Publication Year	2022
Acceptance in OA	2025-02-25T11:22:21Z
Title	Low-mass young stars in the Milky Way unveiled by DBSCAN and Gaia EDR3: Mapping the star forming regions within 1.5 kpc
Authors	PRISINZANO, Loredana, DAMIANI, Francesco, SCIORTINO, Salvatore, FLACCOMIO, Ettore, GUARCELLO, Mario Giuseppe, MICELA, Giuseppina, Tognelli, E., Jeffries, R. D., ALCALA', JUAN MANUEL
Publisher's version (DOI)	10.1051/0004-6361/202243580
Handle	http://hdl.handle.net/20.500.12386/36194
Journal	ASTRONOMY & ASTROPHYSICS
Volume	664

Mathieu, R. D. 1994, ARA&A, 32, 465

Mayne, N. J., & Naylor, T. 2008, MNRAS, 386, 261

McInnes, L., Healy, J., & Astels, S. 2017, JOSS, 2, 205

Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, AJ, 144, 192

Mitra-Kraev, U., Harra, L. K., Güdel, M., et al. 2005, A&A, 431, 679

Miville-Deschênes, M.-A., & Lagache, G. 2005, ApJS, 157, 302

Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, ApJ, 834, 57

Moitinho, A., Alves, J., Huélamo, N., & Lada, C. J. 2001, ApJ, 563, L73

Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, L100

Montalto, M., Piotto, G., Marrese, P. M., et al. 2021, A&A, 653, A98

Otrupcek, R. E., Hartley, M., & Wang, J. S. 2000, PASA, 17, 92

Palla, F., & Stahler, S. W. 1999, ApJ, 525, 772

Palla, F., Randich, S., Flaccomio, E., & Pallavicini, R. 2005, ApJ, 626, L49

Peacock, J. A. 1983, MNRAS, 202, 615

Pecaut, M. J., & Mamajek, E. E. 2013, ApJS, 208, 9

Piecka, M., & Paunzen, E. 2021, BAJ, accepted [arXiv:2107.07230]

Poggio, E., Drimmel, R., Cantat-Gaudin, T., et al. 2021, A&A, 651, A104

Prisinzano, L., Damiani, F., Micela, G., & Sciortino, S. 2005, A&A, 430, 941

Randich, S., Gilmore, G., & Gaia-ESO Consortium. 2013, The Messenger, 154, 47

Randich, S., Tognelli, E., Jackson, R., et al. 2018, A&A, 612, A99

Rebull, L. M., Johnson, C. H., Gibbs, J. C., et al. 2013, AJ, 145, 15

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, ApJ, 885, 131

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

Salpeter, E. E. 1955, ApJ, 121, 161

Scalo, J. 1998, in ASP Conf. Ser. 142: The Stellar Initial Mass Function (38th Herstmonceux Conference), 201

Sitnik, T. G. 2003, Astron. Lett., 29, 311

Soderblom, D. R., Hillenbrand, L. A., Jeffries, R. D., Mamajek, E. E., & Naylor, T. 2014, Protostars and Planets VI, 219

Somers, G., Cao, L., & Pinsonneault, M. H. 2020, ApJ, 891, 29

Spina, L., Randich, S., Magrini, L., et al. 2017, A&A, 601, A70

Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287

Tognelli, E., Prada Moroni, P. G., & Degl'Innocenti, S. 2018, MNRAS, 476, 27

Tognelli, E., Prada Moroni, P. G., Degl'Innocenti, S., Salaris, M., & Cassisi, S. 2020, A&A, 638, A81

Venuti, L., Damiani, F., & Prisinzano, L. 2019, A&A, 621, A14

Yonekura, Y., Dobashi, K., Mizuno, A., Ogawa, H., & Fukui, Y. 1997, ApJS, 110, 21

Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, A&A, 620, A172

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, Res. Astron. Astrophys., 12, 723

Zucker, C., Speagle, J. S., Schlafly, E. F., et al. 2020, A&A, 633, A51

Appendix A: Interstellar reddening effects

In this section, we show the effects of the reddening on the sample selected as described in Section 3. As discussed in Anders et al. (2019), for a generic pass band, *i*, the extinction coefficients A_i/A_V depend on the stellar effective temperature. The subsequent dust-attenuated photometry of very broad photometric passbands, such as the *Gaia* EDR3 ones, is not simply a linear function of A_V. It is also a function of the source spectrum that is its effective temperature.

The PARSEC 1.2S stellar models (Bressan et al. 2012; Chen et al. 2014; Tang et al. 2014) have been implemented to predict tracks and isochrones also at non-zero extinction. As done in Montalto et al. (2021), in order to have an indication of the reddening that affects our data, we used the CMD 3.3 input form web interface, and we constructed a grid of stellar models assuming the 1 Gyr isochrone and A_V=[0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 20.0, 30.0].

Figure A.1 shows how the 1 Gyr isochrone changes by increasing extinction, $A_{V,}$ from 0 to 10, in the CAMD obtained by adopting the different *Gaia* colours (panels a and b). The 1 Gyr isochrone at $A_V = 0$ is highlighted by a thick black line, while the 1 Gyr isochrone at $A_V = 3$ is highlighted by symbols with different shades of pink in the different stellar evolution phases.

We note that the reddened isochrone is not linearly shifted along a reddening direction, which usually happens when adopting a reddening vector. For example, for an object at $M_G = 3$, corresponding to a star with $(G_{BP}\text{-}G_{RP})_0$=0.47, $(G\text{-}G_{RP})_0$=0.30, an effective temperature of 6930 K at 1 Gyr (black empty square in the Figure), and an extinction of $A_V = 3$, the reddening $E(G_{BP}\text{-}G_{RP})$ is equal to 1.24 and $E(G\text{-}G_{RP})$ is equal to 0.55 (blue arrows in the Figure). However, for an object at $M_G = 8$, corresponding to a star with $(G_{BP}\text{-}G_{RP})_0$=1.81, $(G\text{-}G_{RP})_0$=0.90, effective temperature of 3945 K at 1 Gyr (black bullet in the Figure), and an extinction of $A_V = 3$, the reddening $E(G_{BP}\text{-}G_{RP})$ is equal to 1.09 and $E(G\text{-}G_{RP})$ is equal to 0.26 (red arrows in the Figure). Thus, at different temperatures, and for a fixed A_V, the shift in colour due to the reddening is smaller for the colder star. This effect is higher in the G versus $G\text{-}G_{RP}$ diagram, as can be deduced from the different slopes of the blue and red arrows. In this case, for a ~4000 K star, the $E(G\text{-}G_{RP})$ value (equal to ≃0.26) is about half of that (≃0.55) associated with a ~7000 K star. This implies that while a reddened 1-Gyr-old star with an effective temperature of ~ 7000 K can be expected to be found in the PMS region and mimic a star younger than 10 Myr, a colder star of ~4000 K, of the same age, and affected by the same extinction, does not fall in the PMS region (see Fig. A.1, panel b). In conclusion, the effect of uncorrected reddening in terms of contamination of our initial photometric sample by old stars is larger for stars of spectral type F and G than for K and M stars.

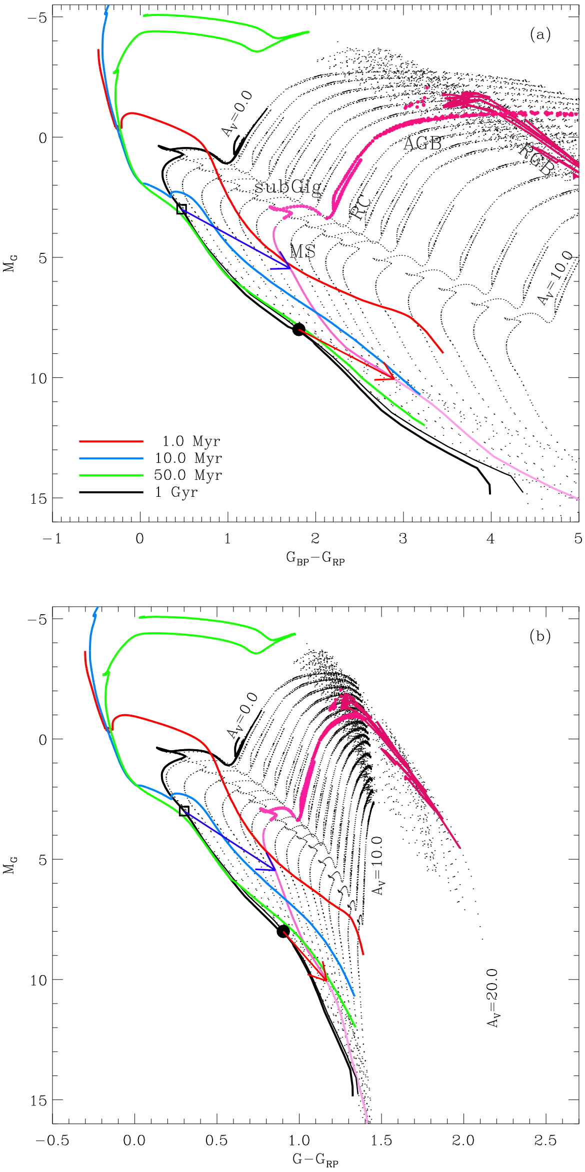

Fig. A.1. PARSEC 1 Gyr isochrone at solar metallicity with extinction, $A_{V,}$ ranging from 0 to 10.0 in the CAMD obtained by adopting the different *Gaia* magnitudes (black dots). The 1 Gyr isochrone at $A_V = 0$ is highlighted by a thick black solid line. The 1 Gyr isochrone at $A_V = 3$ is highlighted by pink coloured lines of different shades during the red giant branch (RGB), asymptotic giant branch (AGB), red clump (RC), sub-giant (subGig), and MS phases. The red, light-blue, and green solid lines are the 1, 10, and 50 Myr Pisa isochrones at solar metallicity. The empty square and the bullet in each panel represent a star of 6930 K and 3945 K, respectively, while the blue and red arrows are the reddening vectors corresponding to $A_V = 3$, for these two representative stars (see text).

Appendix B: Effect of binarity or multiplicity on astrometric selections

At the level of astrometric sensitivity offered by Gaia, the orbital motions of binary (or multiple) stars sometimes become measurable, and also difficult to disentangle from proper motion. This holds both for resolved pairs and for unresolved, unequal-mass pairs where the photocentre displays significant motion (see Kervella et al. 2022). If the binary period resonates with the Gaia sampling frequency, parallax measurements will also be affected. Perhaps the best-studied star-forming region in terms of its binary-star population is Taurus-Auriga, and we refer the reader to the review by Mathieu (1994) for a perspective of the expected range of system parameters. Taurus is one of the few SFRs where lunar occultation techniques were feasible for the detection of close pairs, down to separations of 0.009″(Mathieu 1994, Table A1 and references therein). Therefore, the projected binary separations span a factor of ~ 1000, with no ‘typical’ value. Correspondingly, their orbital periods span a range of a factor of ~ 30 000.

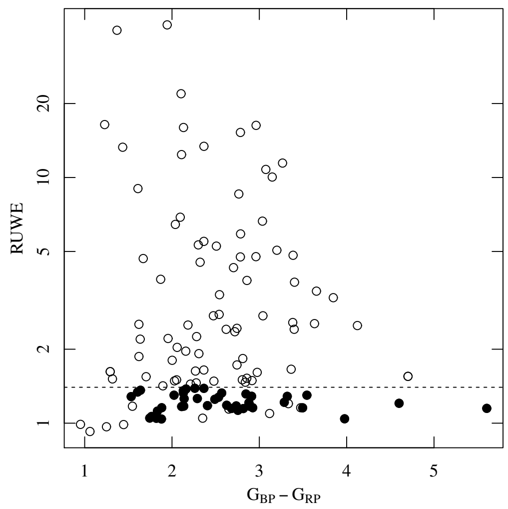

Fig. B.1. Diagram of RUWE values versus the Gaia colour $G_{BP}-G_{RP}$ of the Kraus et al. (2012) Taurus-Auriga binary-star list with *Gaia* EDR3 counterparts. Filled symbols are the binaries also selected in this work, while empty symbols are those rejected.

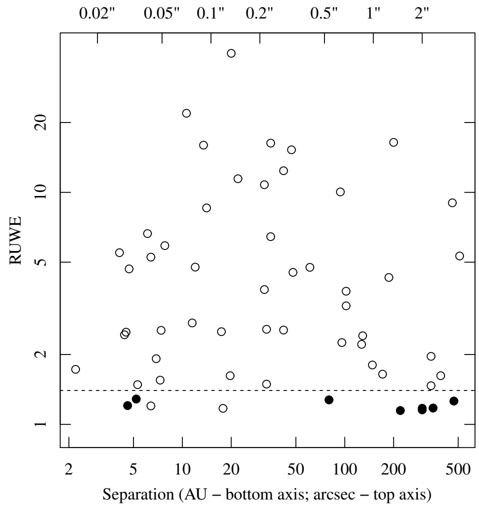

Fig. B.2. Diagram of RUWE values versus the binary separation of the Kraus et al. (2012) Taurus-Auriga binary-star list with *Gaia* EDR3 counterparts. Symbols are as in Fig. B.1.

We empirically checked if the Gaia-based selection used in this paper keeps the binary members of an SFR by matching the Taurus-Auriga binary-star list in Table 1 from Kraus et al. (2012) with the *Gaia* EDR3 catalogue and its subset selected in this paper. Out of 156 stars in Kraus et al., we found 142 Gaia-EDR3 counterparts within 0.5″, 40 of which were selected in this work using DBSCAN.

We then compared the RUWE distributions of the selected versus unselected systems to gain insight into how binary motions impact RUWE and the subsequent selection. Figure B.1 shows a diagram of RUWE versus *Gaia* colour $G_{BP}-G_{RP}$. The horizontal line indicates our maximum accepted RUWE value. Filled symbols are stars passing our selection, while empty symbols are the unmatched binaries, that is, those not retrieved in our catalogue. It should be noted that the cut in absolute-*G* magnitude rejects some stars that would have passed the RUWE constraint. Nevertheless, the vast majority of unmatched stars indeed have RUWE values well above the chosen limiting value and were likely rejected for this reason. There is little or no dependence of RUWE on *Gaia* colour, and therefore mass (although part of colour spread is also due to high extinction towards Taurus-Auriga). Also interesting is the diagram in Fig. B.2 showing RUWE versus binary separation. Here, the absence of any dependence of RUWE on projected separation (when measurable) is very evident, including unresolved pairs, where only the photocentre is affected by orbital motion. This latter diagram only contains six out of 40 stars selected by us, since about half of the Kraus et al. pairs are spectroscopic binaries with no measured separation. We also point out that out of the 76 binaries with no measured separation, 32 pass our selection (42%), while only eight out of the 66 binaries with measured separation pass the selection (12%), probably because photocentre motion has a smaller effect on astrometry compared to the motion of resolved components. Overall, extending this result from Taurus-Auriga to other SFRs at similar distances, we would predict a loss of ∼ 72% of their binary populations due to our selection criteria. Thus, if a binary frequency is as high as 50% (Mathieu 1994), a loss of ∼35% of PMS members can be expected. However, since this work selects stars at distances up to ∼1500 pc, this estimate should not be naively extended to our whole sample: the larger the distance, the smaller the projected binary motions, and hence, the closer they are to our detection limit. We therefore expect a less significant binary member loss for the farther-out SFRs. A more detailed study of these effects would, however, be far beyond the scope of this paper; one must also take into account that the new *Gaia* data release DR3 contains orbital astrometric solutions for 135 760 non-single stars[11].

Appendix C: Literature comparison

C.1. Taurus-Auriga association

The Taurus-Auriga complex is one of the nearest active SFRs of low-mass stars, to which many works have been dedicated. In this region, we identified several sub-structures, as can be seen from Fig. C.1. In order to identify the YSOs associated with the Taurus-Auriga association, we imposed the upper distance limit equal to 225 pc, as was done in Krolikowski et al. (2021), and restricted the spatial region in the ranges of 58.0°< *RA* <86.0° and 10.5°< *Dec* <38.5°. We considered the sub-structures whose members are all within these limits. With these conditions, we identified a total of 313 YSOs associated with six sub-structures. The spatial distributions are shown in Fig. C.1, while proper motions, parallaxes, and the CAMD are shown in Fig. C.2.

The members in the southwest sub-region (light blue plus symbols in the figures) are distributed quite close to the 10 Myr isochrone; thus, they could be part of an older population not selected by us for the photometric cut we used. However, with the exception of this, the members of the other sub-structures

---

[11] See https://www.cosmos.esa.int/web/gaia/

show the typical distribution of stars in PMS. The proper motions of the sub-structures are quite well distint, as well as parallaxes, suggesting a complex 3D structure with the known core including members in the $63.0^\circ \lesssim RA \lesssim 70.0^\circ$ region and $23^\circ \lesssim Dec \lesssim 28^\circ$ (blue star symbols in the Figures) also being on the close side (median distance equal to 132 pc). The easternmost and most populated sub-structure (red square symbols in the figures) is instead the most distant (median distance equal to 171 pc). Marginal evidence of age spread, as found in Krolikowski et al. (2021), is also found with our analysis, but our results cannot be considered conclusive as they are based on reddening-uncorrected photometry.

Krolikowski et al. (2021) very recently compiled the most complete and inclusive census of members of this region found in the literature. Among these, 587 objects have *Gaia* EDR3 counterparts, with 528 having a full astrometric solution. Using the *Gaia* EDR3 identification number given in the Krolikowski et al. (2021) table, we matched the list of the 437 Taurus members in the Krolikowski et al. (2021) table that are included in the photometric limits imposed in our work with the YSOs with $t < 10$ Myr (i.e. classified with flag from 1 to 28), and we found 202 objects in common, which amounts to about 46% (202/437) of the Krolikowski et al. (2021) list and 65% (202/313) of our list of YSOs in this region. We note that the Krolikowski et al. (2021) list was obtained from the compilation of previous works, including local spectroscopic and IR data surveys that do not homogeneously cover the entire region as we have done with *Gaia* data. For example, many of the 111 YSOs not included in the Krolikowski et al. (2021) table belong to clusters 579 and 572 in Table 3,which includes 88 and 30 YSOs, respectively (red squares and light blue symbols in Fig. C.1, top panel), in two sub-regions poorly covered by Krolikowski et al. (2021).

We also compared the list of YSOs in Taurus with the list of members identified as excess of mass (EOM) by Kerr et al. (2021) using *Gaia* DR2 data. Details on the match with our catalogue are given in Sect. C.4. As in our case, Kerr et al. (2021) found sub-structures beyond the distance of known members. To perform a consistent comparison, we restricted the Kerr et al. (2021) catalogue in RA, Dec, and distance, as was done for our catalogue. 429 were identified as EOM in this region. Among these, we considered the ones with $G > 7.5$ and $M_G > 5$ to match the same photometric region adopted for our catalogue. Of the 409 Kerr et al. (2021) YSOs that meet these conditions, we found that 197 (about 48%) are in common with our list of YSOs. We note that a rigorous consistent comparison is very hard to perform, since it strongly depends not only on the adopted clustering techniques but also on the sub-sample of *Gaia* data that is selected as starting point of the subsequent analysis.

The Taurus region is a well-known complex structure in which the membership has been very hard to achieve due to its large spatial extent and strong obscuration by the nebula. The comparison we made is sufficient to assert that about 50% of the selected YSOs in this region are already found in other surveys and that they are distributed in sub-structures that are consistent with those found in other works, and in particular with the results presented by Kerr et al. (2021), which homogeneously cover the entire region.

C.2. Orion complex

Young stellar objects associated with the Orion complex have been identified by selecting objects with $75^\circ < RA < 90^\circ$ and $-11^\circ < Dec < 10^\circ$. In this way, we found 18 840 YSOs associated with seven sub-structures with $t \lesssim 10$ Myr. These are shown in Fig. C.3 [12], where we note the presence of already known sub-structures such as λ and σ Ori, ONC, and 25 Ori. All the main sub-structures covering the Orion A and B nebulae have been merged by our procedure in a single complex including 14 832

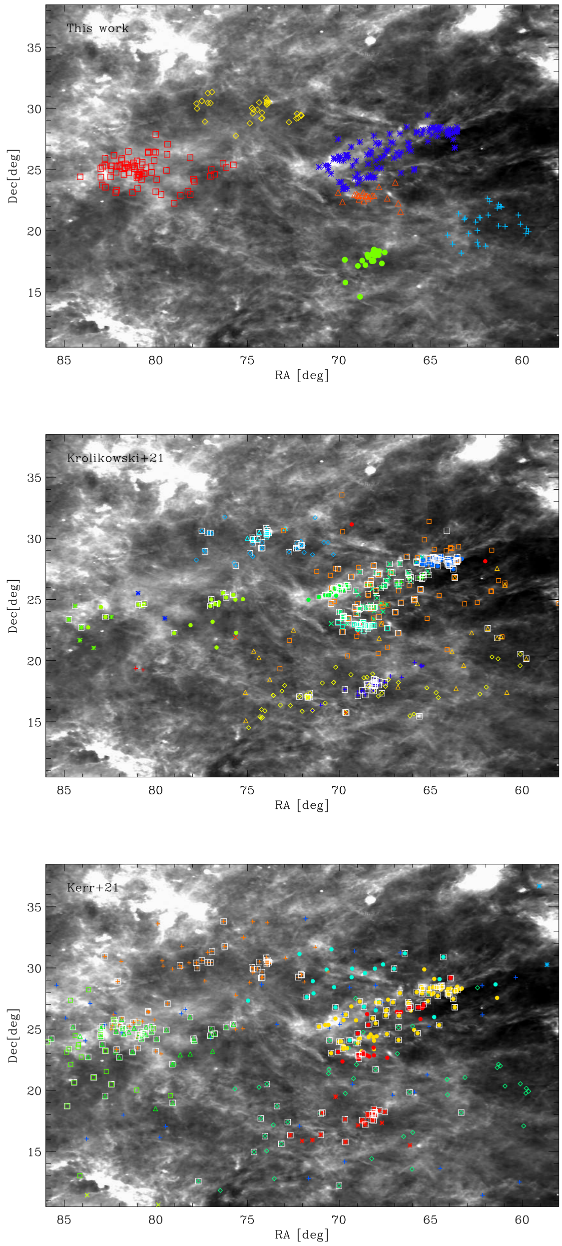

Fig. C.1. YSOs associated with Taurus-Auriga selected in this work (upper panel), Krolikowski et al. (2021; middle panel), and Kerr et al. (2021; lower panel). Colours and symbols indicate the sub-structures we found with DBSCAN, those derived by the Gaussian mixture model (GMM) in Krolikowski et al. (2021), and those derived as EOM by Kerr et al. (2021). White boxes in the middle and lower panels indicate the YSOs in common with our catalogue. YSOs are overplotted on an IRIS 100 μm image (Miville-Deschênes & Lagache 2005).

---

[12] For a direct visual comparison, spatial limits of the figure are the same as those used in Fig. 1 of Kounkel & Covey (2019).

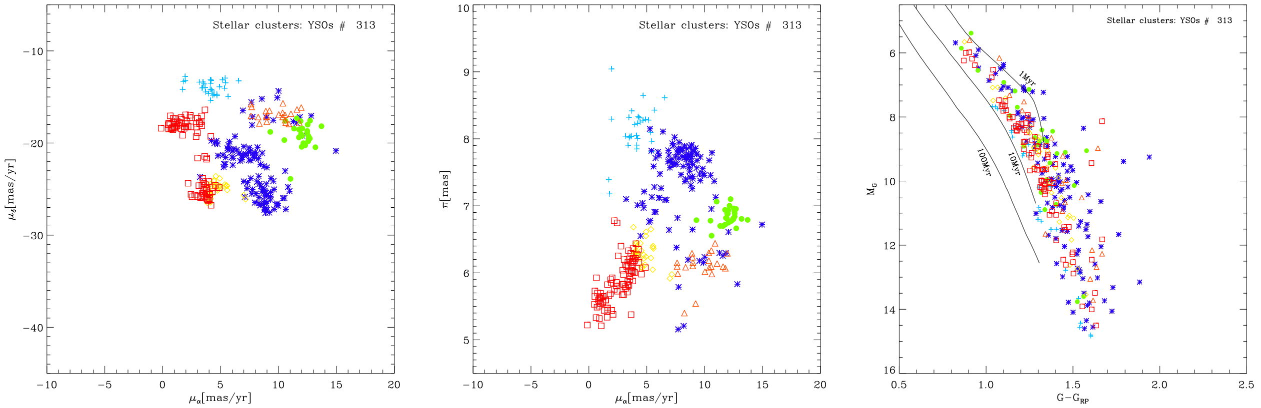

Fig. C.2. Proper motions in RA and Dec, parallaxes, and CAMDs of the clusters associated with the Taurus-Auriga complex. Three representative solar metallicity isochrones from the Pisa models are also shown. Symbols and colours are as in Fig. C.1.

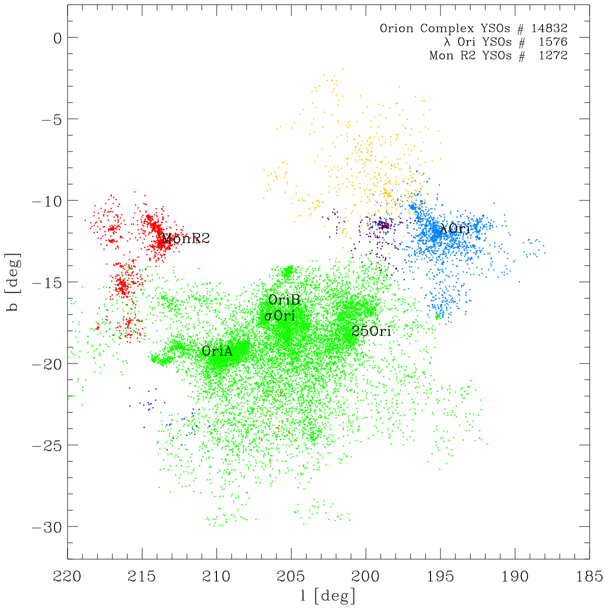

Fig. C.3. Spatial distribution in Galactic coordinates of YSOs associated with the Orion complex. YSOs identified in the seven substructures are drawn with different symbols and colours.

YSOs and a further 1 576 YSOs in the λ Ori cluster. The most distant cluster associated with Monoceros R2 (Mon-R2) is not part of the close Orion complex and includes 1 272 YSOs with a mean distance of 897 pc (σ =123 pc).

Figure C.4 shows proper motions and parallaxes of the substructures found in the Orion area. In particular, the proper motions show a very complex kinematic pattern of the subclusters in this region. However, a detailed analysis of the Orion kinematics is beyond the aims of this work.

Figure C.4 also shows the CAMD of the populations associated with Orion. Even though we cannot rigorously interpret it, as our data are not corrected for reddening, we note an apparent large age spread for all the populations.

We compare our findings in the Orion complex region with the Kounkel & Covey (2019) catalogue. Details on the match between the two catalogues are given in Sect. C.4. To retrieve the YSOs identified by Kounkel & Covey (2019) in the Orion complex, we considered the 16 structures (Theia groups) from their Table 2 falling in the Orion region as defined above. 11 882 and 10 373 YSOs in the Kounkel & Covey (2019) and Kounkel et al. (2020) catalogues are associated with the Theia groups of the Orion complex, respectively. 7 983 (67%) and 7 822 (75%) are in common with the list of Orion members found in this work.

The Orion complex has been extensively investigated with Spitzer IR data. For example, the Megeath et al. (2012) catalogue includes 3 479 YSOs stars[13] that cover a quite extended region of the Orion A and B nebulae. Using the cross-match service provided by CDS, Strasbourg, and a matching radius of 1″, we found that 2 612 IR sources from the Megeath et al. (2012) catalogue have a *Gaia* EDR3 counterpart. From this sample, we only considered those with photometric and astrometric restrictions given in Equation 1, with G-$G_{\rm RP}$ > 0.58 and in the ranges of 203°< l < 216° and −30°< b < 30°, which amount to 1 667 YSOs. Of these, we identified 1 561 (~94%) as members of the Orion complex. The spatial distributions of our members and those found in Megeath et al. (2012) are shown in Fig. C.5. This high percentage proves that *Gaia* data can accurately confirm membership of YSOs in SFRs comparably to IR data. If we consider the sub-sample of 2 612 Megeath et al. (2012) objects with Gaia counterparts, and assume that it includes only genuine YSOs (i.e. 0% contamination), we can conclude that our completeness level is about 60%. This value is the result of the restrictions we imposed on our initial data set to reduce the contamination level. We note that we can have a significant bias against (missed) binary stars. In fact, if we only discard the condition RUWE < 1.4 and retain the other conditions, there are 1 953 Megeath et al. (2012) YSO Gaia counterparts, and this implies that 286 YSOs (1953-1667), that is, about 14% of the total sample (very likely binary systems), are missing from our data set. We do not attempt to estimate the fraction of false positives that could be included in our sample by considering the Megeath et al. (2012) catalogue since it mainly includes Class II stars (i.e. YSOs with IR excess emission from the circumstellar disc), and it is therefore incomplete for the Class III stars, which do not show excess emission in the IR.

---

[13] retrieved at http://astro1.physics.utoledo.edu/~megeath/Orion/The_Spitzer_Orion_Survey.html