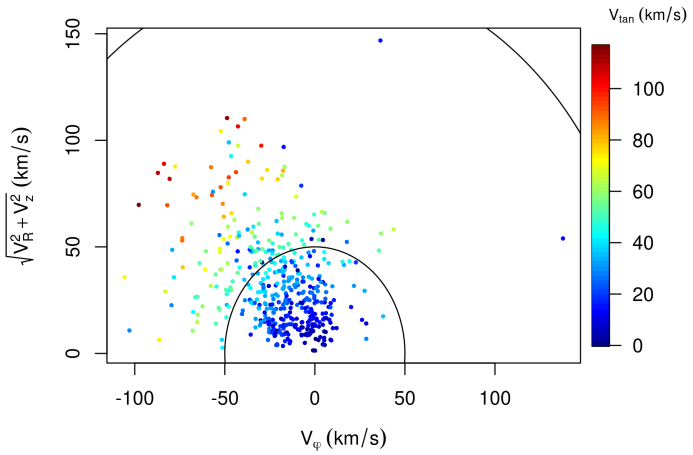

Fig. 12. Toomre diagram of the primaries with UCD companions. The colour code reflects the tangential velocities. Common boundaries for the thin–thick disc and thick disc–halo components are included as half circles at 50 and 180 km s^{-1}, respectively.

even though their parallaxes are very different from the UCD parallaxes. Thirty-two of the 880 candidate primaries are included in the UCD catalogue and thus represent cases of UCD pairs and we find two triple systems candidates.

For the candidate primaries in the WD sequence we do not find differences in the distribution of the separations with respect to main sequence primaries as could be expected if UCDs in wide binaries do not survive long because they are less gravitationally bound. In Fig. 11 we have also marked in orange UCDs with $M_G > 12$ and in green their corresponding primaries. The overluminous absolute magnitudes of these primaries are in agreement with the assumed indication of youth ascribed to the UCD components (see Sect. 6).

We also searched the *Gaia* DR3 archive for radial velocities and astrophysical parameters of the primaries. We find radial velocities available for 465 of the 880 candidate primaries and results from the FLAME module for 586 of them (see Creevey et al. 2023; Fouesneau et al. 2022, for a detailed description of the Apsis modules that produced stellar astrophysical parameters as part of the *Gaia* processing, including FLAME and GSP-Phot). Figure 12 shows the position of the primaries in the Toomre diagram. It shows $\sqrt{V_R^2 + V_z^2}$ and V_ϕ of the sources with radial velocity measurement, where (V_R, V_ϕ, V_z) are the velocity components of the stars in the Galactocentric cylindrical coordinate system, with R pointing from the Galactic centre to the Sun, z along the axis perpendicular to the Galactic plane, and ϕ along the azimuthal direction in the Milky Way disc plane (defined such that V_ϕ is positive for prograde stars in the disc). The calculation of (V_R, V_ϕ, V_z) follows the same assumptions as adopted for the selection of OBA stars in Gaia Collaboration (2023). In particular, we assume the local circular velocity from the `MWPotential2014` Milky Way model (Bovy 2015), which is 219 km s^{-1} at the distance of the Sun from the Galactic centre (8277 pc, GRAVITY Collaboration 2022). The height of the Sun above the disc plane is assumed to be 20.8 pc (Bennett & Bovy 2019) and the peculiar motion of the Sun is assumed to be $(U, V, W) = (11.1, 12.24, 7.25)$ km s^{-1} (Schönrich et al. 2010). The Toomre diagram seems to indicate that the UCDs in binary systems (with radial velocities of the primary measured by *Gaia*) are a mixture of the thin and thick disc components with no conspicuous member of the halo.

FLAME produces amongst other, luminosities, ages, masses and radii based on astrophysical parameters derived by the GSP-Phot module for ≈280 million sources. However, FLAME ages and masses are derived for a set of stellar models that do not include pre-main-sequence stages. They cover from the main sequence to the tip of the red giant branch, for masses between 0.5 and 10 solar masses, and for a solar-metallicity prior. Figure 13 shows the distribution of the masses of the (candidate) primaries with a peak at ∼0.8 $M_\odot$ larger than expected for the solar neighbourhood (around M3 or 0.3 $M_\odot$ according to Gaia Collaboration 2021b; Jao et al. 2018). This bias arises mainly from the requirement of availability of FLAME masses (with a minimum mass of 0.5 $M_\odot$) but also from the parallax S/N cut at $\varpi/\sigma_\varpi > 15$. Figure 14 shows the $T_{\rm eff}$-luminosity scatter plot colour coded by the decadic logarithm of the age when available and only for primaries with UCD companions characterised by $M_G > 12$ mag.

6. UCDs in star-forming regions, clusters, and moving groups

The distribution on the sky of sources in the *Gaia* UCD catalogue has clear overdensities apparent in the top row of Fig. 15. It shows the distribution of sources in the two best ESP-UCD categories (qualities 0 and 1; left panel) and quality 2 sources (right). The former shows overdensities that can be easily identified with open clusters and star-forming regions while the latter is dominated by an overdensity aligned with the Galactic disc that we interpret as residual contamination from bad astrometric solutions due to crowding and reddened sources.

In this section we explain how we determine membership of the UCD candidates to several groups, show the position of the candidate members in several CAMDs and illustrate the differences in the RP spectra as a function of the age (taken from the literature). The sizes and nature of the groups found can vary greatly. We use the terms cluster, association, and group loosely,

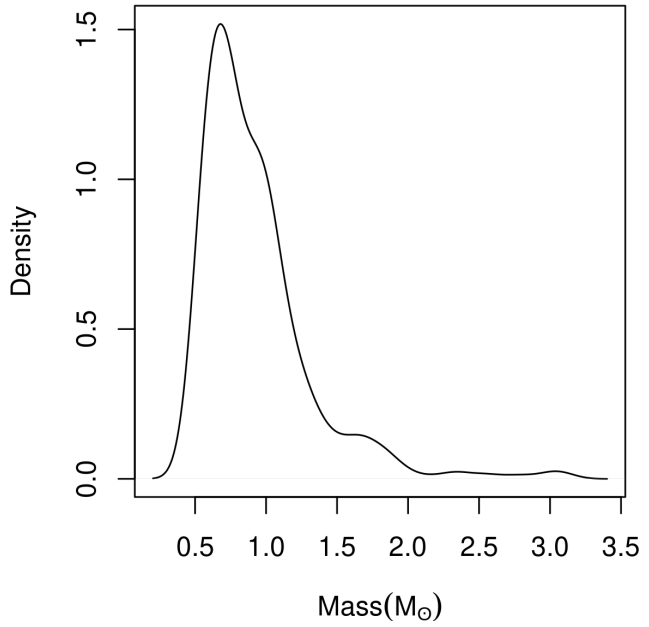

Fig. 13. Kernel density estimate of the mass distribution for primaries with good astrometric measurements ($\varpi/\sigma_\varpi > 15$).

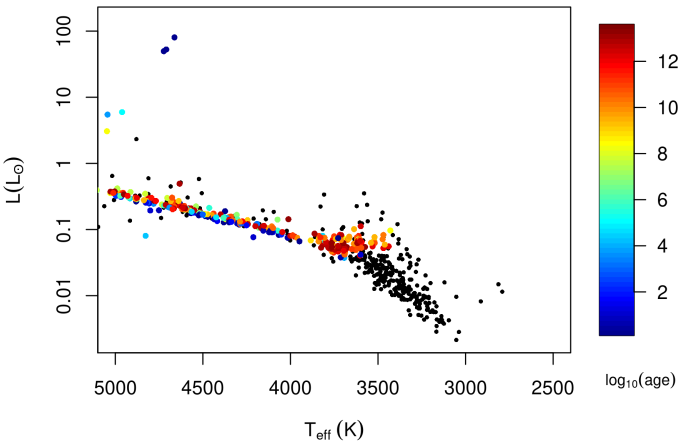

Fig. 14. Scatter plot of effective temperatures from the GSP-Phot module (x-axis) and luminosities from FLAME (y-axis) for primaries in systems where both components have $\varpi/\sigma_\varpi > 15$. Coloured circles correspond to sources with available age estimates from FLAME. The colour code represents the decadic logarithm of the age.

without implying a specific range of sizes or complexities in terms of members.

We identify the above mentioned overdensities by using a clustering technique applied to the set of sources in quality classes 0 and 1 of the *Gaia* UCD catalogue in the five-dimensional space of Galactic coordinates, tangential velocities and parallaxes. The clustering analysis in this five-dimensional space uses the hierarchical mode association clustering (HMAC) algorithm (Li et al. 2007). HMAC defines groups as sets of points associated with each mode (maxima) of the density distribution in the input space. It does not explicitly estimate the density, but it makes use of a kernel inside an iterative loop that associates sources with modes. Each kernel defines a set of modes, with narrower kernels resulting in many modes (that may include noise) and wider kernels reflecting only the larger structures. By using several kernels of increasing size we attain a hierarchical stratification of clustering groupings. The term cluster in this context will designate one of the groups of UCDs identified by the HMAC algorithm as associated with the same mode of the density in the five-dimensional space. The analysis of the substructures revealed by the different levels of the HMAC hierarchy is beyond the scope of this paper. It involves the analysis of the ages, kinematics and star formation histories of these regions as demonstrated by recent studies of some of the regions identified here (see for example Cantat-Gaudin et al. 2019; Kerr et al. 2021; Kounkel et al. 2022). In this section we do not take the subtleties due to the spatial and temporal complexities of the groups found by HMAC into account, and we give a broad picture of the differences as seen by *Gaia*.

We derive equatorial tangential velocities (in km s^{-1}) using the measured parallax and a conversion factor of 4.74. In a preprocessing stage, the data were centred at the median value of the distribution of each variable in the full set of sources in quality classes 0 and 1 and divided by the median absolute deviation (MAD) to avoid the clustering being dominated by the variables with a larger range of values. In our case, we defined isotropic kernels of sizes 0.02 (level 1), 0.04, 0.06, 0.08, and 0.1 (level 5). The bottom panel of Fig. 15 shows clusters with more than ten members identified at level 4 (left) and the distribution of sources not attached to any cluster (right). The latter shows hints of residual overdensities. Some of them are due to the scanning law (that results in more transits, better quality measurements and consequently more sources passing the quality criteria in certain regions of the sky) while two of the others (shown in red and blue) are discussed later below. Figure 16 shows the various clusters depicted in the lower-left panel of Fig. 15 in the space of tangential velocities (using the same colour code).

The significant overdensities identified by HMAC as distinct groups can easily be identified with well-known star-forming regions and clusters. These large groups show clear substructures discussed in the specialised literature. In Sect. 6.1 we discuss clusters identified using the BANYAN Σ software tool (Gagné et al. 2018) and in Sect. 6.2 we analyse HMAC clusters without members identified in any of the BANYAN groups.

But before discussing the group identification we next analyse the residual overdensities not identified as clusters by HMAC. While some are easily recognised as being due to the scanning law, there are two prominent regions that cannot be explained as being related to it. The positions of at least a significant fraction of the sources in these two overdensities on the celestial sphere and several CAMDs is consistent with membership to the Upper Scorpio, ρ Ophiucus, and Taurus star-forming regions, but their velocities are not concentrated and do not correspond to those typical of these regions. Figures C.16 and C.17 show the distribution of these sources in the space of tangential velocities and parallaxes. We find no concentration at the positions expected for these star-forming regions. Figure C.18 provides a potential explanation for these overdensities. It shows a M_G versus $(G - J)$ CAMD including the GCNS (black dots with transparency), the UCD catalogue discussed in this work (salmon dots with transparency), the GUCDS (violet dots), the sources in the two overdensities (red and blue dots) and the corresponding photometry corrected for extinction (green and orange dots) using the Planck generalized needlet internal linear combination (GNILC) map (Planck Collaboration XLVIII 2016). It shows a remarkable coincidence with the GCNS main sequence, indicating that the overdensities are mainly background sources of spectral type earlier than the UCD limit and whose RP spectra appear as those of UCDs due to the associated reddening.

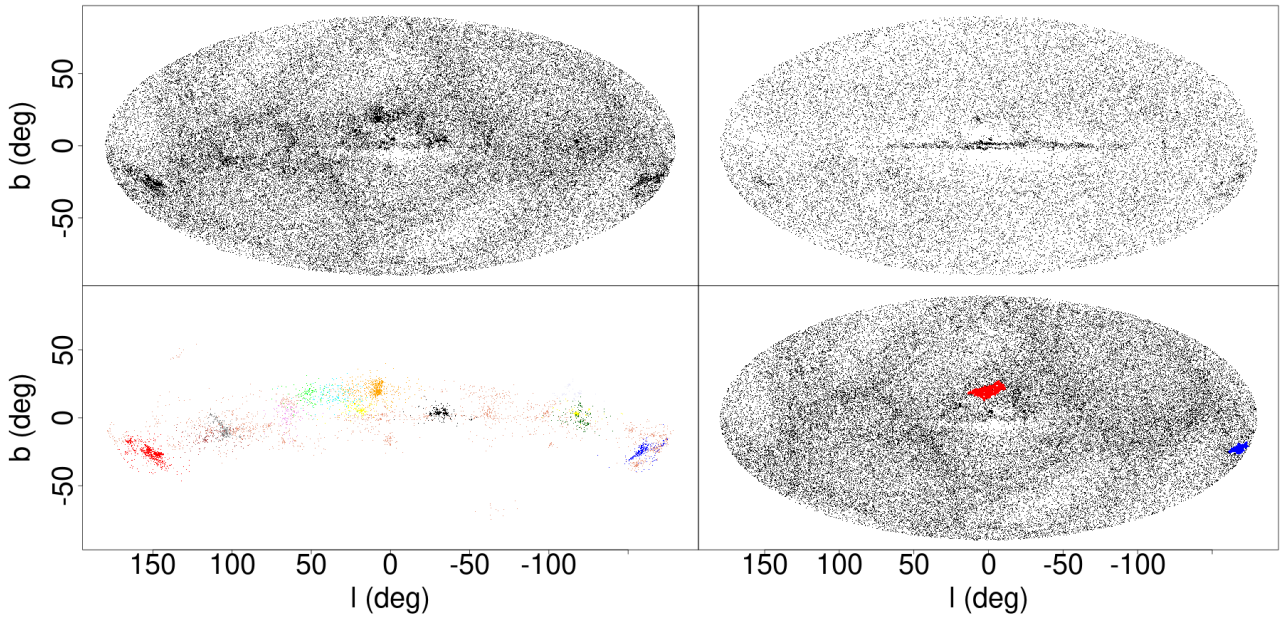

Fig. 15. Distribution in Galactic coordinates of the sources in quality classes 0 and 1 (*upper left*) and 2 (*upper right*) of the *Gaia* UCD catalogue using the Aitoff projection. *Lower-left panel*: clusters at level 4 of the hierarchy with more than ten members identified by HMAC in the set of sources in quality classes 0 and 1. There are 88 such clusters, the largest 13 of which are identified with prominent colours and the rest plotted in salmon for clarity. *Lower-right panel*: rest of the sources in clusters with ten members or fewer. We mark in red and blue the two most prominent overdensities.

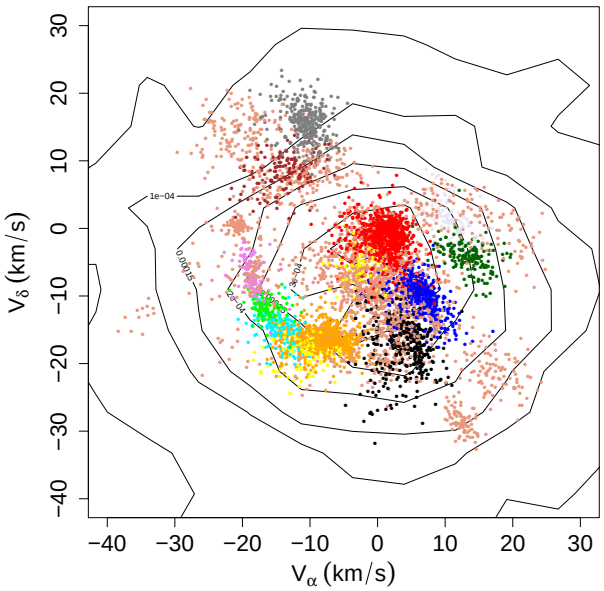

Fig. 16. Distribution in the space of tangential velocities of the sources in clusters with more than ten members. The colour code is the same as used in the lower-right panel of Fig. 15. The underlying contour lines depict the distribution of the sources in clusters with ten members or fewer as estimated using a kernel with $\sigma = 3$.

Table 4. First ten lines of the table containing the clusters assignments of the HMAC and BANYAN groups, including the BANYAN membership probability.

Source ID	HMAC cluster	BANYAN group	BANYAN probability
2572901021957789568	NA	CARN	0.98
1016186483391641216	NA	CARN	0.79
5556620540565785600	NA	ARG	0.95
1954170404122975232	NA	CARN	0.73
3411692668689199744	NA	ARG	0.96
5541111516746730752	NA	ABDMG	0.89
3597096309389074816	NA	CARN	0.99
6031367499416648192	NA	ARG	0.72
5908794218026022144	NA	ARG	1
6118581861234228352	NA	ABDMG	1

Notes. NA is used as a code to denote 'not available' in cases where one UCD was assigned to a group by one of the techniques but not by the other. The full table contains 7630 entries and is available at the CDS.

6.1. Membership according to BANYAN

In this section we discuss the determination of membership to star clusters and moving groups within 150 pc from the Sun as derived by the BANYAN Σ software tool (Gagné et al. 2018). We use the *Gaia* DR3 UCD candidates (their celestial positions, proper motions and parallaxes) as input to BANYAN Σ and find 2840 sources with membership probabilities higher than 0.5, 80% of them higher than 0.8. In the following we use the lower threshold of 0.5 despite indications that this may include contamination by sources from the field (Liu et al. 2016). BANYAN Σ is based on multivariate Gaussian modelling of the groups, which in some cases can be a simplification of the true distribution of sources in the space of measurements. HMAC on the contrary is a non-parametric clustering technique that can identify groups of arbitrary shapes and we prefer to be conservative in the definition of the BANYAN probability threshold for the sake of completeness. The cluster assignments and membership probabilities are included in Table 4. BANYAN Σ only produces membership probabilities to a set of predefined, well-known clusters of stars. As opposed to HMAC, it does not detect clusters or groups of sources with similar properties.

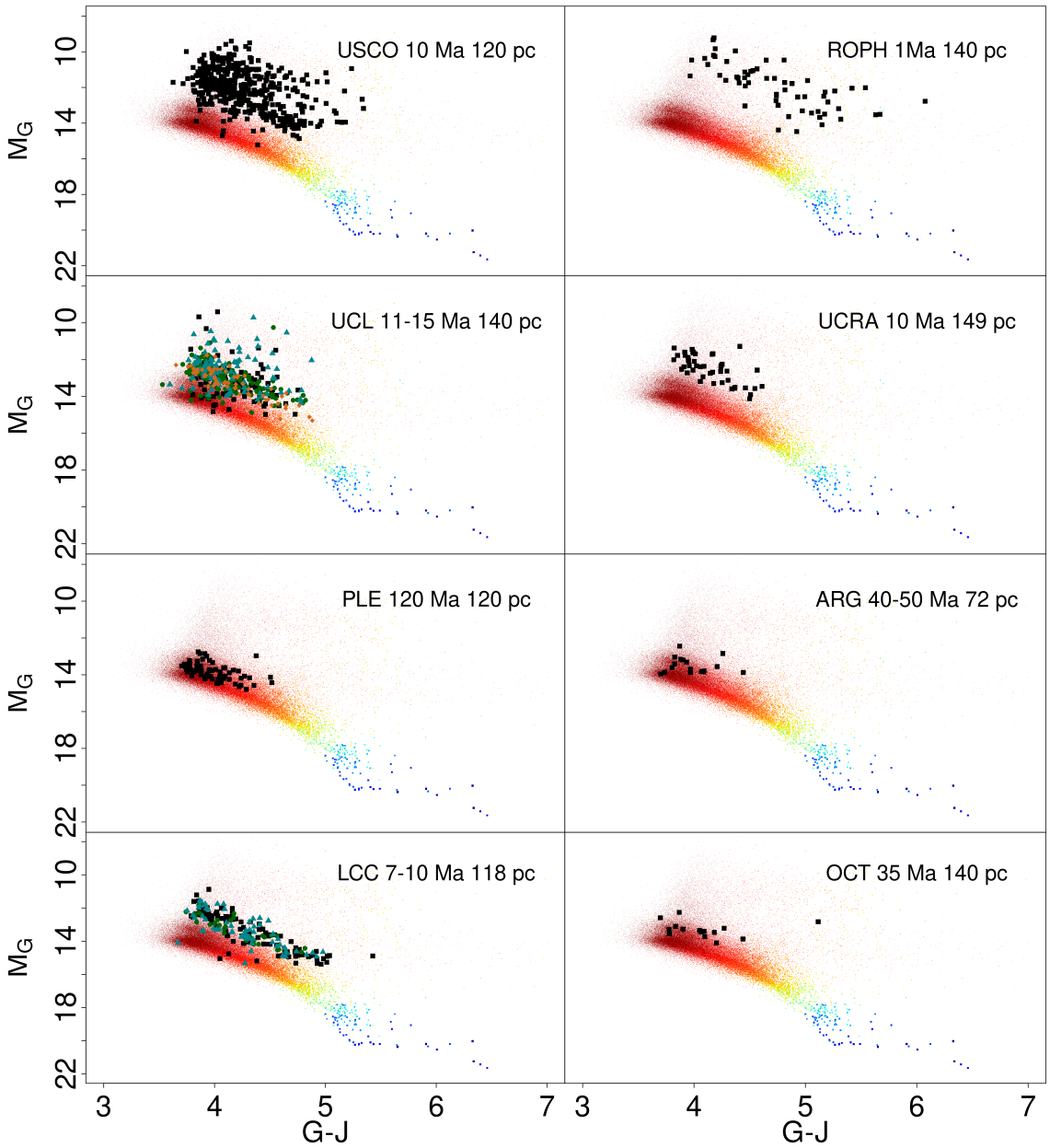

Fig. 17. Position in the CAMD of sources identified as members of one of the BANYAN associations or clusters and assigned to (up to three) HMAC clusters with more than 50 members. We use the same colour scale as in Fig. 2 for the full set of UCD candidates in the catalogue, and black filled squares, turquoise filled triangles, and brown filled diamonds are used to identify the HMAC groups over which BANYAN members can be spread if they correspond to more than one (see Table C.1).

Table C.1 shows the number of sources in common between the BANYAN Σ groups and HMAC clusters, and Fig. 17 shows the position of these sources in a CAMD for a few selected groups.

6.2. Stellar groups without members of the BANYAN clusters

Apart from the 19 HMAC clusters with members in common with one or several BANYAN groups, there are 38 additional HMAC clusters with more than 20 members with no common member with BANYAN groups. Figures C.1–C.14 show the celestial coordinates (leftmost plot), CAMD (mid-left), tangential velocities (mid-right), and distance (estimated by naive parallax inversion; rightmost) distributions of the 14 most numerous groups identified by the HMAC algorithm but without identifications in any of the BANYAN groups. Each figure represents one HMAC cluster at level 4 while the colours represent subclusters with more that five members identified at lower levels of the hierarchy (that is, for narrower kernels). The black circles denote sources in subclusters with five members or fewer. The substructures respond mainly to variations in the space of tangential velocities and not to the space of celestial coordinates where they sometimes mix without clear separations. The tentative identifications provided in the captions are only orientative and do not aim to reflect the spatial complexity of these stellar associations.

Figure C.15 depicts the remaining 24 HMAC clusters with more than 20 members in a similar way as that used in Figs. C.1–C.14 except for the rightmost panel that represents the kernel density estimates of the distance estimated by naive inversion of the parallax.

6.3. RP spectra as a function of age

Figure 18 shows the median RP spectrum in some of the clusters identified in the previous sections. Again, the distances and ages used are only included as a guide to order the spectra. The medians are calculated in bins of increasing effective temperature between 2350 and 2450 K (top left), between 2450 and 2550 K (top right), between 2550 and 2650 K (bottom left), and between 2650 and 2700 K (bottom right). When interpreting these figures, bear in mind that the regression module estimates temperatures using an empirical training set that does not include young or non-solar metallicity sources. Hence, the effective temperatures assigned to the sources in these stellar associations may be biased. Each of the figures includes the median calculated for sources outside the overdensities with a black line labelled 'main sequence' at the bottom of each figure, and with a light grey line superimposed on each association to facilitate the comparison. We interpret these sources as representing evolved examples from the zero-age main sequence and so, well represented in the training set. It is impossible to separate the contribution of the various absorption lines and bands at the low resolution of the RP spectra. Reiners et al. (2007) provide us with high-resolution spectra and line/band identifications in that wavelength range and for spectral types slightly cooler that those corresponding to the temperatures represented in Fig. 18. From them, we hypothesise that the two absorption features visible in these RP spectra are mainly due to TiO (band heads at 758.9, 766.6, 843.2, and 885.9 nm), VO (785.1 and 852.1 nm), CrH (861.1 nm), and the alkali spectral lines of K I (766.5 and 7698 nm), Rb I (780.0 and 794.7 nm), and Cs I (852.1 nm). In all four figures (but more prominently, in the top-right panel) we see that the absorption bands get deeper as the association age becomes older. As mentioned above, the HMAC cluster identifications involves complex groups with several substructures and a range of ages. Orion for example contains star-forming regions with ages estimated between 1 and 10 Ma. In these cases, the median RP spectrum plotted in the figures represents a weighted average of slightly different spectra.

These figures seem to indicate that by 10 Ma the RP spectrum can hardly be distinguished from that of the evolved ones in the main sequence and therefore, low gravity detection can only be accomplished for very young associations. Plans for DR4 include the parameterisation of these changes in the RP spectrum and the inclusion of youth indication flags in the archive for these young UCD candidates.

7. Variability of UCDs

Gaia DR3 includes the results of the processing and analysis of the time series of individual sources. We searched the associated archive variability tables for entries in common with the UCD catalogue discussed here. Figure 19 shows the median G magnitude of the 1109 sources in common as a filled circle and its range of values in the time series (segments) as a function of difference between the mean and the median of the time series. The colour code reflects the $T_{\rm eff}$ value inferred by the ESP-UCD module. We have highlighted sources at the two extremes of the distribution of this difference (mean-median) with thicker lines. It shows that the main bulk of UCD candidates with entries in the vari_summary table concentrates around the origin of the x axis implying that the time series does not show outliers. But it also shows UCD candidates with very asymmetric distributions. On the left hand side of the plot we encounter sources with bright outliers and on the right hand side, sources with faint outliers. Figure 20 shows the position of the UCD candidates that have entries in the vari_summary table (orange) in the *Gaia* CAMD, superimposed on a kernel-based estimate of the density of UCD candidates. Red squares represent the variable UCDs with faint outliers and the blue asterisks, UCDs with bright outliers. The line segments represent the displacement between the values of the G and $G_{\rm RP}$ magnitudes in the main *Gaia* catalogue (used in all sections. of this paper) and the median values from the time series. The CAMD already shows that most of the variable UCD candidates are placed in the region of young, pre-main-sequence sources. In fact, 728 of the 1109 UCD candidates with variability data are identified by HMAC as belonging to overdensities, and 353 by BANYAN Σ as members of the BANYAN clusters. The most numerous HMAC groups are Upper Sco (250 sources), Orion (162), Serpens (53), and the Perseus association (48). Those of the BANYAN class set are Upper Sco (203 sources), Upper Centaurus Lupus (57), and Taurus (45).

We initially interpreted the sources with bright outliers as potential sources with flares, and those with faint outliers as eclipsing binary candidates. However, the UCD candidates with outliers significantly fainter that the median (the examples to the right, which we had interpreted initially as potential eclipsing binaries) are all classified in the vari_classifier_result as young stellar objects (YSOs) except for one source that is classified as long period variable, albeit with a probability of 0.3. This is in agreement with their position in the CAMD.

The sources to the left of Fig. 19 were initially assumed to be candidates for flaring UCDs. We searched the identifiers in the variability tables in *Gaia* DR3 and in SIMBAD. The latter returned seven cross-matches with object type identifications: two high proper motion sources, one low-mass object, two rotating variables, and two YSOs. The former returned five classifications as YSOs (with probabilities greater than 0.85 except in one case); three classifications in the RR Lyrae class (with probabilities below 0.15); and two long period variables (probabilities of 0.06).

Variability of UCDs has been studied in the past (see e.g. Wilson et al. 2014; Apai et al. 2017; Biller 2017; Artigau 2018, and references therein). Some variability might be associated with binarity, but intrinsic atmospheric changes should be present too. We note, however, that young, low mass stars present other phenomena, such as the presence of dips in their light curve possibly due to stellar disc occultations (Stauffer et al. 2015). This explanation cannot be ruled out since most of our variable sources are associated with stellar associations. However, the irregular sampling of the *Gaia* time series and their unavailability in DR3 precludes a deeper analysis in the context of this work.

8. Summary

In this work we present a first overview of the UCD content in the *Gaia* DR3 archive. More specifically: we present the typical (median) RP spectra in spectral type bins and compare them to UCD standards and ground-based high-resolution spectra; we present CAMDs that include *Gaia* and external photometry; we compare the catalogue with previous lists of UCD candidates based on *Gaia* data; we construct a simple hierarchical probabilistic model as a first step towards inferring the spatial and luminosity distribution of UCDs when a deeper knowledge of the catalogue selection function becomes available; we provide a list of candidate companions in wide binary systems, including UCDs from the catalogue; we analyse the overdensities of UCDs in the celestial sphere and identify them (at a coarse grain level)

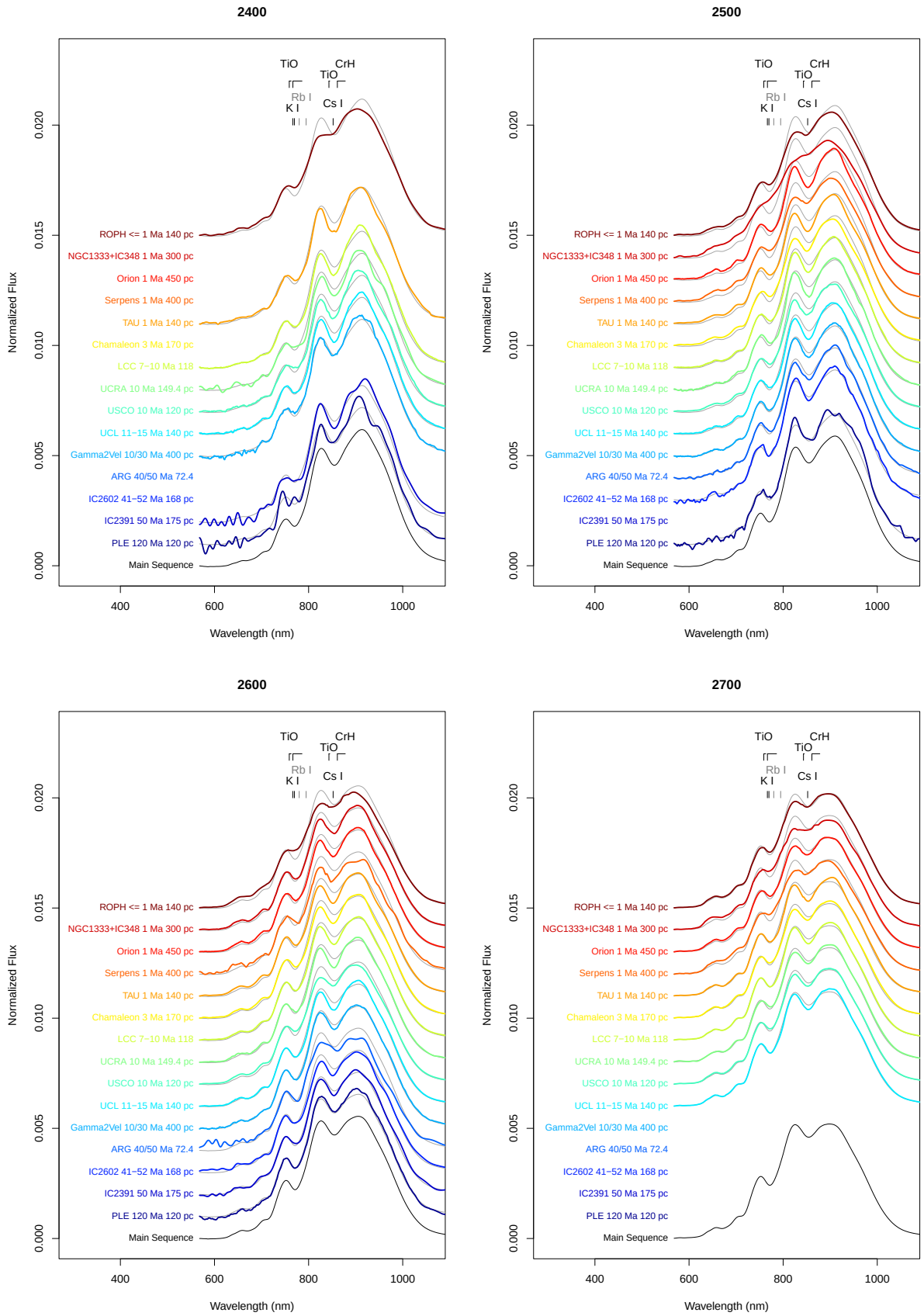

Fig. 18. Median RP spectra calculated for sources in the range between 2350 and 2450 K (*top left*), 2450 and 2550 K (*top right*), 2550 and 2650 K (*bottom left*), and 2650 and 2700 K (*bottom right*) and assigned to the most prominent BANYAN and HMAC groups. The spectra are ordered in age from the top (youngest) to the bottom (oldest). The median RP spectrum for sources outside the overdensities in each temperature bin is labelled 'main sequence' and is shown in black at the bottom and in light grey superimposed on each group.

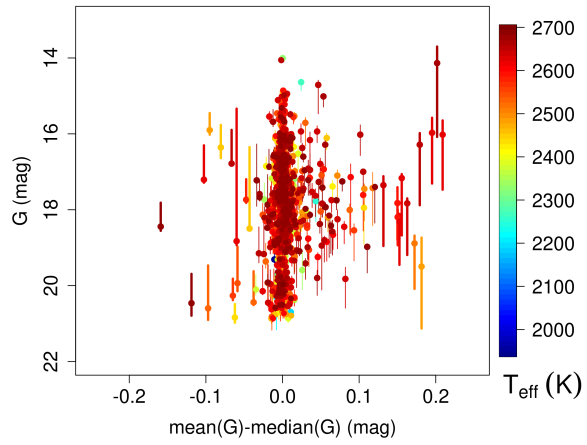

Fig. 19. Ranges of G values in the time series of UCD candidates included in the variability tables of *Gaia* DR3. The filled circle in each segment marks the time series median value.

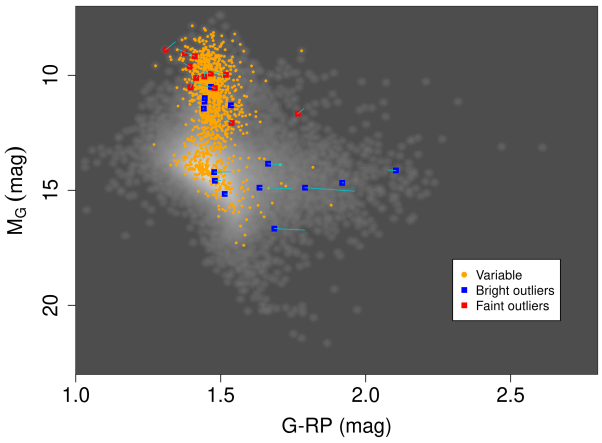

Fig. 20. Kernel density estimate of the distribution of UCD candidates in the *Gaia* CAMD (grey scale). Candidates with entries in the vari_summary table are marked as orange circles; blue and red squares mark the extreme cases of bright and faint outliers shown in Fig. 19, respectively. The segments illustrate the displacement in the CAMD between the main archive values of the photometry and the medians available in the vari_summary table.

with known star-forming regions and stellar associations; and, finally, we briefly review the variability properties of the UCD candidates. These identifications will serve as a basis for the inclusion of youth indicators in future versions of the catalogue.

This global overview will help the community further explore some of the aspects of the catalogue highlighted here and inspire subsequent analyses with complementary observations.

Acknowledgements. This work has made use of results from the European Space Agency (ESA) space mission *Gaia*, the data from which were processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). The *Gaia* mission website is http://www.cosmos.esa.int/gaia. This work was supported by the MCIN (Spanish Ministry of Science and Innovation) through grant PID2020-112949GB-I00; the MINECO (Spanish Ministry of Economy) through grants AyA2017-84089, ESP2016-80079-C2-1-R, ESP2014-55996-C2-1-R, and RTI2018-095076-B-C22 (MINECO/FEDER, UE). This research has been funded by the Spanish State Research Agency (AEI) Projects No.PID2019-107061GB-C61 and No. MDM-2017-0737 Unidad de Excelencia "María de Maeztu"- Centro de Astrobiología (CSIC/INTA). Ground based spectra is from observations made with the Gran Telescopio Canarias (GTC), installed in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias, on the island of La Palma. The GTC data was obtained with the instrument OSIRIS, built by a Consortium led by the Instituto de Astrofísica de Canarias in collaboration with the Instituto de Astronomía of the Universidad Autónoma de México. OSIRIS was funded by GRANTECAN and the National Plan of Astronomy and Astrophysics of the Spanish Government. The programme codes were GTC54-15A0 & GTC8-15ITP. This work has made use of the Python package GaiaXPy, developed and maintained by members of the *Gaia* Data Processing and Analysis Consortium (DPAC), and in particular, Coordination Unit 5 (CU5), and the Data Processing Centre located at the Institute of Astronomy, Cambridge, UK (DPCI). We derived extiction corrections using the Python package dustmap (Green 2018).

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, ArXiv e-prints [arXiv:1603.04467]
Almendros-Abad, V., Mužić, K., Moitinho, A., Krone-Martins, A., & Kubiak, K. 2022, A&A, 657, A129
Apai, D., Karalidi, T., Marley, M. S., et al. 2017, Science, 357, 683
Artigau, É. 2018, in Handbook of Exoplanets, eds. H. J. Deeg, & J. A. Belmonte (Berlin: Springer), 94
Bailer-Jones, C. A. L. 2015, PASP, 127, 994
Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, A&A, 577, A42
Bardalez Gagliuffi, D. C., Burgasser, A. J., Gelino, C. R., et al. 2014, ApJ, 794, 143
Bardalez Gagliuffi, D. C., Burgasser, A. J., Schmidt, S. J., et al. 2019, ApJ, 883, 205
Bate, M. R. 2012, MNRAS, 419, 3115
Bennett, M., & Bovy, J. 2019, MNRAS, 482, 1417
Biller, B. 2017, Astron. Rev., 13, 1
Bonnell, I. A., Clark, P., & Bate, M. R. 2008, MNRAS, 389, 1556
Bovy, J. 2015, ApJS, 216, 29
Burgasser, A. J., Wilson, J. C., Kirkpatrick, J. D., et al. 2000, AJ, 120, 1100
Burgasser, A. J., Kirkpatrick, J. D., Brown, M. E., et al. 2002, ApJ, 564, 421
Burgasser, A. J., Kirkpatrick, J. D., Liebert, J., & Burrows, A. 2003a, ApJ, 594, 510
Burgasser, A. J., Kirkpatrick, J. D., McElwain, M. W., et al. 2003b, AJ, 125, 850
Burgasser, A. J., Cruz, K. L., Cushing, M., et al. 2010, ApJ, 710, 1142
Burningham, B., Pinfield, D. J., Leggett, S. K., et al. 2008, MNRAS, 391, 320
Cantat-Gaudin, T., Jordi, C., Wright, N. J., et al. 2019, A&A, 626, A17
Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86
Castro, P. J., & Gizis, J. E. 2012, ApJ, 746, 3
Castro, P. J., Gizis, J. E., Harris, H. C., et al. 2013, ApJ, 776, 126
Chabrier, G., Johansen, A., Janson, M., & Rafikov, R. 2014, in Protostars and Planets VI, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning (Tucson: University of Arizona Press), 619
Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, A&A, in press, https://doi.org/10.1051/0004-6361/202243688
Cruz, K. L., Reid, I. N., Liebert, J., Kirkpatrick, J. D., & Lowrance, P. J. 2003, AJ, 126, 2421
Cruz, K. L., Reid, I. N., Kirkpatrick, J. D., et al. 2007, AJ, 133, 439
De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, A&A, in press, https://doi.org/10.1051/0004-6361/202243680
Deacon, N. R., & Hambly, N. C. 2007, A&A, 468, 163
Delchambre, L., Bailer-Jones, C. A. L., Bellas-Velidis, I., et al. 2023, A&A, in press, https://doi.org/10.1051/0004-6361/202243423
Delfosse, X., Tinney, C. G., Forveille, T., et al. 1997, A&A, 327, L25
Dillon, J. V., Langmore, I., Tran, D., et al. 2017, ArXiv e-prints [arXiv:1711.10604]
Fan, X., Knapp, G. R., Strauss, M. A., et al. 2000, AJ, 119, 928
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2022, A&A, in press, https://doi.org/10.1051/0004-6361/202243919
Gagné, J., Faherty, J. K., Cruz, K. L., et al. 2015, ApJS, 219, 33
Gagné, J., Mamajek, E. E., Malo, L., et al. 2018, ApJ, 856, 23
Gaia Collaboration (Brown, A. G. A., et al.) 2016a, A&A, 595, A2
Gaia Collaboration (Prusti, T., et al.) 2016b, A&A, 595, A1
Gaia Collaboration (Babusiaux, C., et al.) 2018a, A&A, 616, A10
Gaia Collaboration (Brown, A. G. A., et al.) 2018b, A&A, 616, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2021a, A&A, 649, A1
Gaia Collaboration (Smart, R. L., et al.) 2021b, A&A, 649, A6
Gaia Collaboration, (Creevey, O. L., et al.) 2023, A&A, in press https://doi.org/10.1051/0004-6361/202243800
Geballe, T. R., Knapp, G. R., Leggett, S. K., et al. 2002, ApJ, 564, 466
Gizis, J. E. 2002, ApJ, 575, 484
Gizis, J. E., Monet, D. G., Reid, I. N., et al. 2000, AJ, 120, 1085
GRAVITY Collaboration (Abuter, R., et al.) 2022, A&A, 657, L12
Green, G. 2018, J. Open Source Softw., 3, 695
Hall, P. B. 2002, ApJ, 580, L77
Hawley, S. L., Covey, K. R., Knapp, G. R., et al. 2002, AJ, 123, 3409
Hennebelle, P., & Chabrier, G. 2008, ApJ, 684, 395

Higdon, D. 2002, in Quantitative Methods for Current Environmental Issues, eds. C. W. Anderson, V. Barnett, P. C. Chatwin, & A. H. El-Shaarawi (London: Springer London), 37
Hoffman, M. D., & Gelman, A. 2014, J. Mach. Learn. Res., 15, 1593
Jao, W.-C., Henry, T. J., Gies, D. R., & Hambly, N. C. 2018, ApJ, 861, L11
Kendall, T. R., Delfosse, X., Martín, E. L., & Forveille, T. 2004, A&A, 416, L17
Kendall, T. R., Jones, H. R. A., Pinfield, D. J., et al. 2007, MNRAS, 374, 445
Kerr, R. M. P., Rizzuto, A. C., Kraus, A. L., & Offner, S. S. R. 2021, ApJ, 917, 23
Kirkpatrick, J. D., Henry, T. J., & McCarthy, Donald W., J. 1991, ApJS, 77, 417
Kirkpatrick, J. D., Reid, I. N., Liebert, J., et al. 1999, ApJ, 519, 802
Kirkpatrick, J. D., Reid, I. N., Liebert, J., et al. 2000, AJ, 120, 447
Kirkpatrick, J. D., Martin, E. C., Smart, R. L., et al. 2019, ApJS, 240, 19
Kirkpatrick, J. D., Gelino, C. R., Faherty, J. K., et al. 2021, ApJS, 253, 7
Kounkel, M., Stassun, K. G., Covey, K., & Hartmann, L. 2022, MNRAS, 517, 161
Lépine, S., & Shara, M. M. 2005, AJ, 129, 1483
Lépine, S., Shara, M. M., & Rich, R. M. 2002, AJ, 124, 1190
Li, J., Ray, S., & Lindsay, B. 2007, J. Mach. Learn. Res., 8, 1687
Liu, M. C., Dupuy, T. J., & Allers, K. N. 2016, ApJ, 833, 96
Looper, D. L., Kirkpatrick, J. D., Cutri, R. M., et al. 2008, ApJ, 686, 528
Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, A&A, 616, A9
Luyten, W. J. 1955, Luyten's Five Tenths. (1955, 0)
Luyten, W. J. 1979, LHS catalogue. A catalogue of stars with proper motions exceeding 0″5 annually (Minneapolis: University of Minnesota), 2nd edn.
Mainzer, A., Bauer, J., Grav, T., et al. 2011, ApJ, 731, 53
Marocco, F., Andrei, A. H., Smart, R. L., et al. 2013, AJ, 146, 161
Martín, E. L., Basri, G., Delfosse, X., & Forveille, T. 1997, A&A, 327, L29
Martín, E. L., Delfosse, X., Basri, G., et al. 1999, AJ, 118, 2466
Montegriffo, P., De Angeli, F., Andrae, R., et al. 2023, A&A, in press, https://doi.org/10.1051/0004-6361/202243880
Padoan, P., & Nordlund, Å. 2004, ApJ, 617, 559
Planck Collaboration XLVIII. 2016, A&A, 596, A109
Rasmussen, C., & Williams, C. 2006, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning (Cambridge, MA, USA: MIT Press), 248
Reid, I. N., & Gizis, J. E. 2005, PASP, 117, 676
Reid, I. N., Kirkpatrick, J. D., Gizis, J. E., et al. 2000, AJ, 119, 369
Reid, I. N., Lewitus, E., Allen, P. R., Cruz, K. L., & Burgasser, A. J. 2006, AJ, 132, 891
Reid, I. N., Cruz, K. L., Kirkpatrick, J. D., et al. 2008, AJ, 136, 1290
Reiners, A., Homeier, D., Hauschildt, P. H., & Allard, F. 2007, A&A, 473, 245
Reipurth, B., & Clarke, C. 2001, AJ, 122, 432
Reylé, C. 2018, A&A, 619, L8
Riaz, R., Vanaverbeke, S., & Schleicher, D. R. G. 2018, MNRAS, 478, 5460
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Rix, H.-W., Hogg, D. W., Boubert, D., et al. 2021, AJ, 162, 142
Salim, S., Lépine, S., Rich, R. M., & Shara, M. M. 2003, ApJ, 586, L149
Sarro, L. M., Berihuete, A., Carrión, C., et al. 2013, A&A, 550, A44
Saumon, D., Geballe, T. R., Leggett, S. K., et al. 2000, ApJ, 541, 374
Schmidt, S. J., Cruz, K. L., Bongiorno, B. J., Liebert, J., & Reid, I. N. 2007, AJ, 133, 2258
Schneider, A. C., Cushing, M. C., Kirkpatrick, J. D., et al. 2014, AJ, 147, 34
Scholz, R. D., & Meusinger, H. 2002, MNRAS, 336, L49
Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smart, R. L., Marocco, F., Caballero, J. A., et al. 2017, MNRAS, 469, 401
Stamer, T., & Inutsuka, S.-i. 2019, MNRAS, 488, 2644
Stauffer, J., Cody, A. M., McGinnis, P., et al. 2015, AJ, 149, 130
Stephens, D. C., Leggett, S. K., Cushing, M. C., et al. 2009, ApJ, 702, 154
Tinney, C. G., & Reid, I. N. 1998, MNRAS, 301, 1031
Veras, D., & Raymond, S. N. 2012, MNRAS, 421, L117
Whitworth, A. P., & Stamatellos, D. 2006, A&A, 458, 817
Whitworth, A. P., & Zinnecker, H. 2004, A&A, 427, 299
Wilson, P. A., Rajan, A., & Patience, J. 2014, A&A, 566, A111
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zahnle, K. J., & Marley, M. S. 2014, ApJ, 797, 41

[1] Dpto. de Inteligencia Artificial, UNED, c/ Juan del Rosal 16, 28040 Madrid, Spain
e-mail: lsb@dia.uned.es

[2] INAF – Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese (TO), Italy
[3] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[4] INAF – Osservatorio astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy
[5] CNES Centre Spatial de Toulouse, 18 avenue Édouard Belin, 31401 Toulouse Cedex 9, France
[6] Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
[7] Observational Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden
[8] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[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] Institut d'Astrophysique et de Géophysique, Université de Liège, 19c, Allée du Six-Août, 4000 Liège, Belgium
[12] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[13] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, 40129 Bologna, Italy
[14] Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[15] 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
[16] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[17] Centre for Astrophysics Research, University of Hertfordshire, College Lane, AL10 9AB, Hatfield, UK
[18] 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
[19] Data Science and Big Data Lab, Pablo de Olavide University, 41013 Seville, Spain
[20] Department of Astrophysics, Astronomy and Mechanics, National and Kapodistrian University of Athens, Panepistimiopolis, Zografos, 15783 Athens, Greece
[21] Dipartimento di Fisica e Astronomia Ettore Majorana, Università di Catania, Via S. Sofia 64, 95123 Catania, Italy
[22] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 Place Jules Janssen, 92190 Meudon, France
[23] Université Rennes, CNRS, IPR (Institut de Physique de Rennes) - UMR 6251, 35000 Rennes, France
[24] IPAC, Mail Code 100-22, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[25] Applied Physics Department, Universidade de Vigo, 36310 Vigo, Spain
[26] Departamento de Astrofísica, Centro de Astrobiología (CSIC-INTA), ESA-ESAC. Camino Bajo del Castillo s/n. 28692 Villanueva de la Cañada, Madrid, Spain
[27] Institut UTINAM CNRS UMR6213, Université Bourgogne Franche-Comté, OSU THETA Franche-Comté Bourgogne, Observatoire de Besançon, BP1615, 25010 Besançon Cedex, France
[28] Depto. Estadística e Investigación Operativa, Universidad de Cádiz, Avda. República Saharaui s/n, 11510 Puerto Real, Cádiz, Spain

Appendix A: Alternative colour-absolute magnitude diagrams

This Sect. includes two additional CAMDs to provide a complementary view of Fig. 2. In Fig. A.1 we show in the x axis alternative colour indices constructed from the G band and the 2MASS H and K bands (instead of J; left and middle plots), and the WISE W_2 band (instead of W_1; right plot). In Fig. A.2 we show colour indices combining 2MASS and WISE bands but not the *Gaia* G band. In the left and middle panels, the well-known elbow towards bluer colour indices is visible. This blue turn for the latest spectral types is due to the appearance of methane absorption bands in the H and K bands at the L/T spectral type transition and the silicate clouds transition from above to below the photosphere resulting in a J band brightening. We interpret the larger scatter of the $J-H$ and $H-K$ colour indices as due to their comparatively short baseline and the relative faintness of the UCDs. Furthermore, these colours are in the peak of energy distribution for the hotter UCDs and the behaviour of water vapour absorption across the H-band can produce additional scatter (see for example Almendros-Abad et al. 2022), in contrast to the relative insensitivity of bands such as $W1$ and $W2$, which are on the Rayleigh-Jeans tail of the energy distribution.

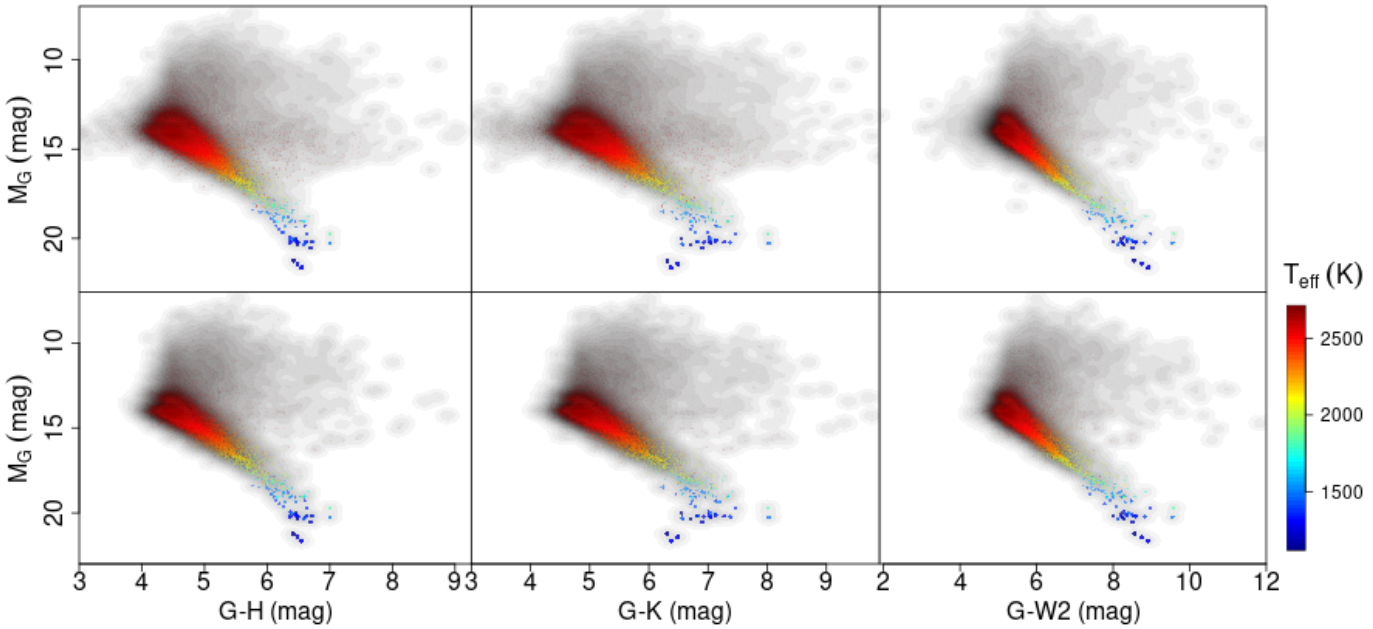

Fig. A.1: Same as Fig. 2 but for additional colour indices, including the *Gaia* G band magnitude and 2MASS H, K or AllWISE W_2. A kernel density estimate is shown using a grey scale.

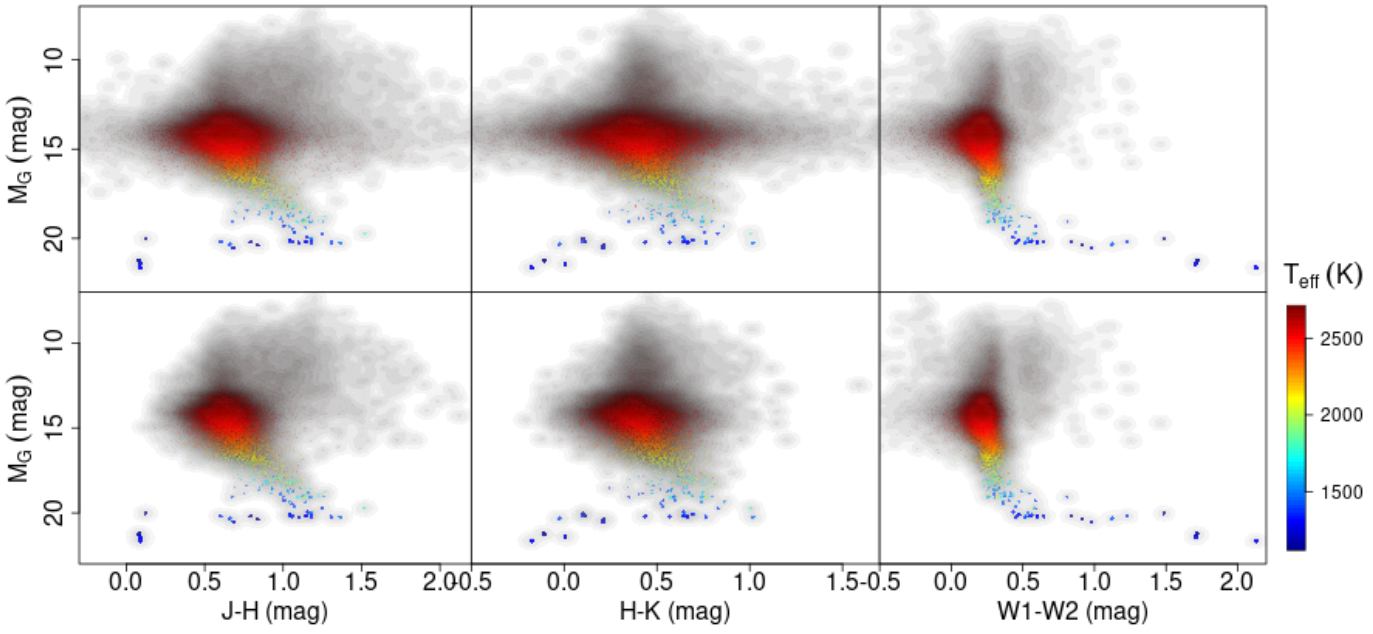

Fig. A.2: Same as Fig. 2 but for additional colour indices, excluding the *Gaia* G band magnitude. A kernel density estimate is shown using a grey scale.

Appendix B: Details of the probabilistic model

Inference with the probabilistic model described in Sect. 4 was carried out using Tensorflow libraries (Abadi et al. 2015; Dillon et al. 2017). These libraries adopt the single-program multiple-data (SPMD) paradigm. The SPMD paradigm allows abstractions for scaling the code to configurations such as tensor processing unit pods or clusters of graphics processing units that allowed us to compute the likelihood in a distributed computing framework. This was required due to the complexity of the hierarchical Bayesian model and the large number of parameters and observations.

Posterior distributions have been characterised by drawing samples using the no U-turn sampler (NUTS) algorithm, which is an adaptive variant of the Hamiltonian Monte Carlo method for MCMC (Hoffman & Gelman 2014).

Appendix B.1. The modified PERT distribution

Several of the prior probabilities for parameters of the model described in Sect. 4 were defined as a modified PERT distribution. It is defined by four parameters, usually referred to as low, high, peak and temperature. It has a compact support in an interval between the low and high values and a most probable value at the peak, which usually indicates the expert's most frequent prediction. The temperature parameter controls the sharpness of the peak. Specifically, a modified PERT distribution has the shape

$$\text{PERT}(\text{low}, \text{peak}, \text{high}, \text{temperature}) \equiv \mu + \sigma \cdot \text{Beta}(\alpha, \beta),$$

where

$$\begin{aligned} \mu &= \text{low}, \\ \sigma &= \text{high} - \text{low}, \\ \alpha &= 1 + \text{temperature} \cdot \frac{\text{peak} - \text{low}}{\text{high} - \text{low}}, \\ \beta &= 1 + \text{temperature} \cdot \frac{\text{high} - \text{peak}}{\text{high} - \text{low}}, \end{aligned} \tag{B.1}$$

with temperature > 0.

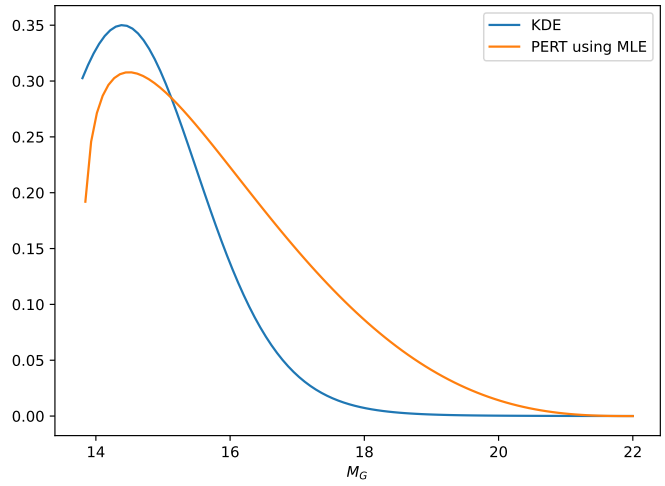

Fig. B.1: Modified PERT distribution for the M_G prior and the kernel density estimate of the observed M_G distributions derived using the naive inversion of the parallax.

Appendix B.2. Approximate Gaussian process

In this paper we use the approximation proposed by Higdon (2002), for constructing a Gaussian process $f(x)$ over a general region $x \in \mathcal{X}$ by convolving a continuous white noise process $\beta(x)$ with a smoothing kernel $k(x)$ so that

$$f(x) = \int_{\mathcal{X}} k(u - x)\beta(u)\,\mathrm{d}u, \text{ for } x \in \mathcal{X}. \tag{B.2}$$

In Higdon (2002) this integral is approximated for a basic spatial model. In brief, let $y_1, y_2, \ldots y_n$ be data recorded over spatial locations $x_1, x_2, \ldots x_n$ in $\mathcal{X}$, and consider a simple spatial model such as

$$y = \mu + z + \epsilon, \tag{B.3}$$

where the elements of $z = (z_1, z_2, \ldots z_n)^T$ are the restriction of the Gaussian process $f(x)$ to the data locations $x_1, x_2, \ldots, x_n$. In this case, the Gaussian process $f(x)$ has zero mean and, instead of being defined by its covariance function, it is determined by the latent process $\beta(x)$ and the smoothing kernel $k(x)$.

The latent process $\beta(x)$ is restricted to be non-zero at spatial locations $\omega_1, \omega_2, \ldots \omega_m$, and in $\mathcal{X}$, and we consider $\beta_j = \beta(\omega_j)$, $j = 1, \ldots, m$. Each β_j is modelled as an independent draw from a $\mathcal{N}(0, \sigma_x)$ distribution. Given Eq. B.2 the continuous Gaussian process is approximated as

$$f(x) = \sum_{j=1}^{m} \beta_j k(x - \omega_j), \tag{B.4}$$

where $k(\cdot - \omega_j)$ is a kernel centred at ω_j. In this paper we choose $k(\cdot)$ to be a Gaussian density, which implies a smooth radially symmetric kernel. Therefore, Eq. B.3 results in the linear model

$$y = \mu \mathbf{1}_n + K\boldsymbol{\beta} + \epsilon, \tag{B.5}$$

where $\mathbf{1}_n$ is the n-vector of 1's, and the elements of K are given by

$$\begin{aligned} K_{ij} &= k(x_i - \omega_j) \\ \boldsymbol{\beta} &\sim \mathcal{N}(0, \sigma_x I_m) \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \sigma_\epsilon I_n). \end{aligned}$$

From a statistical point of view this model is a basic mixed effects model that can be included as part of our Bayesian model.

Appendix C: Stellar association complementary material

In this Sect. we include additional material with details about the cross-match between HMAC and BANYAN groups (Sect. C.1) and several diagrams (sky positions, CAMD, tangential velocities, and histograms of the inverse of the parallax) showing the properties of the HMAC clusters without members in common with the BANYAN groups in Sect. C.2.

Appendix C.1. Intersection between HMAC and BANYAN clusters

Table C.1 shows the number of sources in common between BANYAN and HMAC groups. The leftmost column includes numerical identifiers of the HMAC clusters.

Appendix C.2. HMAC clusters without members in any of the BANYAN groups

Figures C.1–C.15 show the distribution in equatorial coordinates (left), the CAMD (middle left), the tangential velocities (middle right), and the histogram of the inverse of the parallax (right) for several HMAC clusters without members in any BANYAN groups. Tentative identifications are included in the top of the left plot and in the captions.

Appendix C.3. Residual overdensities

Figures C.16–C.18 include additional diagnostic plots for the interpretation of the residual overdensities shown in Fig. 15.

Table C.1: Number of sources in common between the BANYAN groups (top row) and the clusters identifiers obtained using HMAC (leftmost column).

	ARG	CRA	EPSC	IC2391	IC2602	LCC	OCT	PL8	PLE	ROPH	TAU	UCL	UCRA	USCO
2	0	0	0	0	0	0	0	0	0	59	0	117	0	469
7	0	0	0	0	0	2	0	0	0	0	0	139	0	2
8	0	0	0	0	0	0	0	0	0	0	0	103	0	0
9	0	0	0	0	0	0	17	0	0	0	0	0	0	0
10	0	0	5	0	0	106	0	0	0	0	0	1	0	0
11	0	0	0	0	0	33	0	0	0	0	0	68	0	0
15	0	0	0	0	0	0	0	0	70	0	0	0	0	0
18	0	0	0	0	0	64	0	0	0	0	0	1	0	0
19	0	7	0	0	0	0	0	0	0	0	0	0	44	0
20	17	0	0	19	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	52	0	0	0
22	0	0	0	0	0	1	0	16	0	0	0	0	0	0
24	0	0	0	0	10	0	0	2	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	17	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	12	0	0
46	0	0	0	0	2	0	0	0	0	0	0	0	0	0
47	4	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	5	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	4	0	0

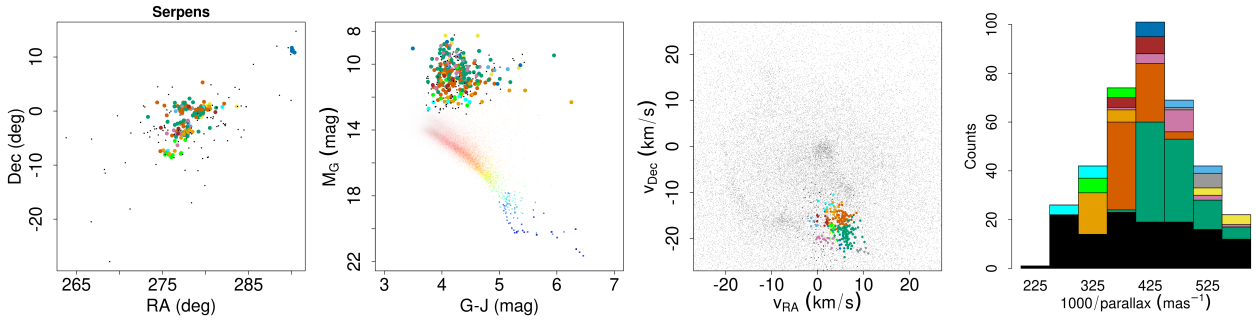

Fig. C.1: Properties of one (level 4) HMAC cluster tentatively identified with Serpens (but see text) with no sources in common with BANYAN associations. From left to right: Equatorial coordinates with various colours separating subclusters at a hierarchy level lower (and hence finer) than 4; CAMD (points represent the full set of UCD candidates in the catalogue, and circles identify the HMAC group at level 4; black dots represent sources not assigned to any subcluster in the lower level); tangential velocities; and the stacked histogram of the inverse of the parallax in all subclusters.

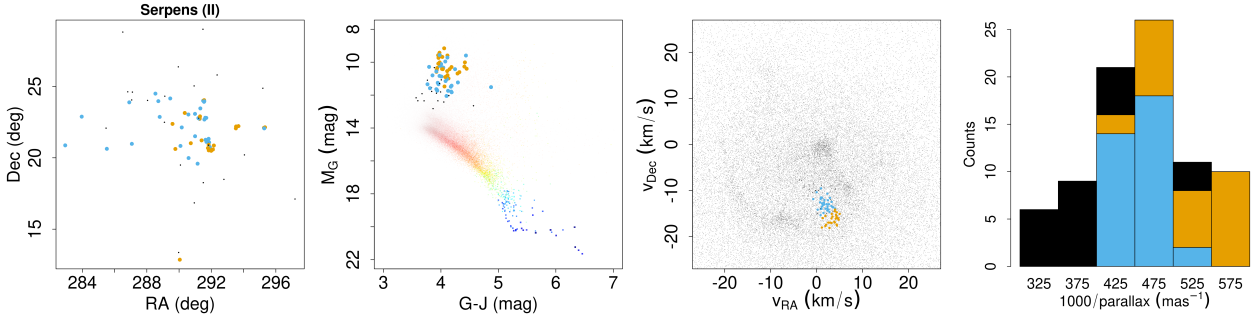

Fig. C.2: Same as Fig. C.1 but for Serpens (II)

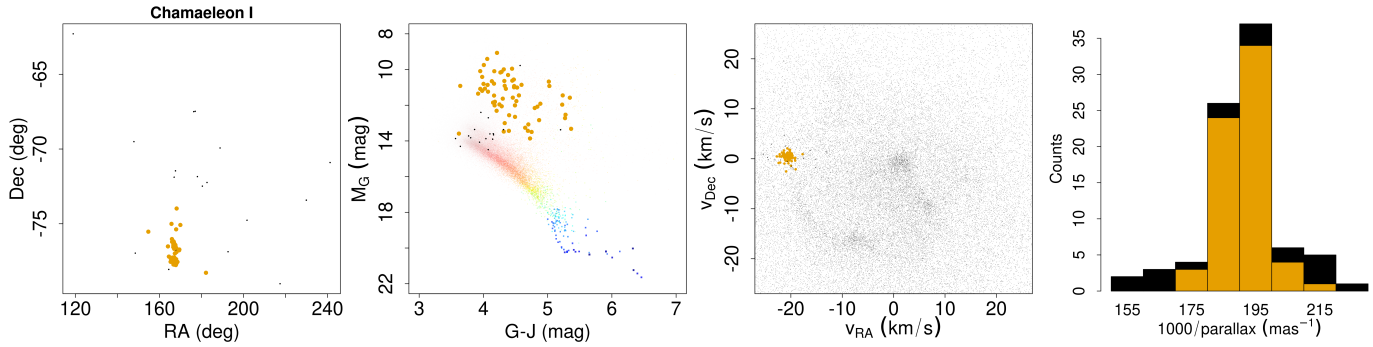

Fig. C.3: Same as Fig. C.1 but for Chamaeleon I.

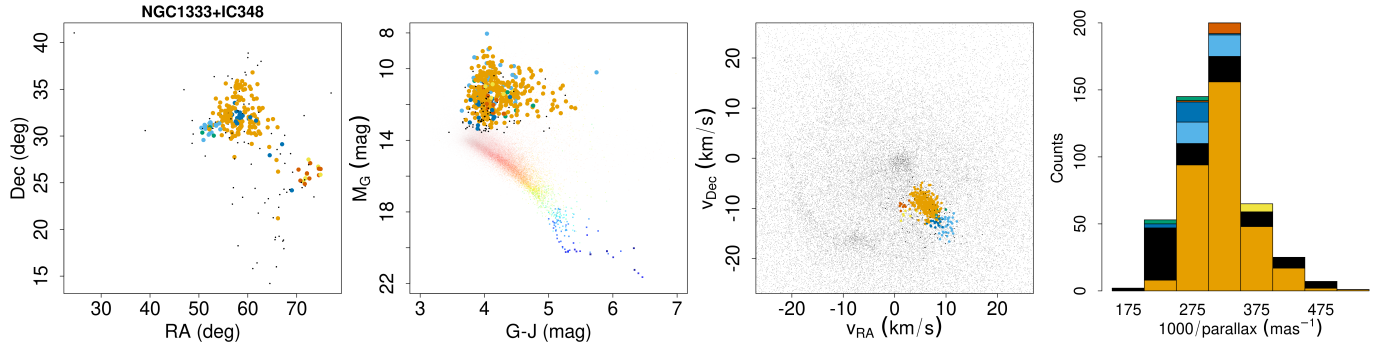

Fig. C.4: Same as Fig. C.1 but for the Perseus region.

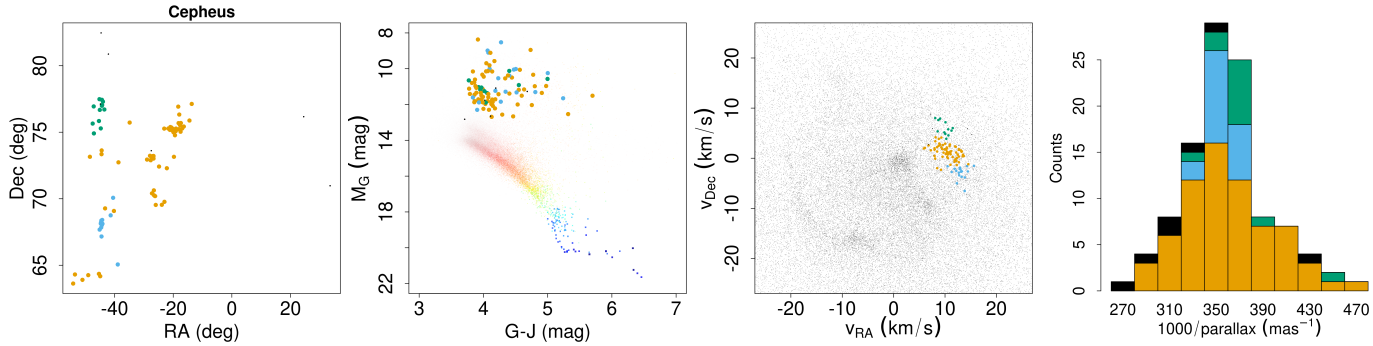

Fig. C.5: Same as Fig. C.1 but for the Cepheus region.

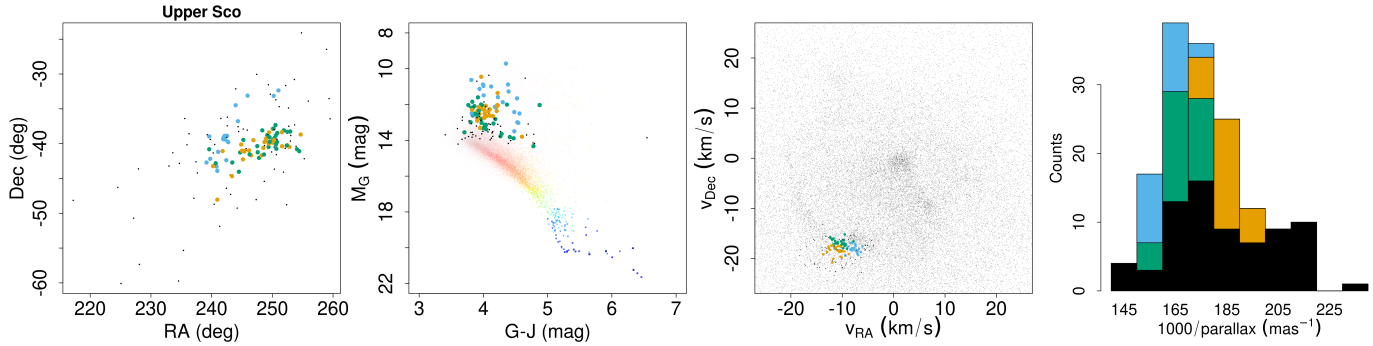

Fig. C.6: Same as Fig. C.1 but for Upper Scorpio.

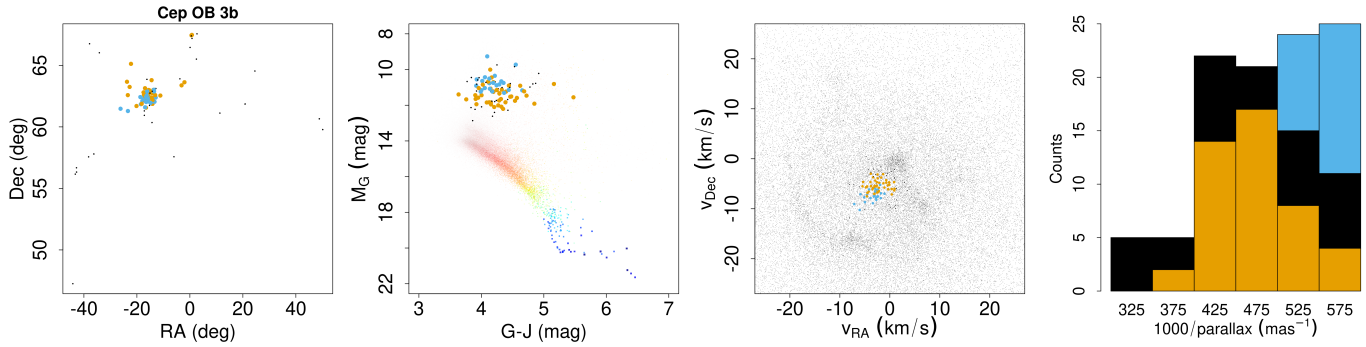

Fig. C.7: Same as Fig. C.1 but for Cep OB 3b.

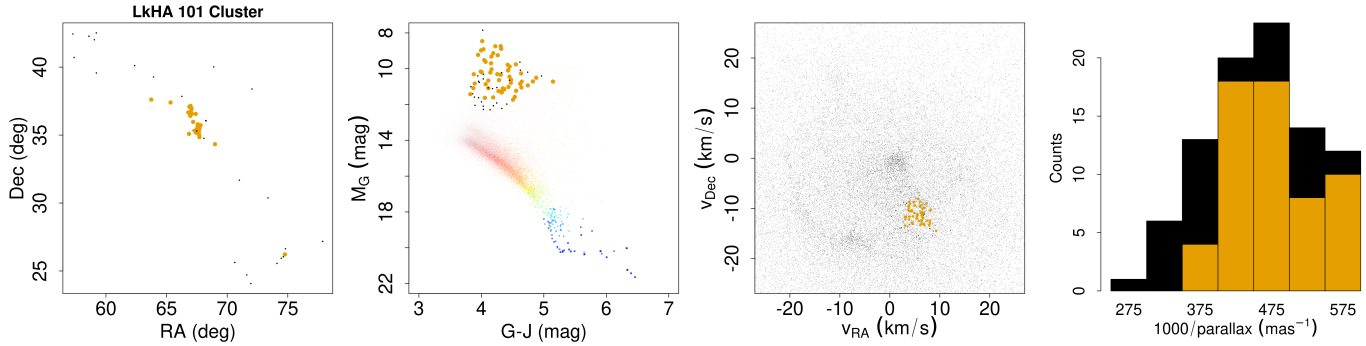

Fig. C.8: Same as Fig. C.1 but for the LkHA 101 Cluster.

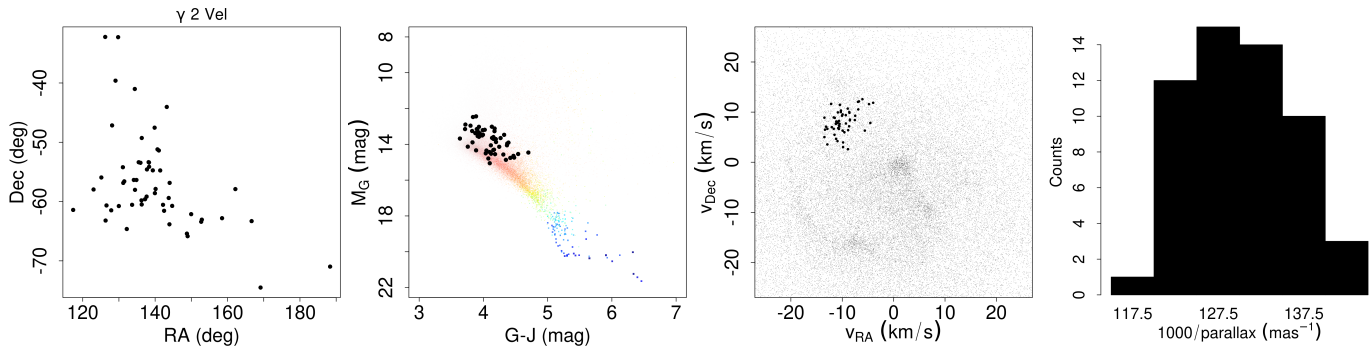

Fig. C.9: Same as Fig. C.1 but for γ2 Vel (I).

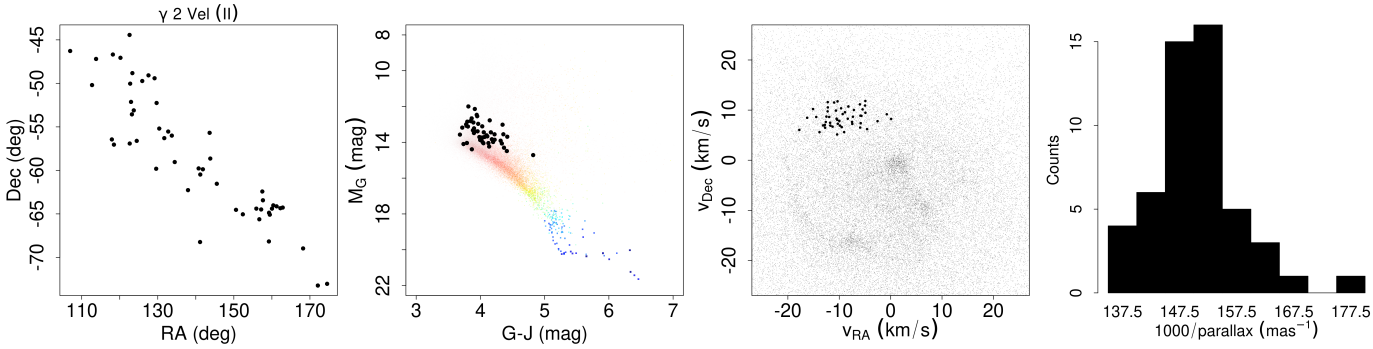

Fig. C.10: Same as Fig. C.1 but for γ2 Vel (II).

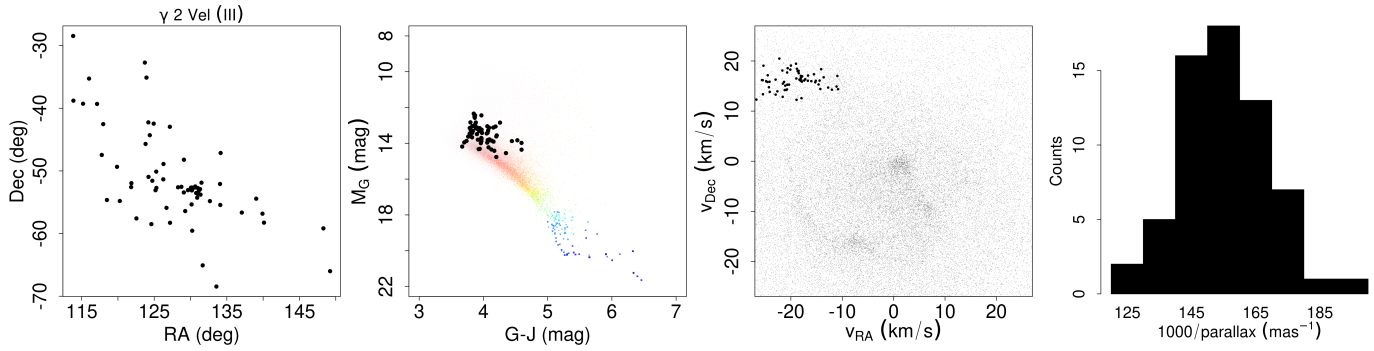

Fig. C.11: Same as Fig. C.1 but for γ2 Vel (III).

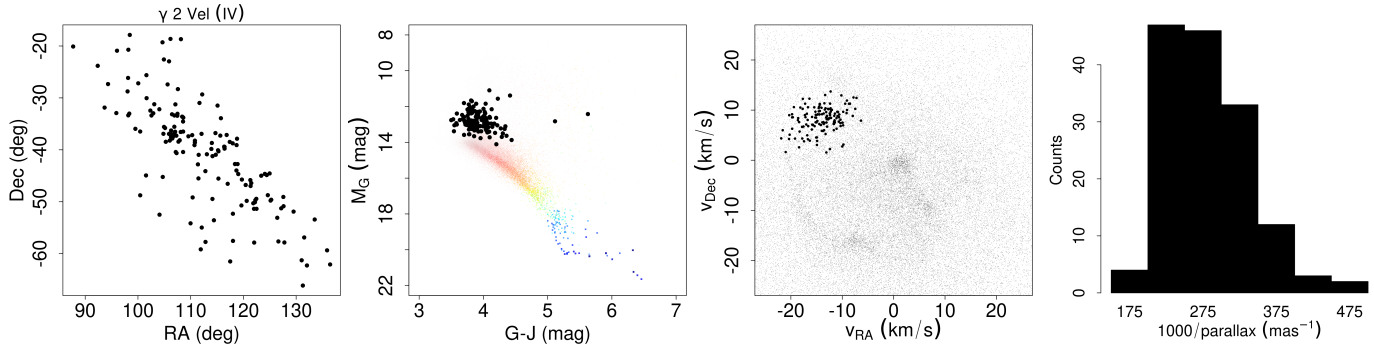

Fig. C.12: Same as Fig. C.1 but for γ2 Vel (IV).

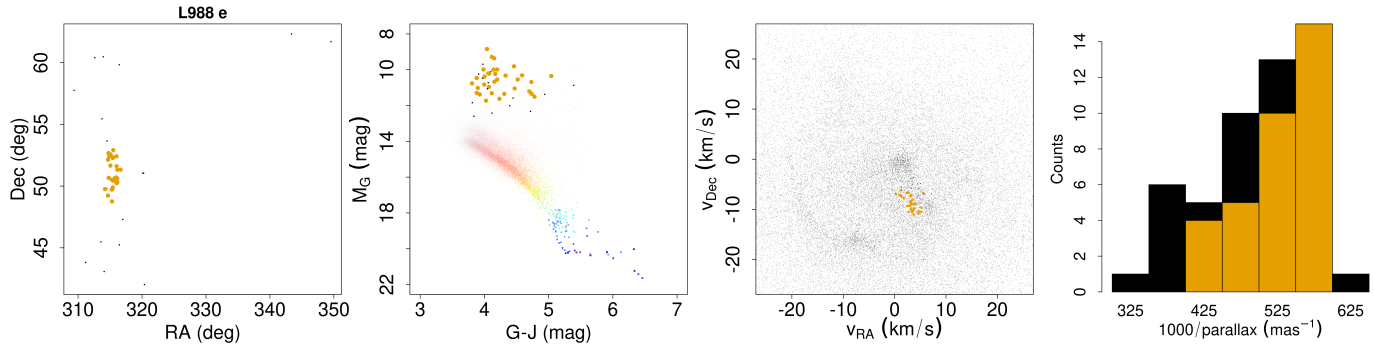

Fig. C.13: Same as Fig. C.1 but for L988 e.

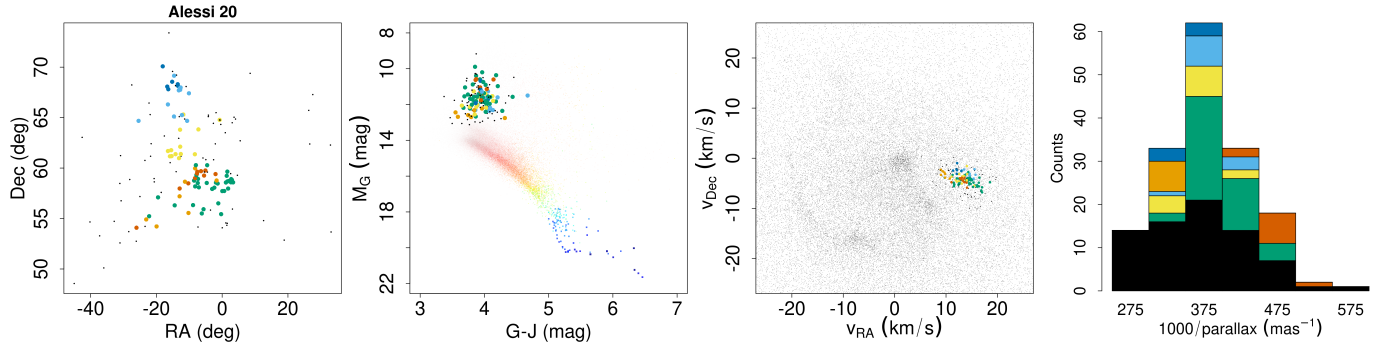

Fig. C.14: Same as Fig. C.1 but for the region near Alessi 20.

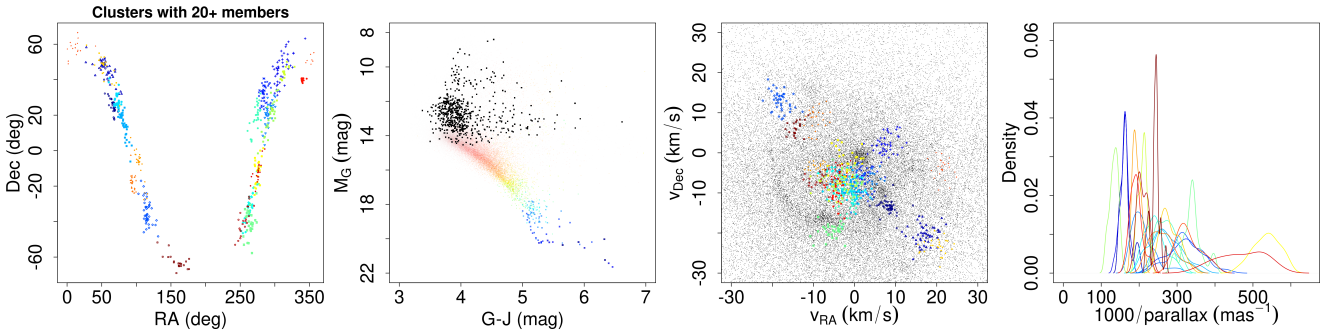

Fig. C.15: Same as Fig. C.1 but for the rest of clusters with more than 20 members. We use different colours for the various clusters, except for the CAMD, to avoid confusion with the colour code that reflects $T_{\rm eff}$ along the main sequence.

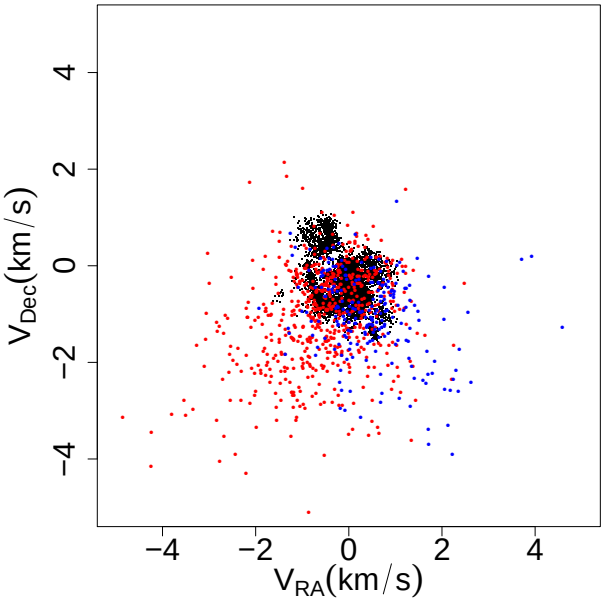

Fig. C.16: Distribution in the space of tangential velocities of the sources in residual overdensities shown in the lower-right panel of Fig. 15 using the same colour code.

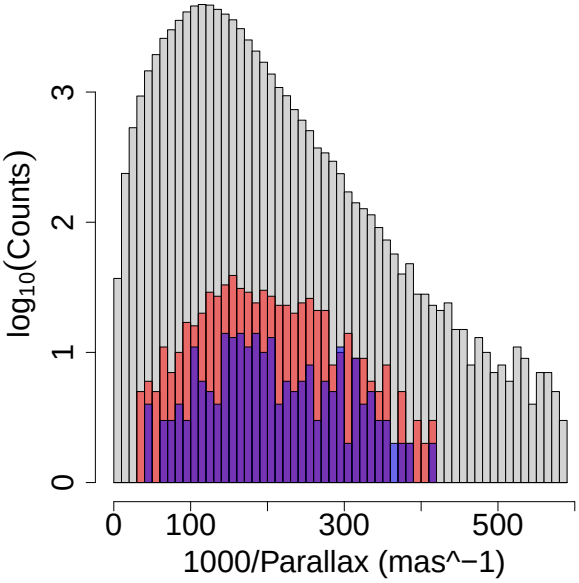

Fig. C.17: Histogram of the inverse of the parallax of candidate UCDs (white) and of sources in the residual overdensities marked in red and blue in the lower-right panel of Fig. 15 using the same colour code.

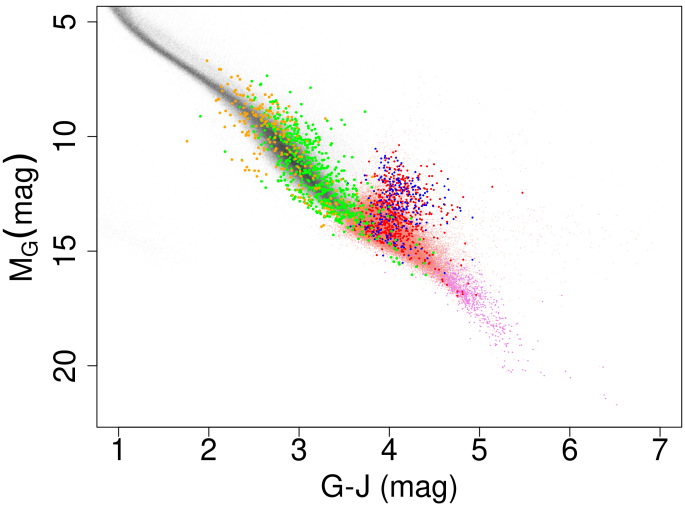

Fig. C.18: Distribution of sources in the residual overdensities marked in red and blue in the lower-right panel of Fig. 15 in the CAMD diagram of $(G-J)$ and M_G. It shows the main sequence derived from the GCNS as black dots (using transparency); the list of sources in our catalogue of UCDs in HMAC groups of fewer than ten members (salmon dots with transparency); the GUCDS (violet dots); and the position of the sources in the blue and red overdensities after de-reddening using the *Planck* GNILC (Planck Collaboration XLVIII 2016) dust map as orange and green dots, respectively.