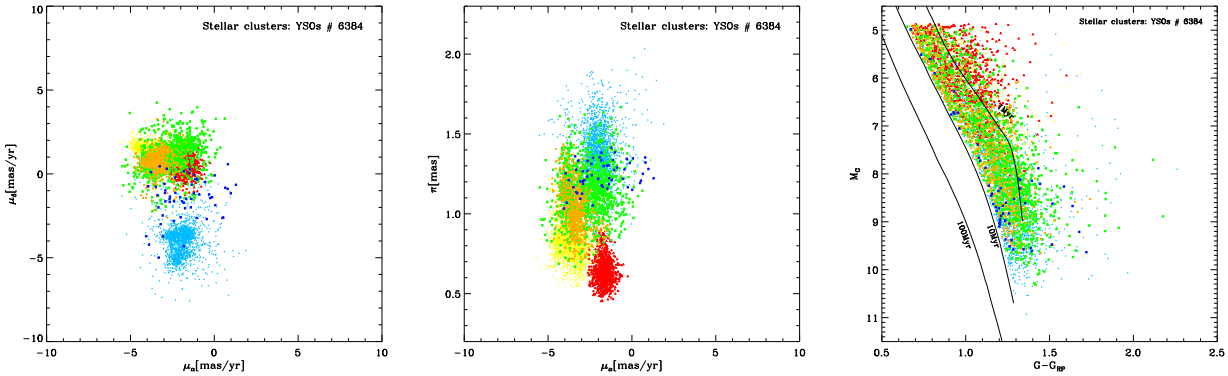

Fig. 9. Proper motions in RA and Dec, parallaxes, and CAMDs of the SFRs falling in the field of view of NGC 2264. The symbol colours of the clusters are as in Fig. 8. Three representative solar metallicity isochrones computed from the Pisa models are also shown.

now discuss these validated findings in the context of the GP structure within ∼1.5 kpc of the Sun. The maps of the young stellar clusters recognised by the DBSCAN clustering algorithm, most of them already known in the literature, are shown in the previous sections, and specific spatial and kinematic details are presented for some of them.

To identify clusters extended on scales larger than the 5° × 5° boxes used in the analysis, we merged adjacent clusters with consistent proper motions and distances. This procedure has been applied to identify extended SFRs as a whole, as in the case of the Orion complex or Sco OB2 UCL, with r_{50} equal to ≃17°and ≃15°, respectively, which are among the most extended structures resolved in this work. In several cases, it identifies clusters that encompass multiple populations, as in the case of NGC 2264, which was identified as a unique structure also including the close cluster IC 446 and other YSOs in the surrounding region. A more in-depth analysis of the two clusters shows that their proper motions can be distinguished into slightly different sub-populations. Thus, our overall procedure used to define clusters tends to include multiple sub-populations sharing similar properties, which are likely associated with the progenitor molecular cloud.

The question of cluster and sub-cluster identification is a very complex issue that can be dealt with at the different spatial precision levels required for a given analysis. This was done, for example, for the MYStiX project in Feigelson (2018), where a parametric statistical regression approach providing hierarchical ellipsoid structures was adopted. The evidence of a wide range of central surface densities found in the MYStIX maps is in agreement with the different spatial morphology of the SFRs identified in this work.

Figure 10 shows the spatial distribution of the young stellar clusters found in this work in three different distance bins: [100, 600] pc, [600, 2000] pc, and [100, 2000] pc. The young clusters are drawn by distinguishing them in the age bins $t < 10$ Myr, 10 Myr $< t <$ 100 Myr, and $t <$ 100 Myr. We note that clusters with 10 Myr $< t <$ 100 Myr were only found in the solar neighbourhood (<600 pc) and thus are only shown in the [100, 600] pc distance range.

The distribution of SFRs ($t < 10$ Myr) within 600 pc is dominated by the presence of big young structures crossing the GP such as the Orion and Perseus complexes, Gamma Velorum (Pozzo 1), and Lac OB1, which are under the GP, BH 23 (corresponding to Theia 80 in Kounkel & Covey 2019); and RSG 8, which is close to the GP, Serpens, Alessi 62, Collinder 359, and Rho Ophiuchi (over the GP). The clusters with ages of 10 Myr $< t <$ 100 Myr in the same distance range definitely appear more diffuse. Apart from the well-known Sco-Cen association covering ∼60° in longitudes, we detected the similarly huge association in the Vela-Puppis region as a unique complex, including Trumpler 10, γ Velorum, NGC 2457, and NGC 2451B, as well as the associations around NGC 2232, Roslund 5, and Alessi 19. Their positions appear to be connected to the clusters with $t < 10$ Myr since they follow a spatial pattern crossing or one very close to that of the SFRs. This suggests that they likely belong to a common star formation process encompassing at least two generations of YSOs, with the first generation including extended populations of dissolving young clusters and associations.

The large Sco-Cen association is connected to the Vela and Orion complexes, confirming what was already found by Bouy & Alves (2015) with Hipparcos data. These three regions are described there as three large-scale stream-like structures.

Going towards larger distances ($d \gtrsim 600$ pc), the SFRs show a more regular pattern, which is approximately parallel to the GP. The most prominent SFRs are ASCC 32 and Cep OB3b in the Cepheus, respectively under and over the GP at a distance of ∼800–900 pc. Among the most distant SFRs with more than 300 members and distances ≳1400 pc, we detected NGC 2244, NGC 6530, NGC 6531 , NGC 2362, and FSR 0442.

The overall distribution of YSOs in SFRs with $d \lesssim 600$ pc traces a complex 3D pattern in the solar neighbourhood. In particular, in the Z versus X edge-on Galactic projection (see bottom left panel in Fig. 4 and top left panel in Fig. 10) we find evidence of a projected inclined structure, mainly traced by the Orion, Vela OB2, and Rho Ophiuchi star forming complexes in the third and fourth Galactic quadrants and by the Serpens, Lacerta OB1 and Perseus in the first and second Galactic quadrants. However, the SFRs falling in the Cepheus region do not follow this pattern. A global view of these structures and their spatial correlation with the surrounding nebular emission suggests a pattern consistent with the results found in Molinari et al. (2010), where massive proto-clusters and entire clusters of YSOs in active SFRs are associated with clouds that collapse into filaments.

As already found in Zari et al. (2018), current data reveal a very complex 3D structure that cannot be simply described with the Gould Belt, that is, the giant flat structure inclined by ∼20° with respect to the GP, first pointed out by Gould (1879). This insight was already suggested by Guillout (2001), who presented the first detection of the Gould Belt late-type star population and proposed the alternative scenario of a Gould disc.

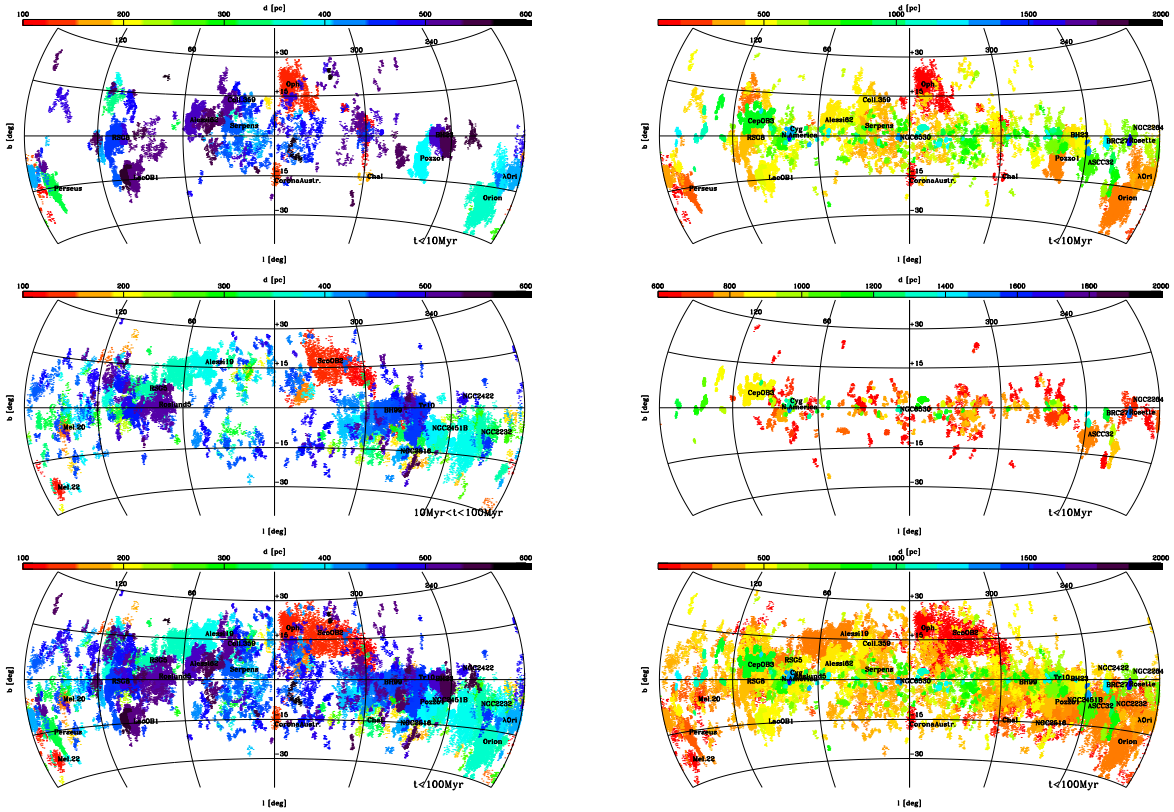

Fig. 10. Aitoff projections in Galactic coordinates of the YSOs in the different age bins ($t < 10$ Myr, 10 Myr $< t <$ 100 Myr and $t <$ 100 Myr), with distance in the range ([100, 600] pc (*left panels*), [600, 2000] pc (*mid right panel*) and [100, 2000] pc) (*upper and bottom right panels*). Colour codes indicate stellar distances.

A more detailed representation of the young Galactic component in the Solar neighbourhood was recently proposed by Alves et al. (2020), who determined the 3D distribution of all local cloud complexes by deriving accurate distances to about 380 lines of sight. They suggested that such 3D distribution could be described by a damped sinusoidal wave, which they call the Radcliffe wave, with an amplitude of ∼160 pc and a period of ∼2 kpc. It crosses Orion (around a minimum), Cepheus (crest), North America, and Cygnus X. This structure is separated and distinct from a second structure, indicated as a "split", crossing the Sco-Cen, Aquila, and Serpens clouds. They propose that the Gould Belt is a projection effect of two linear cloud complexes. The spatial distribution of YSOs associated with SFRs that has been identified in our work shows much more complex and diffuse structures, but the two elongated linear structures suggested by Alves et al. (2020) approximately cross the borderline of the two separated structures visible in the X, Y map of Fig. 4, delimited by the SFRs indicated by Alves et al. (2020). This leads us to confirm that the local young Galactic component is very complex. While our data are broadly consistent with the Alves et al. (2020) findings, further investigations, including a more detailed analysis of the kinematics of the structures based on the 3D space coordinates (X, Y, Z) and velocities (U, V, W; e.g. de Zeeuw et al. 1999), are required to confirm the scenario and to find additional insights regarding their origin.

To gain further insights concerning the star formation history of the SFRs, it is crucial to derive more accurate stellar ages. However, we do not attempt to derive stellar ages of the selected YSOs for several reasons. First of all, we lack a suitable photometric system. In fact, the large *Gaia* EDR3 G and RP photometric bands used for this work are not sensitive to the fundamental stellar parameters (effective temperatures, stellar ages, etc.), especially for low-mass stars. However, future *Gaia* releases, overcoming issues related to the BP bands at faint magnitudes, could be crucial to this aim. Secondly, we lack the spectroscopic data needed to derive individual stellar reddening values to appropriately place these YSOs on the HR diagram. Alternatives such as the use of 3D reddening maps (e.g. Bovy et al. 2016; Lallement et al. 2019) require careful analysis, since the integrated extinction tends to be underestimated in the molecular clouds, where SFRs are typically located. A detailed analysis is deferred to future works based on the combination of *Gaia* and spectroscopic data from available surveys, such as *Gaia*-ESO (Gilmore et al. 2012; Randich et al. 2013), LAMOST (Zhao et al. 2012), or APOGEE (Majewski et al. 2017), or future surveys such as WEAVE (Dalton et al. 2012) and 4MOST (Guiglion et al. 2019).

7. Summary and conclusions

We used the machine learning unsupervised clustering algorithm DBSCAN to systematically identify all SFRs with ages of $t \lesssim 10$ Myr within ∼1.5 kpc of the Sun. The density-based clustering algorithm was applied to the *Gaia* EDR3 positions, parallaxes, and proper motions of a photometrically selected starting sample.

A pattern-match procedure based on a template data set including typical clusters detected within the photometric sample was used to distinguish very young clusters from the

contaminant old clusters and from photometrically unphysical clusters. We provide here a catalogue with the main parameters (positions, spatial extent, median distance and number of members) of the 354 SFRs with ages of $t \lesssim 10$ Myr. The parameters of the 322 young clusters with ages of 10 Myr $\lesssim t \lesssim$ 100 Myr are also given. We also provide the list of 124 440 and 65 863 YSOs found in the SFRs and the young clusters, respectively, mainly including late-type K-M stars. A substantial number of YSOs have been recognised for the first time. Based on the comparison of our list of YSOs in the well-known Sco-Cen region and in NGC2264, we roughly estimate that within our observational limits the completeness of the census of cluster members obtained with our analysis is $\gtrsim$85%, at least in very rich and concentrated SFRs. For low-density regions, such as the Taurus-Auriga association (see Appendix C), this completeness figure is expected to be around 50%. The mass-function coverage of each cluster strongly depends on the cluster distance and is set by the observational limit. Compact regular clusters and SFRs in large complexes such as Taurus, Orion, Sco-OB2, Perseus, and Cygnus, were identified with a high level of efficiency, as estimated from the comparison with other available catalogues (see Appendix C).

The overall distribution of these clusters in the Galaxy context shows that they are distributed along a very complex 3D pattern that seems to connect them at least within 500–600 pc. Outside of this distance, the clusters appear to be more regularly and closely distributed along the GP.

As far as we know, the catalogue of YSOs presented in this work is the sole all-sky catalogue based on the most recent *Gaia* EDR3 data, which benefit from major improvements with respect to *Gaia* DR2. This catalogue represents a step forwards in the census of SFRs and can be used, for example, for further detailed interpretations of their spatial distribution in the context of the spiral arm model (Reid et al. 2019), since it covers a substantial region crossed by the Local Arm and, marginally, some regions of the Perseus and Sagittarius-Carina arms (Poggio et al. 2021). Future and photometric deep surveys, such as the Rubin Legacy Survey of Space and Time (LSST) will allow us to extend these limits.

We note that at this stage these results are not suitable for studies of IMF, star formation history or cluster dynamics, those based on the full space 3D velocity determination, since the census of the SFRs is not complete, and accurate masses and ages, as well as radial velocities can not be determined, until further data are available. Nevertheless, the dominant component of the SFRs has been detected, and thus these results can be used as driving samples for the extraction of complete populations from *Gaia* data by relaxing the stringent constraints adopted in this work. Finally, the SFRs identified in this work are defined well enough to allow detailed studies of circumstellar disc evolution and direct imaging of young giant planets based on multi-band analyses of available or future additional observations (X-rays or IR or radio) targeting some of the individual clusters.

Acknowledgements. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. E.T. acknowledges Czech Science Foundation GAČR (Project: 21-16583M). J.M.A. acknowledges financial support from the project PRIN-INAF 2019 "Spectroscopically Tracing the Disk Dispersal Evolution". The authors are very grateful to the anonymous referee, for providing constructive comments and suggestions which significantly contributed to improving this publication.

References

Alves, J., Zucker, C., Goodman, A. A., et al. 2020, Nature, 578, 237
Anders, F., Khalatyan, A., Chiappini, C., et al. 2019, A&A, 628, A94
Avedisova, V. S. 2002, Astron. Rep., 46, 193
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
Bica, E., Pavani, D. B., Bonatto, C. J., & Lima, E. F. 2019, AJ, 157, 12
Bonito, R., Prisinzano, L., Guarcello, M. G., & Micela, G. 2013, A&A, 556, A108
Bouy, H., & Alves, J. 2015, A&A, 584, A26
Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, ApJ, 818, 130
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Cantat-Gaudin, T., & Anders, F. 2020, A&A, 633, A99
Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1
Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A&A, 618, A59
Castro-Ginard, A., Jordi, C., Luri, X., et al. 2020, A&A, 635, A45
Chabrier, G. 2003, PASP, 115, 763
Chen, W. P., & Lee, H. T. 2008, in Handbook of Star Forming Regions, ed. B. Reipurth, 4, 124
Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525
Cody, A. M., Stauffer, J., Baglin, A., et al. 2014, AJ, 147, 82
Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, in SPIE Conf. Ser., 8446, 84460P
Damiani, F., Micela, G., Sciortino, S., et al. 2006, A&A, 460, 133
Damiani, F., Prisinzano, L., Pillitteri, I., Micela, G., & Sciortino, S. 2019, A&A, 623, A112
Dell'Omodarme, M., Valle, G., Degl'Innocenti, S., & Prada Moroni, P. G. 2012, A&A, 540, A26
de Zeeuw, P. T., Hoogerwerf, R., de Bruijne, J. H. J., Brown, A. G. A., & Blaauw, A. 1999, AJ, 117, 354
Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871
Dutra, C. M., & Bica, E. 2002, A&A, 383, 631
Ercolano, B., Picogna, G., Monsch, K., Drake, J. J., & Preibisch, T. 2021, MNRAS, 508, 1675
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Second International Conference on Knowledge Discovery and Data Mining, eds. J. Simoudis, E. Han, & U. Fayyad (Menlo Park, CA: AAAI Press), 226
Fasano, G., & Franceschini, A. 1987, MNRAS, 225, 155
Feigelson, E. D. 2018, in Astrophysics and Space Science Library, The Birth of Star Clusters, ed. S. Stahler, 424, 119
Fernández-López, M., Arce, H. G., Looney, L., et al. 2014, ApJ, 790, L19
Flaccomio, E., Micela, G., & Sciortino, S. 2006, A&A, 455, 903
Franciosini, E., Tognelli, E., Degl'Innocenti, S., et al. 2021, A&A, 659, A85
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
Gould, B. A. 1879, Result. Observ. Nacional Argentino, 1, I
Guiglion, G., Battistini, C., Bell, C. P. M., et al. 2019, The Messenger, 175, 17
Guillout, P. 2001, in Astronomical Society of the Pacific Conference Series, From Darkness to Light: Origin and Evolution of Young Stellar Clusters, eds. T. Montmerle, & P. André, 243, 677
Gullbring, E., Hartmann, L., Briceno, C., & Calvet, N. 1998, ApJ, 492, 323
Jackson, R. J., Jeffries, R. D., Wright, N. J., et al. 2022, MNRAS, 509, 1664
Jeffries, R. D., Jackson, R. J., Franciosini, E., et al. 2017, MNRAS, 464, 1456
Kerr, R. M. P., Rizzuto, A. C., Kraus, A. L., & Offner, S. S. R. 2021, ApJ, 917, 23
Kervella, P., Arenou, F., & Thévenin, F. 2022, A&A, 657, A7
Kounkel, M., & Covey, K. 2019, AJ, 158, 122
Kounkel, M., Covey, K., & Stassun, K. G. 2020, AJ, 160, 279
Kraus, A. L., Ireland, M. J., Hillenbrand, L. A., & Martinache, F. 2012, ApJ, 745, 19
Krolikowski, D. M., Kraus, A. L., & Rizzuto, A. C. 2021, AJ, 162, 110
Kronberger, M., Teutsch, P., Alessi, B., et al. 2006, A&A, 447, 921
Krone-Martins, A., & Moitinho, A. 2014, A&A, 561, A57
Lada, C. J. 2006, ApJ, 640, L63
Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, A&A, 625, A135
Le Duigou, J. M., & Knödlseder, J. 2002, A&A, 392, 869
Lindegren, L., Bastian, U., Biermann, M., et al. 2021a, A&A, 649, A4
Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021b, A&A, 649, A2
Liu, L., & Pang, X. 2019, ApJS, 245, 32
Luhman, K. L. 2022, AJ, 163, 24
Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, A&A, 616, A9
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94