

3D mapping of young stars in the solar neighbourhood with *Gaia* DR2

E. Zari¹, H. Hashemi¹, A. G. A. Brown¹, K. Jardine², and P.T. de Zeeuw¹

¹ Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, the Netherlands;

² Consultant, Radagast Solutions, Simon Vestdijkpad 24, 2321 WD Leiden, the Netherlands; galaxymap.org

October 24, 2018

ABSTRACT

We study the three dimensional arrangement of young stars in the solar neighbourhood using the second release of the *Gaia* mission (*Gaia* DR2) and we provide a new, original view of the spatial configuration of the star forming regions within 500 pc from the Sun. By smoothing the star distribution through a gaussian filter, we construct three dimensional density maps for early-type stars (upper-main sequence, UMS) and pre-main sequence (PMS) sources. The PMS and the UMS samples are selected through a combination of photometric and astrometric criteria. A side product of the analysis is a three dimensional, *G*-band extinction map, which we use to correct our colour-magnitude diagram for extinction and reddening. Both density maps show three prominent structures, *Scorpius-Centaurus*, *Orion*, and *Vela*. The PMS map shows a plethora of lower mass star forming regions, such as *Taurus*, *Perseus*, *Cepheus*, *Cassiopeia* and *Lacerta*, which are less visible in the UMS map, due to the lack of large numbers of bright, early-type stars. We report the finding of a candidate new open cluster towards $l, b \sim 218.5^\circ, -2^\circ$, which could be related to the Orion star forming complex. We estimate ages for the PMS sample and we study the distribution of PMS stars as a function of their age. We find that younger stars cluster in dense, compact clumps, and are surrounded by older sources, whose distribution is instead more diffuse. The youngest groups that we find are mainly located in *Scorpius-Centaurus*, *Orion*, *Vela* and *Taurus*. *Cepheus*, *Cassiopeia*, and *Lacerta* are instead more evolved and less numerous. Finally, we find that the three dimensional density maps show no evidence for the existence of the ring-like structure which is usually referred to as the Gould Belt.

Key words. Stars: distances - stars: formation - stars: pre-main sequence - stars: early-type - Galaxy: solar neighbourhood - Galaxy: open clusters and associations

1. Introduction

Since the second half of the 19th century, it was recognised by Herschel (1847) and Gould (1874) that the brightest stars are not distributed randomly in the sky, but seemed to form a belt (which afterwards became known as the Gould Belt) with an inclination of $\sim 20^\circ$ with respect to the plane of the Milky Way. Furthermore, O and B type stars clustered in loose groups that were named 'associations' by Ambartsumian (1947). The Gould Belt was subsequently found to be associated with a significant amount of interstellar material (Lindblad 1967), interpreted as an expanding ring of gas (Olano 1982; Elmegreen 1982). Giant molecular clouds were also found to be related to the most prominent OB associations (Sancisi et al. 1974; Kutner et al. 1977; de Geus 1992; Dame 1993). This agrees well with the fact that OB associations are young, as supported by the ages derived from color-magnitude diagrams.

The origin of the Belt is debated, and various formation scenarios have been proposed. Comeron & Torra (1992) and Comeron et al. (1998) proposed that the Gould Belt was formed after the oblique impact of a high velocity cloud on the galactic disk. Poppel (1997) suggested instead a cascade of supernova explosions. Alternatively, Olano (2001) proposed that a $2 \times 10^7 M_\odot$, 400 pc size supercloud is the common precursor of the Sirius super cluster, the Gould Belt, and the Local Arm. The breaking and compression of the supercloud would have produced the latter two, while the cluster, unaffected by friction would have moved on, away from the gas system. Finally, Bekki (2009) sug-

gests that the Belt was formed after the collision between a gas cloud of $\sim 10^6 M_\odot$ and a $\sim 10^7 M_\odot$ dark matter clump, based on numerical simulations of the collision.

Many studies have described the structure and the kinematics of the Gould Belt. Thanks to the data of the *Hipparcos* satellite, the definition and characterization of nearby OB associations and open clusters was improved (de Zeeuw et al. 1999; de Bruijne 1999a; Hoogerwerf & Aguilar 1999; Elias et al. 2006a,b, 2009; Bouy & Alves 2015) and our knowledge of the structure of the solar neighbourhood amplified.

In particular, Elias et al. (2006a) first studied the three dimensional spatial distribution of early type stars within 1 kpc from the Sun, by modelling the star distribution with two interacting discs, the Gould Belt and the Local Galactic Disc.

Bouy & Alves (2015) revisited the distribution of OB stars in the solar neighbourhood by constructing a 3D map of their spatial distribution. They found three stream-like structures (named *Scorpius-Canis Major*, *Vela*, and *Orion*), not only coherent in space but also characterized by monotonic age sequences. The main conclusion emerging from Elias et al. (2006a) and Bouy & Alves (2015) is that there is no evidence of a ring-like structure in the three dimensional configuration of young, bright stars in the solar neighbourhood. The Gould Belt as perceived by Herschel and Gould would be due to a projection effect according to Bouy & Alves (Orion and Sco-Cen causing the apparent tilt due to their locations below and above the plane.)

In this work, we make use of the second data release of the *Gaia* mission, hereafter *Gaia* DR2, to study the three di-

mensional configuration of the solar neighbourhood, focusing on young groups and OB associations. We also study the star formation history of the solar neighbourhood by estimating the ages of the young groups that we find.

In Sec. 2 we give a short description of the data, which we divide in two samples, the upper main sequence (UMS) and the pre-main sequence (PMS). We further describe the selection procedure that we used to derive astrometrically 'clean' samples, and the photometric and kinematic selection criteria that we apply. In Sec. 3 we describe the methods used to obtain a three dimensional map of the solar neighbourhood, and we study the three dimensional distribution of the UMS and PMS samples in terms of ages. In Sec. 4 we discuss our findings. Finally in Sec. 5 we summarize our results and draw our conclusions.

2. Data

In this section we present the selection criteria that we used. We refer to Gaia Collaboration et al. (2016, 2018b) and Lindegren et al. (2018) for a detailed description of the data. The queries that we used to retrieve the data from the *Gaia* archive are reported in Appendix A.

We selected all the stars within $d = 500$ pc from the Sun ($\varpi \geq 2$ mas), and we divided them in two samples, the upper main sequence sample (UMS) and the pre-main sequence sample (PMS). There are two reasons for this division. The first reason concerns the data analysis procedure: dividing the initial sample allows to apply different selection criteria that are more suitable for one sub-sample or the other. The second reason has instead a scientific justification: it is indeed interesting to study UMS and PMS as two separate samples in order to compare the distribution of young, high-mass stars and low-mass sources.

Both samples are selected by combining photometric and astrometric criteria. With regards to the photometric criteria, the first step in our procedure consists of correcting for extinction and reddening the colour-magnitude diagrams. The method that we apply to do such a correction is presented in Section 2.1, and applied to the UMS and PMS samples in Section 2.2 and 2.3 respectively. The final result of the data selection consists of a catalogue of UMS and PMS stars, which will be available on CDS. We shortly describe the catalogue columns in Appendix F.

2.1. Extinction correction

G band extinction, A_G , and colour excess, $E(G_{BP} - G_{RP})$, are reported in the *Gaia* DR2 catalogue for a sub-set of sources, with measured parallax. Although single extinction and/or reddening values are inaccurate on a star-by-star level, they are mostly unbiased and can be used reliably at the ensemble level (Andrae et al. 2018). We can therefore compute extinction (and colour excess) as a function of position and distance, create a three dimensional A_G map, and assign to the stars without measured extinction and colour excess a value of A_G and $E(G_{BP} - G_{RP})$ based on the 3D map. In this way, we aim at producing a dereddened colour magnitude diagram, to better isolate young star forming regions. We use *Gaia* DR2 extinction and reddening values mainly for two reasons. On the one hand, cross-matching with other catalogues, such as 2MASS (see e.g. Katz et al. 2018; Poggio et al. 2018), significantly reduces the number of sources, while we aim to use as many sources as possible. On the other hand, although three dimensional extinction maps are available, they generally report extinction values in the V band. Thus, one should transfer the V band extinction to the *Gaia* DR2 bands through photometric transformation (or vice-versa). Even

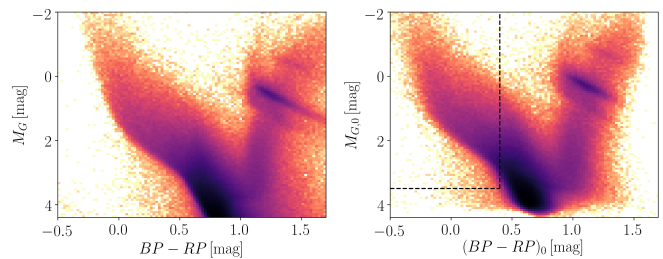


Fig. 1. Upper main sequence colour magnitude diagrams. Left: colour magnitude diagram before correcting for extinction and colour excess. Right: colour magnitude diagram after accounting for extinction and reddening. The dashed lines limit the region we considered as upper-main sequence in this study.

though this is in principle possible, it is very error-prone as the transformation between A_V and A_G and between $E(B - V)$ and $E(G_{BP} - G_{RP})$ is non-trivial due to the very wide photometric bands used by *Gaia* (see Andrae et al. (2018) for more details). To create the map, we proceed as follows. We query all the sources with $\varpi > 2$ mas, $\varpi/\sigma_\varpi > 5$ and a measured A_G value. Then, we compute the source galactic Cartesian coordinates, x, y, z . We define a volume $N = 1000 \times 1000 \times 1000$ pc centred on the Sun and we divide it in cubes n of $10 \times 10 \times 10$ pc each. For each cube, we compute the average extinction and colour excess. In this way, we obtain a crude map that however delivers better results than the alternatives described above. Finally, we assign to all the sources the appropriate extinction and colour excess values according to their position in space, and we correct the observed M_G vs. $G_{BP} - G_{RP}$ colour magnitude diagram.

2.2. Upper Main Sequence

To construct the sample, we first downloaded from the *Gaia* archive bright and blue sources, nominally closer than $d = 500$ pc from the Sun:

$$\begin{aligned} M_G &\leq 4.4 \text{ mag;} \\ (G_{BP} - G_{RP}) &\leq 1.7 \text{ mag;} \\ \varpi &\geq 2 \text{ mas;} \\ \varpi/\sigma_\varpi &> 5 \end{aligned} \quad (1)$$

By using the extinction A_G and colour excess $E(G_{BP} - G_{RP})$ values computed in Section 2.1, we correct the colour-magnitude diagram for extinction and reddening, and apply the following selection criteria:

$$\begin{aligned} M_{G,0} &\leq 3.5 \text{ mag;} \\ (G_{BP} - G_{RP})_0 &\leq 0.4 \text{ mag;} \end{aligned} \quad (3)$$

The first and second condition aim at selecting sources whose colours are consistent with being of spectral type O, B, or A. The condition $\varpi/\sigma_\varpi > 5$ is primarily motivated by the fact that in the rest of the paper we compute distances simply by inverting parallaxes, ($d = 1000/\varpi$ pc), and this holds only when parallax errors are small (Bailer-Jones 2015). Fig. 1(left) shows the initial colour-magnitude diagram used for the selection. Fig. 1(right) shows the conditions on colour and magnitude as black dashed lines.

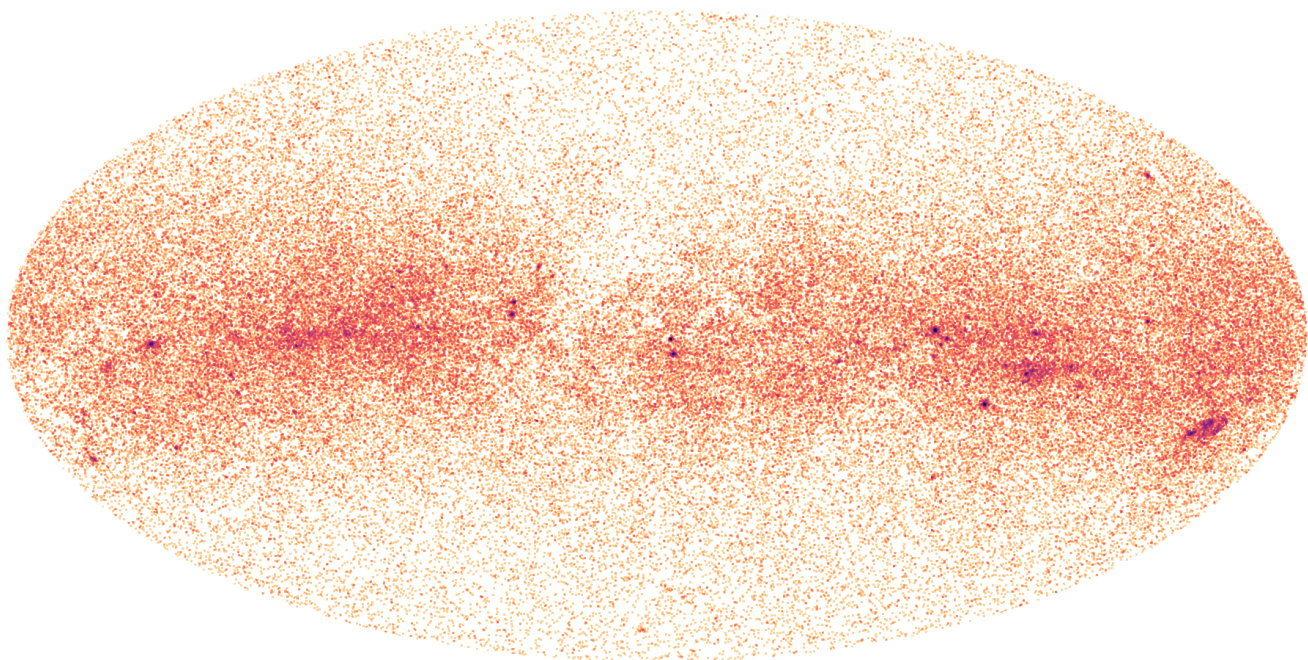


Fig. 2. Upper main sequence sources selected by applying the conditions in Sec. 2.2. The sources are concentrated towards the galactic plane, and their density decreases towards the poles. Clumps corresponding to known open clusters and associations are visible.

2.2.1. Tangential velocities

Fig. 2 shows the distribution of the UMS sources selected in Sec. 2.2. The density of sources increases towards the galactic plane, and some known clusters are visible. Members of clusters and associations share the same spatial velocity, with a small velocity dispersion that varies from a few tenths to some km/s respectively. In proper motion or tangential velocity space, they appear as density enhancements with respect to the underlying, broad field star distribution. Therefore, to clean our sample, we study the tangential velocities distribution ($v_{l,b} = A\mu_{l,b}/\varpi$, where $A = 4.74047 \text{ km s}^{-1} \text{ yr}^{-1}$) of the stars we have selected so far.

Fig. 3 shows an unsharp mask of the tangential velocity distribution of the UMS sample. We use a two-dimensional gaussian filter, with bandwidth = 30 km s^{-1} to smooth the tangential velocity distribution. This produces a blurred ('unsharp') mask of the original distribution. The unsharp mask is subtracted from the original tangential velocity distribution, which was smoothed as well with a gaussian filter of bandwidth = 1 km s^{-1} . Finally we compute the quantity:

$$S = \frac{I_1 - I_{30}}{I_{30}}, \quad (4)$$

where I_x represents the smoothed tangential velocity distribution. S is then a measure of the contrast of the density enhancements with respect to a uniform, smooth distribution. We selected the stars within the $S = 1$ levels, shown as black solid lines in Fig. 3. Fig. 4 shows the distribution in the sky of the sources selected in this fashion. The number of sources at high galactic latitudes visibly decreases with respect to Fig. 2: this indicates that the tangential velocity selection is useful to reduce the contamination level of our sample, since we expect young stars to be mainly located towards the galactic plane. On the other hand, such a selection will reject young stars with peculiar tangential velocities (such as binaries or runaways): we stress however that the scope of this study is to focus on the bulk of the

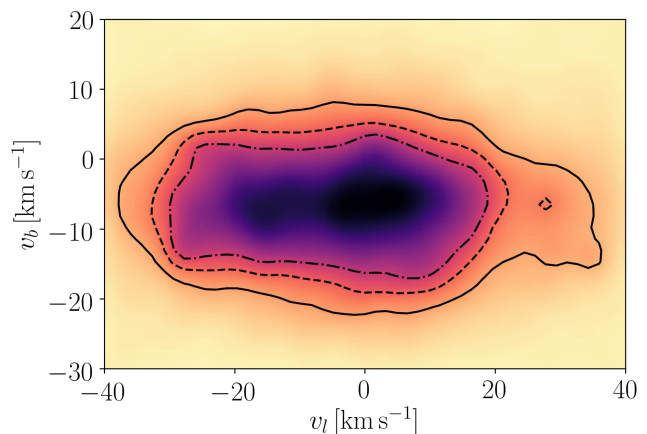


Fig. 3. Smoothed tangential velocity distribution of the UMS sample, defined in Eq. 3 in the text. The contours represent the $S = 1, 2, 3$ levels. The density enhancements correspond to known clusters and associations. Note also that the distribution is not centred in $v_l, v_b = (0, 0)$ due to the solar motion.

early-type population and not on the kinematic outliers, which represent a small fraction of the population.

2.3. Pre-Main Sequence

To select the pre-main sequence (PMS) sample, we first downloaded from the *Gaia* archive all the sources nominally within $d = 500 \text{ pc}$. Due to the large number of sources, the query can not be executed as a single query, but the data has to be divided, for example in parallax bins. After joining all the separate tables, we proceed as follows.

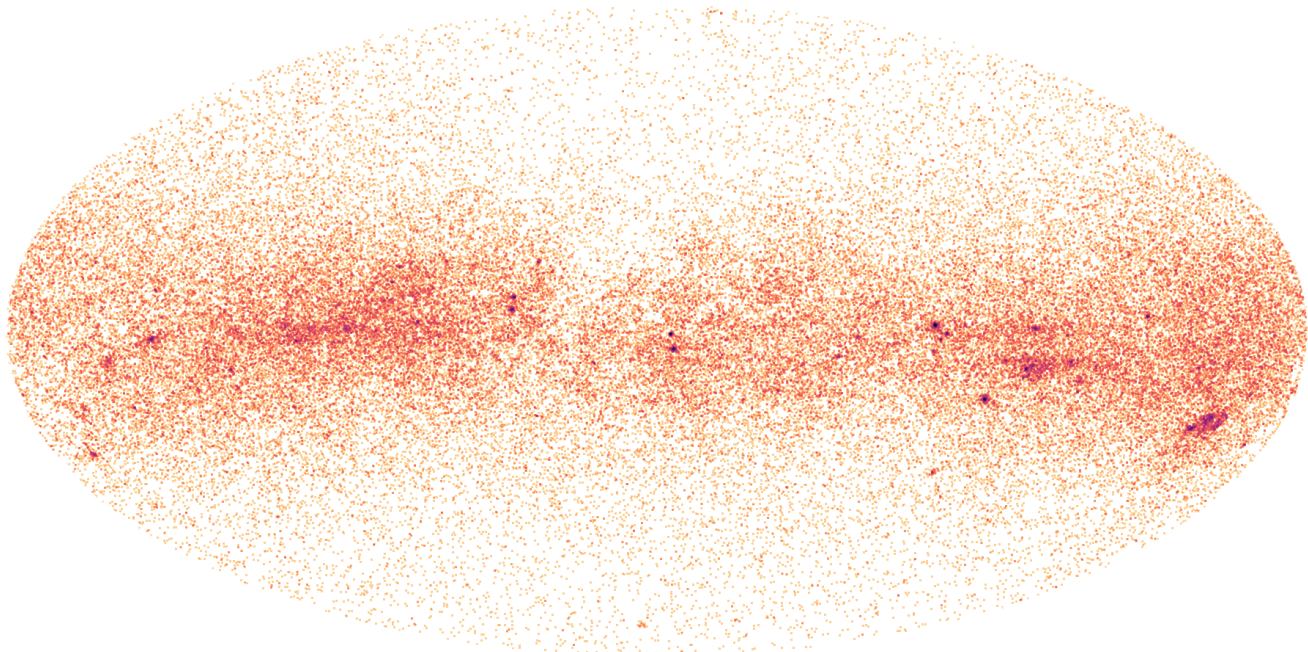


Fig. 4. Distribution of the sources in the Sky after the selection based on tangential velocities. The number of sources at high galactic latitudes has decreased with respect to Fig. 2, which indicates that many contaminants have been discarded.

2.3.1. Astrometrically 'clean' subset

We first applied Eq. C.1 and C.2 in Lindegren et al. (2018), and required that $\varpi/\sigma_\varpi > 5$. Eq. C.1 and C.2 were used by Lindegren et al. (2018) to produce a 'clean' HR diagram of nearby stars ($d < 100$ pc). Eq. C.1 is meant to remove sources with spuriously high parallax. Eq. C.2 deals with the photometric errors in the BP and RP bands, affecting in particular faint sources and crowded areas. We selected stars with small parallax error ($\sigma_\varpi/\varpi < 20\%$) with the same motivations as for the UMS sample. Finally we decided to restrict our sample to stars following the disc kinematics. Thus we required the total tangential velocity to be lower than 40 km s^{-1} :

$$v_t = \sqrt{v_l^2 + v_b^2} < 40 \text{ km s}^{-1}.$$

The condition on the tangential velocity follows Gaia Collaboration et al. (2018a). Usually the cut to select thin disc stars is $v_{\text{TOT}} < 50 \text{ km s}^{-1}$ (e.g. Bensby et al. 2014), however we only have two velocity components instead of three, thus we adapted the cut to take this into account.

2.3.2. Extinction correction and selection of the PMS

We first corrected for extinction and reddening using the procedure described in Section 2.1. Then, we used the PARSEC Isochrones (Bressan et al. 2012) version 1.2S (Chen et al. 2014, 2015; Tang et al. 2014) with $A_V = 0$ mag and solar metallicity ($Z = 0.0152$) to define the main sequence track and the binary sequence (which is brighter than the main sequence by 0.75 mag), and we selected all the stars brighter than the binary sequence. We further restrict our sample to sources with $M_{G,0} > 4$ mag: this cut is motivated by the need to exclude sources that are located on the main sequence turn-off and on the faint end of the giant branch. Fig. 5 shows the color-magnitude diagram of the selection. We note that for $M_{G,0} \sim 7$ mag the binary sequence (black dashed line) and the 20 Myr isochrone (grey dotted line) overlap: thus we expect that region of the color-magnitude diagram to be contaminated by old binaries (see

Section 3.4 for a more detailed discussion). In general, the area of the color-magnitude diagram next to the binary sequence is bound to be subject to contamination from unresolved binaries, but also from reddened main sequence sources: to partially eliminate the issue, we decided to restrict further our sample to the sources brighter (and thus younger) than the 20 Myr isochrone¹. Fig. 6 (top) shows the position in the sky of the sources selected with this procedure. Some groups can be easily identified:

- *Orion*, on the rightmost side at $l < 220^\circ$;
- *Vela*, at $240^\circ < l < 270^\circ$;
- *Scorpius-Centaurus* and *Ophiucus*, at $l > 280^\circ$ and positive b ;
- *Chamaeleon*, at $l, b \sim (300^\circ, -16^\circ)$;
- The *Aquila rift*, at $l, b \sim (30^\circ, +3^\circ)$;
- *Lacerta*, at $\sim (100^\circ, -20^\circ)$;
- *Cepheus* and *Cassiopeia*, at $l > 100^\circ$, above and slightly below the galactic plane;
- *Taurus* and *Perseus*, at $l > 140^\circ$, below the galactic plane.

The source distribution follows the dust features located in the galactic plane: while on the one hand it is expected that young sources follow the outline of the molecular clouds, on the other hand it is likely that our sample is still contaminated by main sequence stars located behind the molecular clouds. Thus, to remove the last contaminants we discarded all the sources with $A_G > 0.92$ mag. We chose this threshold after studying the extinction distribution of our sample: the median of the distribution is 0.51 mag, while the 16th percentile is 0.30 mag and the 84th percentile is 0.92 mag. Thus, we excluded all the sources with extinction larger than the 84th percentile. This is a rough cut, that might exclude not only reddened main sequence sources, but also young sources embedded in the clouds, however it is on average effective in removing contaminants (see also Appendix E). Fig. 6 (centre) shows the distribution in the sky of the sources remaining after the extinction cut.

¹ We also tested whether we would obtain different results by considering, for instance, the luminosity above the main sequence as an age proxy: this was not the case.

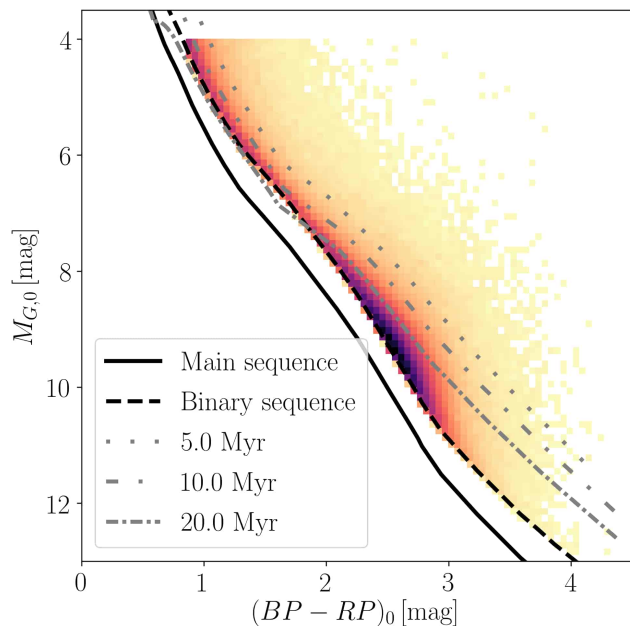


Fig. 5. $G_{BP} - G_{RP}$ vs. M_G colour-magnitude diagram of the sources selected in Section 2.2.2. The density of sources increases towards the binary sequence.

2.3.3. Tangential velocities

As in Sec. 2.2.1, we finally perform a selection in tangential velocity space, relying on the fact that the young clusters and associations that we are interested in share the same kinematic properties. Fig. 7 shows the tangential velocity distribution defined in Eq. 3 of the sources selected in Section 2.2.2. The contour lines represent the $S = 1, 2, 3$ levels. Analogously as with the UMS sample, we selected all the sources within the $S = 1$ level. The final PMS sample is shown in Fig. 6, bottom. As mentioned in the previous Section, the extinction correction reduces the imprint of the molecular clouds on the star distribution. The tangential velocity selection instead mostly reduces the number of sources at high galactic latitudes.

3. Three dimensional mapping of young stars in the solar neighbourhood

In this section we describe the method we use to make three-dimensional density maps of the solar neighbourhood. We make two maps, one for the UMS sample and one for the PMS sample. The maps are then discussed and compared in this Section and in Section 4.

3.1. Method

Similarly to what we did in Section 2.2.3, the first step of creating the maps is to compute galactic Cartesian coordinates, x, y, z , for all the sources and to define a box $V = 1000 \times 1000 \times 700$ pc centred on the Sun. We divide the cube in volume elements v of $3 \times 3 \times 3$ pc. After computing the number of stars in each volume n , we estimate the star density $D(x, y, z)$ by smoothing the distribution by means of a three dimensional gaussian filter, using a technique similar to that used by Bouy & Alves (2015).

The gaussian width (equal on the three axes) is $w = 3$ pc for PMS stars and $w = 4$ pc for UMS stars, and the gaussian is

truncated at 3σ ²). The choice of a certain w value is arbitrary. A high w value produces a smooth, less detailed map, while a low w value results in a noisy map. We finally normalize the density distribution by applying the *sigmoidal logistic* function:

$$f(\mathbf{x}) = \frac{L}{1 + e^{-k(\mathbf{x}-x_0)}} - 1, \quad (5)$$

where $\mathbf{x} = D(x, y, z)$, where D is the not normalized density distribution. The parameters we chose are: $L = 2, x_0 = 0, k = 30$ for PMS stars; $L = 2, x_0 = 0, k = 40$ for UMS stars. In this way, $f(\mathbf{x})$ ranges from 0 to 1 as x ranges between 0 and infinity. A low k value reveals more detail at higher densities and a high k value reveals more detail at lower densities. The choice of the appropriate gaussian w value and logistic k value depends upon the desired map presentation. We have chosen the best values to visualize stellar concentrations for the UMS and PMS maps.

3.2. Results

Fig. 8 (left) shows the density distribution of pre-main sequence sources younger than 20 Myr on the galactic plane (X is directed towards the galactic Centre, Y towards galactic rotation, the Sun is at $(0, 0, 0)$). Fig. 8 (right) shows the density distribution perpendicular to the plane. Fig. 9 shows the density distribution of the UMS sample. The axes are the same as in Fig. 8.

Three main density enhancements, visible in both maps are:

1. *Scorpius-Centaurus (Sco OB2)*: $0 < X < 250$ pc and $-200 < Y < 0$ pc. Due to its proximity ($d \sim 140$ pc, de Zeeuw et al. 1999), the Sco OB2 has been extensively studied (de Bruijne 1999b; Rizzuto et al. 2011; Pecaut et al. 2012; Wright & Mamajek 2018). The association is usually divided in three subgroups, *Upper Scorpius (US)*, *Upper Centaurus-Lupus (UCL)*, and *Lower Centaurus-Crux (LCC)*, with median ages of 11, 16, and 17 Myr (Pecaut & Mamajek 2016).
2. *Vela (Vel OB2)*: $-100 < X < 100$ and $-100 < Y < -450$ pc. Vel OB2 has a distance of $d \sim 410$ pc. Sacco et al. (2015); Jeffries et al. (2014) and Damiani et al. (2017) studied the stellar population towards the Gamma Vel cluster and NGC 2547, finding kinematically distinct populations. By using respectively *Gaia* DR1 and *Gaia* DR2, Armstrong et al. (2018) and Beccari et al. (2018) recently found that the association is composed of many young clusters. In particular Beccari et al. (2018) discovered four new clusters, in addition to Gamma Vel and NGC 2547; four of these clusters are coeval and formed ~ 10 Myr ago, while NGC 2547 and a newly discovered cluster formed ~ 30 Myr ago. Cantat-Gaudin et al. (2018) also characterized the distribution of Vel OB2 on a large spatial scale, and found that the distribution of young stars traces the IRAS Vela Shell. This might suggest a common history for Vel OB2 and the Vela Shell: a previous star formation event caused the expansion of the shell and likely triggered the formation of the clusters composing the association.
3. *Orion (Ori OB1)*: $-300 < X < -200$ and $-200 < Y < -100$ pc. Orion is the nearest ($d \sim 400$ pc) giant molecular cloud complex and it is a site of active star formation, including high mass stars (e.g. Bally 2008, and references therein). Zari et al. (2017) used *Gaia* DR1 to explore the arrangement and

² The python function used for the smoothing is `scipy.ndimage.filters.gaussian_filter()`

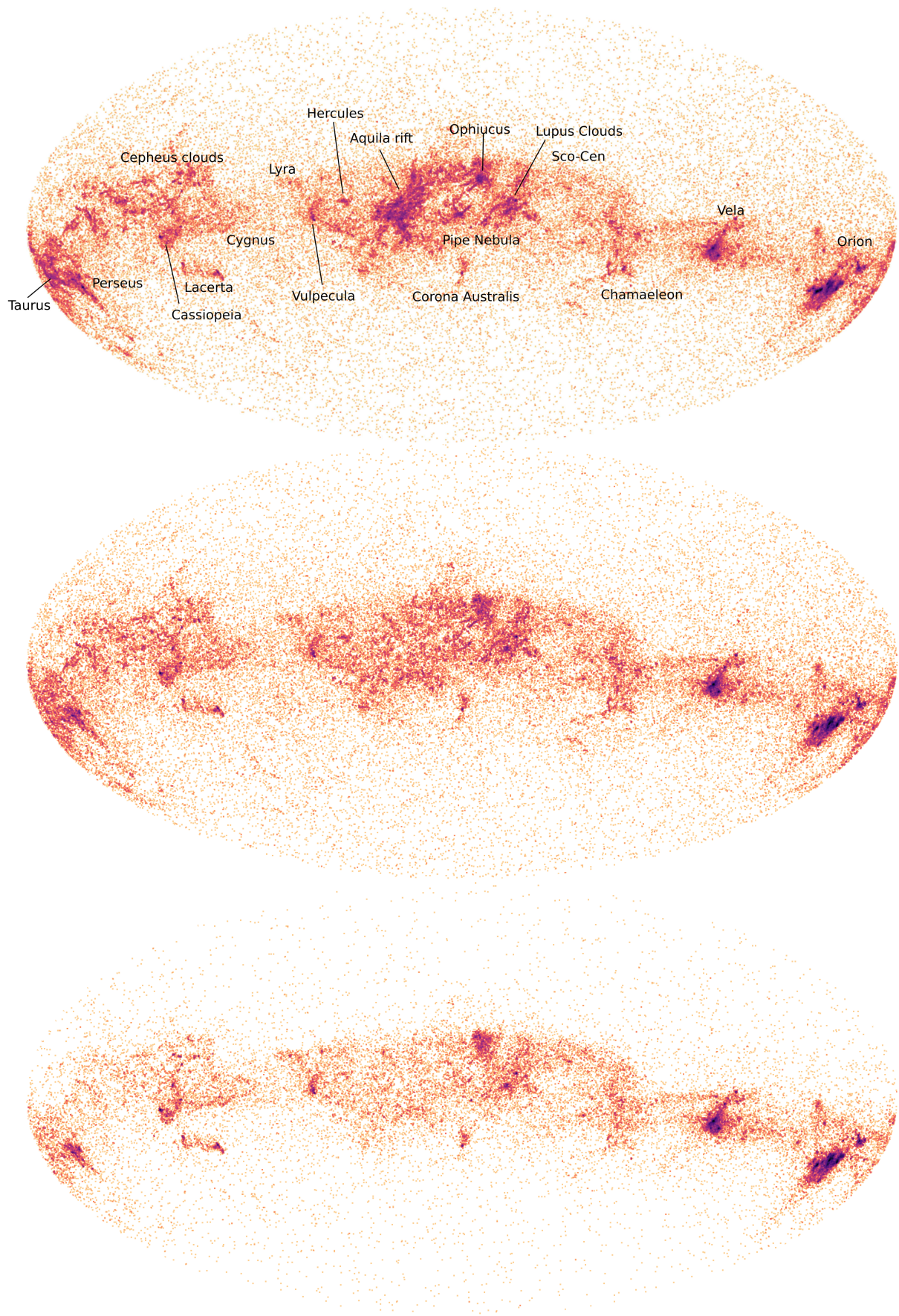


Fig. 6. Top: pre-main sequence sources younger than 20 Myr. Centre: pre-main sequence sources younger than 20 Myr, with $A_G < 0.92$ mag. Bottom: pre-main sequence sources younger than 20 Myr, with $A_G < 0.92$ mag, and within the $S = 1$ level of Fig. 7.

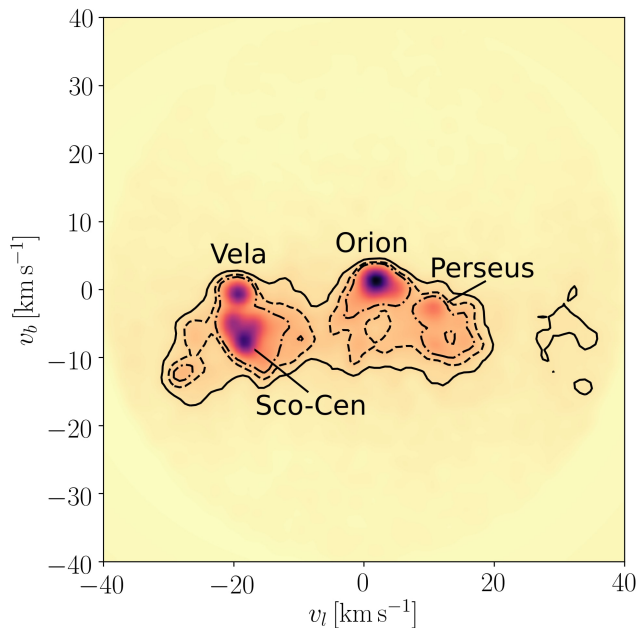


Fig. 7. Tangential velocity plot of the PMS sample selected in Section 2. Many clumps are visible and correspond to known associations and clusters. The four most prominent structures are: Orion, Sco-Cen, Vela and Perseus. Note the gap around 20 km s^{-1} , visible also in Fig. 11.

the age ordering of the numerous stellar groups towards the Orion OB association. Kounkel et al. (2018) used *Gaia* DR2 and APOGEE-2 to identify spatially and kinematically distinct young groups.

The pre-main sequence population of Sco OB2, Vel OB2, and Ori OB1 is predominantly concentrated in the dense areas of the upper main-sequence population. The latter appears instead more diffuse, almost connecting the three regions. A few, more evolved clusters are also visible in Fig. 9: IC 2602, IC 2391, NGC 2451, NGC 2516, NGC 3532, NGC 2422, NGC 6475, NGC 6405, IC 4756, NGC 6633, NGC 7092, Stock 2, α Per, and Pleiades. Some of these clusters appear embedded in the low density levels of the UMS density distribution: this might suggest a relation between current star forming regions and previous star formation episodes. Finally, it is particularly interesting to notice the presence of a diffuse population in front of the Orion complex (visible in both the UMS map of Fig. 9 and the PMS map of Fig. 8). This population was already observed by Bouy & Alves (2015); Zari et al. (2017) and Kounkel et al. (2018), and here we confirm those findings. Further, we would like to draw some attention to the little cluster at $(x, y) \sim (-250, -250) \text{ pc}$ ($l, b \sim 218.5^\circ, -2^\circ$) of Fig. 8. A preliminary inspection of the proper motion and the colour-magnitude diagram (see Appendix C) indicates that this is probably an open cluster, previously unidentified (to the knowledge of the authors) due to its vicinity to the galactic plane. The presence of a new open cluster next, and possibly related, to the Orion star forming region, adds a new piece to the puzzle of the star formation history of Orion.

Some density enhancements are visible only or mostly in the PMS map. This is because those are low or intermediate mass star forming regions, with very few early type stars.

1. *Taurus and Perseus (Per OB2)*: $x - 300 < x < -50$ and $0 < y < 100 \text{ pc}$. Taurus (Kenyon et al. 1994; Scelsi et al. 2007) lacks massive OB-type stars and has therefore become

a prototype to study low-mass star formation processes. Belikov et al. (2002b,a) studied an area of $\sim 20^\circ$ diameter centred on the Perseus OB association, identifying over 800 members by their common proper motion and distances. Surprisingly, even harbouring one of the major associations in the solar vicinity (de Zeeuw et al. 1999; Bally et al. 2008), Per OB2 is only barely visible in the UMS map of Fig. 9, probably because of the lower number of massive stars it contains with respect to Orion, Vela, and Sco-Cen.

2. *Cepheus, Cassiopeia, and Lacerta (Lac OB1)*: $-200 < x < -50$ and $250 < y < 500 \text{ pc}$. Cepheus contains several giant star forming molecular complexes, located at various distances (Kun et al. 2008). According to their distance they can be arranged in different subgroups: at $d < 500 \text{ pc}$ there are the clouds located in the Cepheus flare (see Fig. 2 in Kun et al. 2008), while the associations Cep OB2, Cep OB3 and Cep OB4 (de Zeeuw et al. 1999) are located between 600 and 900 pc, and therefore beyond the boundaries of our region. The groups in Fig. 8 are associated to the Cepheus flare and follow closely the gas structures. Lac OB1 is an association in its final stage of star formation (Chen & Lee 2008). The groups that we identified in our maps are: LBN 437 (also known as Gal 96-15) and Gal 110-13. These are the only regions with recent star formation activities. Cassiopeia contains a few nearby star forming molecular clouds (Kun et al. 2008). In the maps it is possible to identify a group related to LkH α 198 and associated with the dark cloud L 1265, plus other small cluster in the same area.

3. *Aquila*: $x > 100$ and $50 < y < 200 \text{ pc}$.

A few density enhancements are visible towards the Aquila Rift. In general they follow the dust structures, with some small clumps. The density enhancements are not related to the open clusters identified in the UMS map, as the estimated ages of those are older than 20 Myr. We therefore conclude that stars in that region of the PMS map are mainly main sequence contaminants that survive the selection process or older PMS sources.

A peculiar region is that of *Lyra and Cygnus*: $0 < x < 200$ and $250 < y < 500$.

Lyra is predominantly visible in Fig. 8, while Cygnus is visible in both Fig. 8 and 9, although the density enhancements have a slight offset. The reason of these differences might be due to the way we select the samples: indeed, we select density enhancements in tangential velocities and we then study their density in space, therefore some groups might get lost in the process, especially if they do not stand out significantly with respect to the background. This is further discussed in Section 4. We note here that Cyg OB4 and Cyg OB7 de Zeeuw et al. (1999) are beyond the region studied in this work ($d > 500 \text{ pc}$). The density enhancements we find lie towards the 'Northern Coalsack', towards the Cygnus constellation, and towards the δ Lyra cluster. As for Sco OB2, Vel OB2, and Ori OB1, the UMS star distribution is broader than the PMS distribution, and seems to connect different groups. Note that, towards the same line of sight, two open clusters are present: Roslund 6 (Roslund 1960) and Stock 1 (Osborn et al. 2002). However, they are both too evolved (their age is around 300 Myr) to appear in the PMS maps.

By comparing the map contour levels at lower densities, we further notice that the overall star distribution presents some differences. In particular, the PMS distribution shows a clear gap in the region surrounding the Sun. This is not unexpected, as in the innermost 50–100 pc groups younger than 20 Myr are not present. In the same area the UMS distribution looks instead smoother,

even though the area surrounding the Sun does not contain dense clumps in the distribution (which is consistent with the PMS distribution). This is further discussed in the Section 4. The overall source distribution in the X, Z plane appears inclined with respect to the galactic plane, however the tilt is dominated by Sco OB2 and Ori OB1. Again, this is further discussed in Section 4. Finally, we note that the maps might look different because different values of w and k were used, however the main features that we described above remain visible for different k and w parameters.

3.3. Ages of the PMS sample

We now study the ages of the PMS sample selected in Section 3.3. During the pre-main sequence, younger stars are also brighter. For this reason it is quite straightforward to infer age gradients by studying colour-magnitude diagrams of pre-main sequence sources.

Following the procedure outlined in Section 3.1, we made density maps of the PMS stars, dividing them according to their position in the colour magnitude diagram. We divided the PMS sample in three sub-samples, according to the age (τ) suggested by the PARSEC isochrones:

1. $\tau \leq 5$ Myr;
2. $5 \leq \tau \leq 10$ Myr;
3. $10 \leq \tau \leq 20$ Myr;

Fig. 10 shows the density distribution of stars ≤ 5 Myr (red), ≤ 10 Myr (green), ≤ 20 Myr (blue). Not unexpectedly the older population is also more dispersed, while younger sources are tightly clustered. The age gradient observed in Sco-Cen by many authors (e.g., Pecaut & Mamajek 2016) is evident. In Vela, some young clumps are present, however on average the population is older than in the Orion region. This is not unexpected, as Jeffries et al. (2009) find an age of ~ 10 Myr for the PMS population in Vela. In Perseus, the young cluster IC 348 is visible. The red cluster in $(X, Y) \sim -30, 0$ pc belongs to the Taurus star forming region. The groups at large positive Y values are instead more evolved.

3.4. Caveats

By performing the source selection that we described in Section 2, we applied different cuts to the data (photometric and astrometric) to clean our sample. In this paper we do not attempt to estimate the purity nor the completeness of the catalogue. The users can make stricter selections based on tangential velocity to obtain a purer sample, at the expense of completeness.

Through extinction mapping we corrected the observed colour-magnitude diagrams and we excluded extincted main sequence sources that contaminated our sample. On one hand, this procedure is necessary to obtain maps that truly trace the distribution of young sources in the solar neighbourhood. On the other hand, the maps might be affected by selection biases introduced by creating the sample, in particular the truncation on relative parallax uncertainty and the application of the extinction correction.

Relative parallax uncertainty. Selecting sources through their relative parallax uncertainty has at least two effects.

- The ecliptic poles ($|b| > 45^\circ$) are preferred in terms of number of sources due to *Gaia*'s scanning law. This implies that by selecting sources through their relative parallax errors, there might be a 'fake' over-density of sources towards the

ecliptic poles (see Appendix B). The effect of that would be an over-density in the 3D maps corresponding to those areas or, analogously, an under-density in the other areas. A possible signature of this selection bias might be found in the shape of the low-density contour of the $X - Z$ projection of the PMS distribution (Fig. 8, right): the density does not look as a uniform slab (compare with the UMS distribution of Fig. 9, right) but presents peculiar 'cavities' along Z . This bias - if present - influences the low-density levels and the global source distribution of the maps but not the compact groups that we focus on in this study.

- Parallax uncertainties in *Gaia* DR2 increase as a function of increasing G (Gaia Collaboration et al. 2018b). Thus, faint sources at large distances are more easily excluded by the parallax uncertainty selection. This makes our sample incomplete for faint G values. The (in)completeness level is a function of distance (for fixed G): for example, a star with $G = 21$ mag and parallax error $\sigma_\varpi \sim 1$ mas (see Fig. 7 in Gaia Collaboration et al. 2018b), would be considered part of our sample until $\varpi = 5$ mas ($d = 200$ pc) and excluded for smaller parallaxes ($d > 200$ pc). While the completeness of the sample needs to be thoroughly analysed when studying the properties of each star formation region (such as the initial mass function), it should not affect the spatial structures that we observe in the 3D maps.

Extinction correction. While Fig. 2 and 4 show essentially a uniform distribution of sources on the galactic plane, without any evident sign of extinction, Fig. 6 (top) clearly show the outline of nearby molecular clouds. To exclude extincted sources we resolved to eliminate all the PMS sources with $A_G > 0.92$ mag. This cut aims at excluding background, heavily extincted stars, however in practice it removes also young stellar objects still embedded in their parental molecular clouds, or actual pre-main sequence stars that lie behind a dense cloud (e.g. potential young groups behind the Aquila rift). By comparing the maps of Fig. 8 and E.1 (where in the latter the condition $A_G < 0.92$ mag is not applied), we notice substantially the same main density enhancements (see Section 3.2 and Appendix E for more details), thus we conclude that the extinction correction that we are applying is satisfactory for our PMS sample, but should not be applied blindly

4. Discussion

In the previous sections, we analysed the spatial distribution and the age ordering of young stellar population within $d = 500$ pc from the Sun. In this section, we put our findings in the context of the star formation history of the solar neighbourhood.

The Gould Belt's definition varies from author to author. It is however striking how we do not find any evidence of a belt-like structure, neither for the PMS sample, nor for the UMS sample. The tilt observed with respect to the galactic plane is dominated by Ori OB1 and Sco OB2, which are below and above the galactic plane respectively. This is particularly evident by the X vs. Z projections of Fig. 8 and 10. As Bouy & Alves (2015) proposed, the existence of a belt of star forming regions gives a poor description of the spatial distribution of the stars revealed by our analysis, calling for a new interpretation of the distribution of stellar groups in the solar neighbourhood. Referring to the UMS distribution, we confirm the presence of three large structures, Scorpius-Centaurus, Vela and Orion, hundreds of parsecs long, which Bouy & Alves (2015) identified and called 'blue-streams'. The distribution of the pre-main sequence stars follows closely

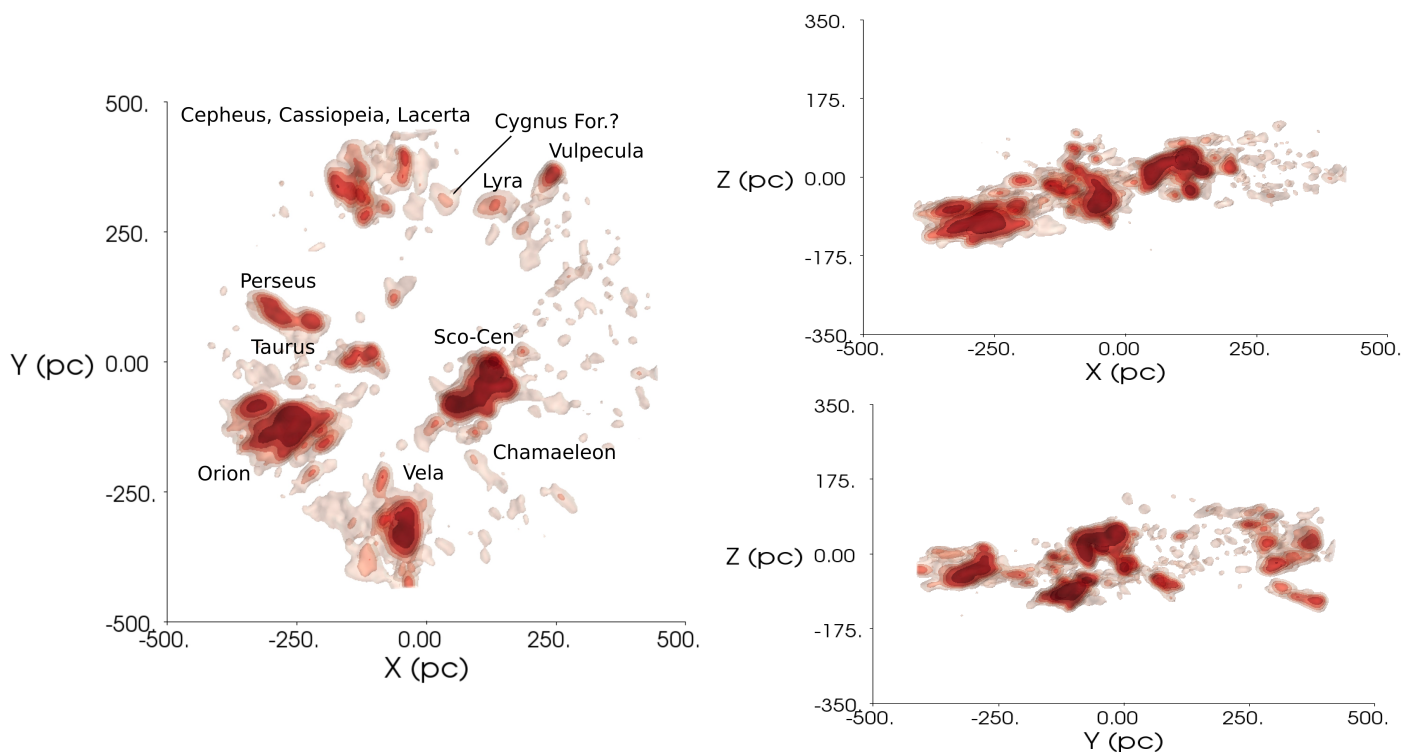


Fig. 8. Left: 3D density distribution of PMS sources younger than 20 Myr on the galactic plane. The Sun is in (0, 0), the x-axis is directed towards the galactic centre, and the y-axis towards the direction of the galactic rotation. The z-axis is perpendicular to the plane. The contours represent the 0.2, 0.4, 0.6, 0.8, and 1 density levels. Right, top: 3D density distribution of the PMS sample (age < 20 Myr) perpendicular to the galactic plane. Contour levels are the same as on the left. Right, bottom: 3D density distribution of the PMS sample (age < 20 Myr) along the rotation axis.

the OB distribution and defines the dense and young regions of the blue-streams. By using *Gaia* DR2 data, we extend Bouy & Alves (2015) study to include the regions at positive Y values in the maps. Perseus and some clusters in Taurus, as well as Lacerta and Cepheus, are well visible in our PMS and UMS maps and were not identified by Bouy & Alves (2015), probably because they do not host a large number of early type stars. The distribution shown in the maps present some differences: for example, some density enhancements are prominent in only one map. As discussed in Section 3.2, the UMS map shows many open clusters that do not appear in the PMS map because they are older than 20 Myr. In the region corresponding to Taurus we do not observe any density enhancement in the UMS map, as Taurus lacks early-type stars.

To further confirm that the main structures that we identify in the PMS map actually correspond to those in the UMS map, we study the groups in a parameter space that we have not used yet. Fig. 11 shows the tangential velocities along galactic latitude of the UMS (top) and the PMS sample (bottom) older than 20 Myr, before (left) and after (right) the tangential velocity selection of Section 2.2.4. The solid orange line shows the projection of the solar motion ($U_{\odot}, V_{\odot}, W_{\odot} = (11.1, 12.24, 7.25) \text{ km s}^{-1}$, Schönrich et al. 2010). The location of the groups in the v_t vs. l plane is primarily due to the projection of the solar motion in different directions. The deviations from the solar motion are due to the peculiar motions of the star forming regions. Clumps and elongated structures are visible, corresponding to the groups mentioned in Section 3. The features in the PMS panels correspond to those in the UMS panels, although in the latter they are less well defined. Indeed, PMS groups have a smaller velocity dispersion than UMS sources. This agrees with the fact that PMS groups are clustered in denser structures in the 3D maps. Fur-

ther, by definition, the UMS sample contains also more evolved sources, which are expected to have a larger velocity dispersion. The reason of the discrepancies in the maps might thus be due to the density contrast of different groups. Indeed the stellar population of some groups is more abundant (such as in Sco OB2 or Ori OB1), and/or more compact (in the case of the open clusters observed in the UMS distribution): the density will peak in these regions, making them stand out more than others. Fig. 11 also shows that the tangential velocity selection is useful to exclude a large number of contaminants, but that still retains a good number of spurious sources. Note that the gap visible especially in the right, bottom panel of the Fig. 11 is due to the tangential velocity selection. One of the goals of this work is to provide catalogues of PMS and UMS sources, which can be used for future works on the global properties of solar neighbourhood or on specific star forming regions. We decided to not impose stricter criteria on our selection to avoid as much as possible to exclude interesting sources. On the other hand, this means that future users should be careful when using the data, and should combine spatial, kinematic and photometric data to select accurately the stellar population of one region.

The most apparent difference in the 3D maps involves the global source distribution. As already noted in Section 3, PMS stars show a gap in their distribution in the inner $\sim 50 \text{ pc}$. This is not unexpected as vicinity of the Sun ($d < 50 \text{ pc}$) is essentially free of stars younger than 20 Myr, except for a few small groups that are difficult to pick up on our maps (e.g. the β Pictoris moving group). On the contrary the distribution of UMS sources looks uniform, with a small under-density next to Sun that loosely traces the gap observed for the PMS distribution. The fact that the density of early-type stars decreases in the solar vicinity is consistent with the PMS distribution. The distribution

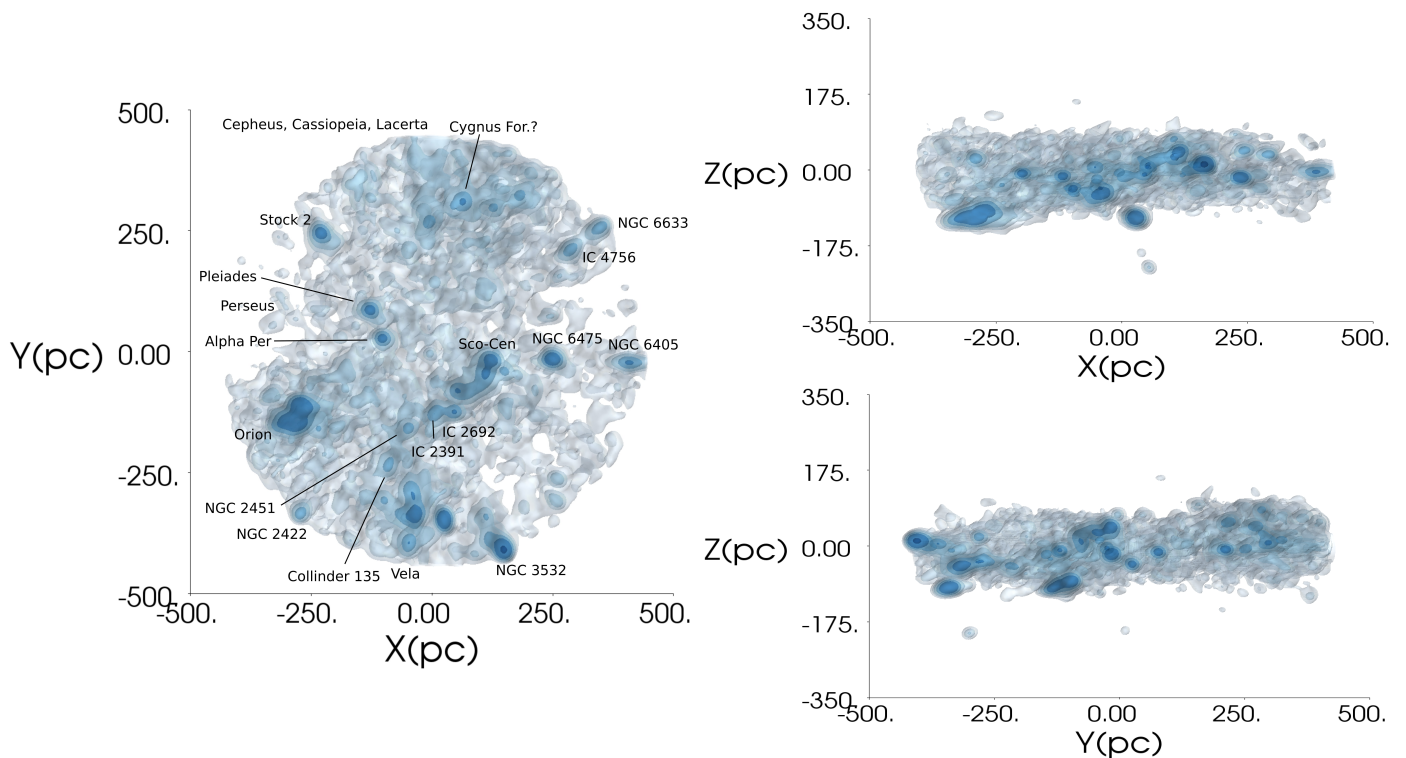


Fig. 9. Same as Fig. 8, but for the upper-main sequence sample selected in Section 2.1. The contours represent the 0.2, 0.3, 0.4, 0.6, 0.8, and 1 density levels.

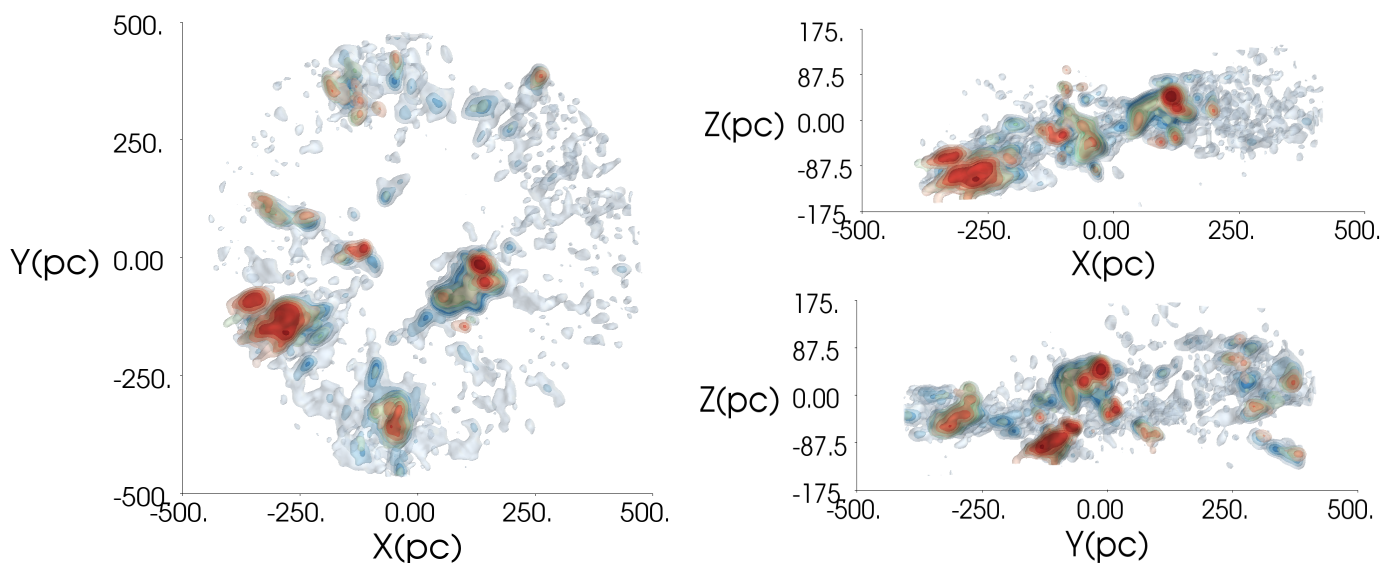


Fig. 10. 3D maps of sources younger than 20 Myr and older than 10 Myr (blue), younger than 10 Myr and older than 5 Myr (green), and younger than 5 Myr (red). The contours are the same as in Figs. 8. In Fig. B.1 we show separate maps of the $X - Y$ plane for each of the age intervals.

is however more uniform for two reasons: the first is related to the smoothing parameters that we used to create the map. Since the number of early-type sources is smaller than that of pre-main sequence stars, we had to use a larger value of σ to smooth the density distribution (see Section 3.1). The second is related to the age of early-type stars. As we already mentioned above, the UMS consists also of stars whose age is larger than 20 Myr because of the way we selected the sample. For this reason the distribution of the UMS sample is intrinsically more spread out than that of the PMS sample.

The age map of Fig. 7 suggests that multiple star formation episodes can occur within the same region and give limits on the duration of a single star formation episode. We notice that a global trend between the different star forming groups is not present, and that, within each group, older and younger stars are spatially mixed. This is also visible in Fig. 12, which shows the same sources as in Fig. 10, projected in the sky (older to younger from top to bottom). Younger stars are clustered in denser clumps, usually surrounded by the older, more diffuse population. Note that in our age maps we do not take binarity into account. As discussed in Zari et al. (2017), unresolved bi-

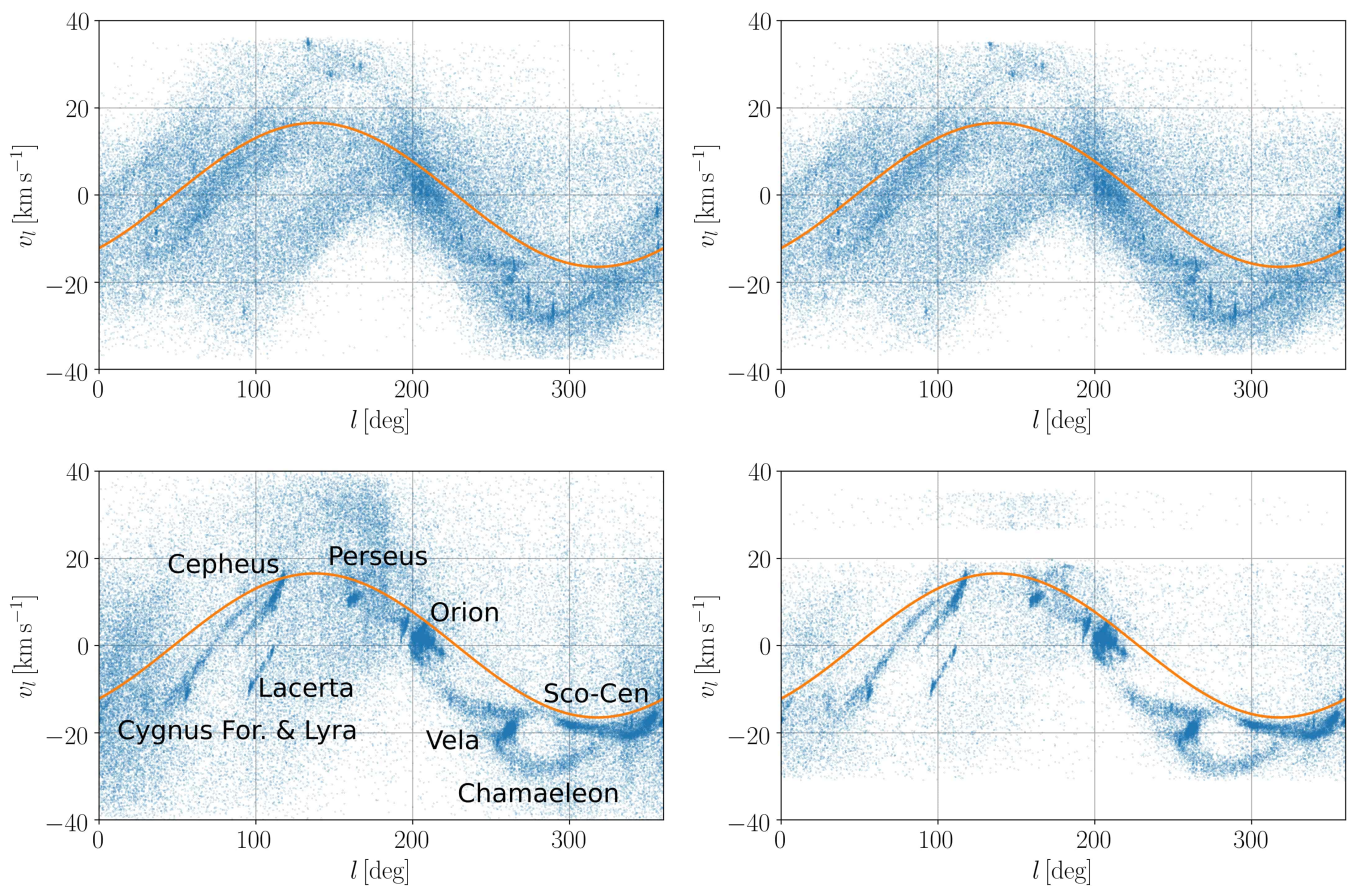


Fig. 11. Tangential velocity along galactic latitude vs. latitude for the UMS (top) and the PMS (bottom) sample, before (left) and after (right) the tangential velocity selection. The solid orange line shows the projection of the Sun motion. The ‘gaps’ in the scatter plots on the left are due to the tangential velocity selection (see Section 2 in the text).

naries stand out as a separate sequence, which, being brighter by ~ 0.75 mag with respect to the main sequence, might look like a younger population. This is a major cause of age spreads, and could affect absolute age estimates. However, binarity should affect our data in the same way in all directions and distances, making relative age estimates quite robust. In fact, significant age spreads have been observed in young clusters. Da Rio et al. (2012) observed an age spread as large as 10 Myr in the Orion Nebula Cluster (ONC). More recently, Beccari et al. (2017) reported three separated pre-main sequences towards the ONC, indicative of three different episodes of star formation, each separated by about a million year. Kroupa et al. (2018) explained such observation by outlining a scenario where subsequent burst of star formation are regulated by stellar feedback and dynamical ejections of high mass stars. According to this scenario, after the first episode of star formation, the newly formed stars ionise and suppress star formation in the embedded cluster. However, high mass stars are soon ejected from the cluster, thus gas inflow can resume. This sequence of events can be repeated until the maximum lifetime of a molecular cloud (around 10 Myr) is reached. Albeit with some stretch of the imagination (the groups we observe in the maps are more extended than the ONC, and the over-densities could encompass more than one cluster), this scenario might explain also our observations: indeed younger groups occupy in general the central regions of the density enhancements and are surrounded by a more diffuse population.

The age map also shows age gradients. In Sco OB2, the youngest

groups correspond to Upper Scorpius, while Upper Centaurus Lupus and Lower Centaurus Crux (see also Pecaut & Mamajek 2016) appear older. In Fig. 12 we observe a density enhancement at coordinates $l, b \sim 343^\circ, +5^\circ$: this cluster has been reported by Röser et al. (2018); Villa Vélez et al. (2018) and Damiani et al. (2018) and is traditionally not within the boundaries of Sco OB2. We confirm that given its distance and age, the cluster is likely related to the association. Krause et al. (2018) combined gas observations and hydrodynamical simulations to study the formation of the Scorpius-Centaurus super-bubble, and suggest a refined scenario for the evolution of the OB association. Dense gas is originally distributed in an elongated cloud, which occupies the current area of the association. The star formation events in UCL and in LCC give origin to super-bubbles that expand, surrounding and compressing the parental molecular cloud, triggering star formation in US. This scenario predicts the formation of kinematically coherent sub-groups within the associations that move in different directions, which is similar to the observed kinematics in Sco-Cen (Wright & Mamajek 2018). Krause et al. (2018) predict also that young groups could occur also in regions of older stars, and that several young groups with similar ages might form over large scales. This is consistent with what we observe, not only in Sco-Cen, but also in the other groups. In the Orion region, old stars appear to cluster on the sides and in front of the young population (see Fig. 11). The candidate open cluster at $l, b \sim 220^\circ, -2^\circ$, $X, Y \sim (-250, -250)$ pc, has an age > 10 Myr and might be related to the Orion dust ring dis-

covered by Schlafly et al. (2015). Cantat-Gaudin et al. (2018) found that young stars in Vel OB2 trace the gas and dust features of the IRAS Vela Shell and proposed that intense supernova activity coming from the Trumpler 10 and NGC 2451B released enough energy to create a cavity and power the expansion of the IRAS Vela Shell, which subsequently produced a shock in the interstellar medium, which then triggered a second burst of star formation. This agrees roughly with what shown in Fig. 12: young stars in the central panel appear slightly more concentrated on the area corresponding to the shell than older stars in the top panel. This should be however further investigated, as Fig. 11 shows an overlap of the sources in the three different age intervals. The star forming regions at positive Y values appear in general more evolved, and their stellar content is less numerous than that of the groups that we have discussed above. However as they are located towards well known and rich star forming regions, such as the Cepheus and Cygnus OB associations, they might be the extremities that lie closer to the Sun of those groups. This should be further investigated by extending the map towards further distances, but it is beyond the scope of this paper.

Finally, we consider the PMS sources that, according to the isochrones in Fig. 5, are older than 20 Myr and we select them using the same method outlined in Section 2.2.2 and 2.2.3. The spatial distribution of the sources is shown in Fig. 13. The density map presents many interesting features. First, we note that the Orion young population has completely disappeared from the map, while the evolved clusters on its sides are still visible. The Vela and Scorpius-Centaurus populations are still traced by the density distribution, although the density levels appear broader than in the maps of Fig. 8. At positive Y values, the sources related to Cassiopeia, Cepheus, and Chamaeleon are barely visible, however those in the Cygnus foreground and related to the Lyra open cluster are present. This suggests that these regions are quite evolved, and raises some doubts on the connection of the Cygnus foreground to the Cygnus associations. The global source distribution is very similar to that presented in the UMS map (Fig. 9). The region surrounding the Sun presents a lack of sources, which is however less pronounced than in the PMS map of Fig. 8. This represents additional evidence that there is a real gap for the youngest stars, extending out to ~ 100 pc towards Scorpius-Centaurus and reaching ~ 200 pc towards Cygnus and, in the opposite direction, towards Vela and Orion. The gap could thus be a consequence of any star forming gas having been cleared out 20 – 30 Myr ago due to the events that created the Local Bubble (Alves et al. 2018; Lallement et al. 2014; Puspitarini et al. 2014).

5. Conclusion

We used *Gaia* DR2 to study the three dimensional configuration of early-type, upper-main sequence (UMS) and pre-main sequence (PMS) stars in the solar neighbourhood, within $d = 500$ pc from the Sun.

- We select the data through a combination of astrometric and photometric criteria. A side product of the data selection procedure is a three dimensional G -band extinction map which we use to correct our data for extinction and reddening. The final UMS and PMS samples are available on-line.
- By using a gaussian filter smoothing technique, we create 3D density maps for both the UMS and the PMS samples.
- The PMS map (Fig. 8) of the sources younger than 20 Myr shows a gap in the innermost 50 – 100 pc. This is due to

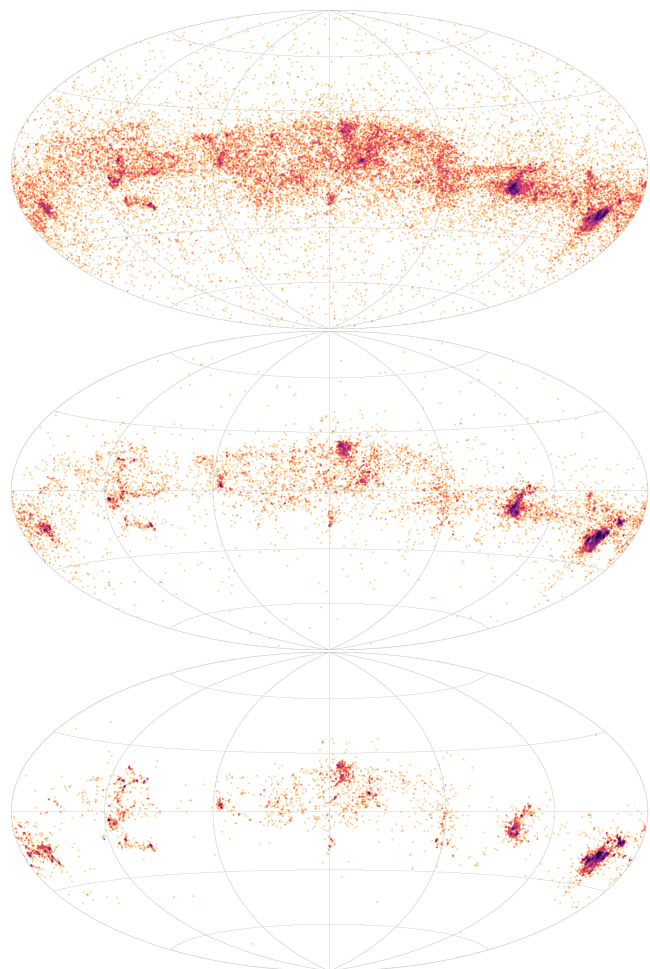


Fig. 12. Sky projection of sources with different ages. Top: sources with $10 < t < 20$ Myr; centre: sources with $5 < t < 10$ Myr; bottom: sources with $t < 5$ Myr.

the absence of young (with age < 20 Myr) groups in the vicinity of the Sun. The same gap appears also in the UMS distribution (Fig. 9), although not as clearly. Due to the way it is constructed, the UMS sample contains indeed also sources older than 20 Myr. This has two effects:

1. the low-density distribution appears smoother;
 2. more evolved open clusters are visible.
- Three structures are recognizable in both the maps of Fig. 8 and 9: Scorpius-Centaurus, Vela, and Orion. The PMS distribution in this regions follows the distribution of the UMS sources, and defines its dense, inner regions.
 - Taurus, Perseus, Lacerta, Cassiopeia, and Cepheus emerge clearly in the PMS map. Taurus does not host any young, massive source, therefore it is not visible in the UMS map. Perseus, Lacerta, Cassiopeia, and Cepheus are instead visible as low-level density enhancements.
 - A peculiar density enhancement is that in the foreground of Cyg OB4 and Cyg OB7: the enhancement is present in both maps, even if with a slight off-set. We exclude that the PMS density enhancement are related to the open clusters Stock 1 and Roslund 6, as their estimated age is much older 20 Myr. The groups in the foreground of the Cygnus (and Cepheus) associations might therefore represent their extremities that are closer to the Sun.
 - We report the discovery of a young cluster at coordinates $l, b \sim 220^\circ, -2^\circ$. Due to its position, distance, and age, this cluster might be related to the Orion star forming complex.

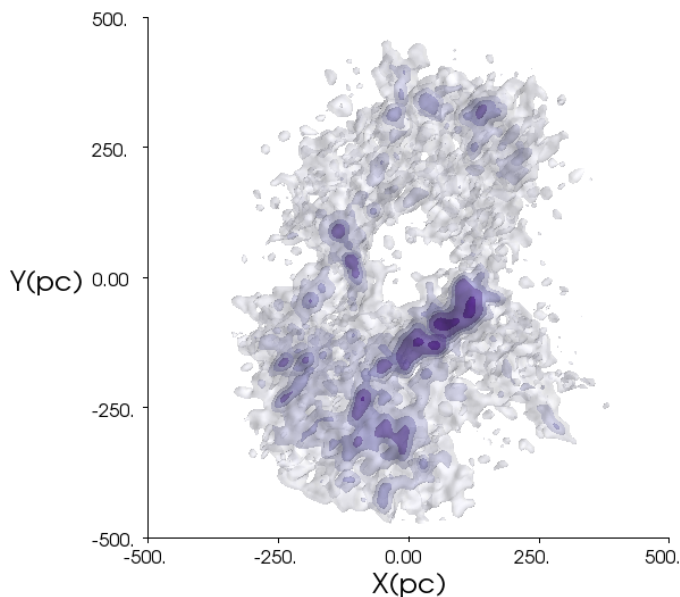


Fig. 13. 3D map of sources older than 20 Myr. The contours represent the 0.2, 0.4, 0.6, 0.8, and 1 density levels.

- We divide the PMS sources in three sub-sets, corresponding to different age ranges (< 5 Myr, $5 < t < 10$ Myr, $10 < t < 20$ Myr), which we compute by using the PARSEC isochrones. We find that sources in the youngest age sub-sets are more concentrated in space, while those in the oldest age sub-sets are globally more diffuse. Age gradients are visible in many regions, particularly in Scorpius-Centaurus, while in others, such as Vela, stars with different ages appear to overlap in space.
- We study the spatial density distribution of the PMS sources older than 20 Myr. At low densities, the density distribution appears similar to the UMS density distribution. The young stellar populations in Orion, Perseus, Cassiopeia, Cepheus, and Chamaeleon are not visible in the map, while Vela and Scorpius-Centaurus are traced by broad density enhancements. At positive Y values, the map shows over-density related to Lyra and to the Cygnus foreground: this implies that those groups are quite evolved and puts into questions the relation of the Cygnus foreground to the Cygnus associations.

In conclusion, we find that the three dimensional configuration of the star forming regions in the solar neighbourhood is far from being described by a ring-like structure such as the Gould Belt, but it is complex and filamentary. A detailed analysis is required to precisely order all the star forming regions according to their ages. In future work we will combine *Gaia* data and other spectroscopic surveys to analyse the kinematic properties of the young stars in the Solar Neighbourhood, which here we have only touched upon. The study of the kinematics and internal velocity patterns (such as expansion or contraction) of the concentrations of young stars will provide deeper insights into their origin.

Acknowledgements. We thank the referee for their constructive comments, which improved the quality of this manuscript. This project was developed in part at the 2018 NYC *Gaia* Sprint, hosted by the Center for Computational Astrophysics at the Simons Foundation in New York City.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC; <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the

Gaia Multilateral Agreement. This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration, 2013). This work has made extensive use of IPython (Pérez & Granger 2007), Matplotlib (Hunter 2007), astroML (Vanderplas et al. 2012), scikit-learn (Pedregosa et al. 2011), and TOPCAT (Taylor 2005, <http://www.star.bris.ac.uk/~mbt/topcat/>). This work would have not been possible without the countless hours put in by members of the open-source community all around the world.

References

- Alves, M. I. R., Boulanger, F., Ferrière, K., & Montier, L. 2018, *A&A*, 611, L5
 Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, *A&A*, 616, A8
 Armstrong, J. J., Wright, N. J., & Jeffries, R. D. 2018, *MNRAS*, 480, L121
 Bailer-Jones, C. A. L. 2015, *PASP*, 127, 994
 Bally, J. 2008, Overview of the Orion Complex, ed. B. Reipurth, 459
 Bally, J., Walawender, J., Johnstone, D., Kirk, H., & Goodman, A. 2008, The Perseus Cloud, ed. B. Reipurth, 308
 Beccari, G., Boffin, H. M. J., Jerabkova, T., et al. 2018, *ArXiv e-prints* [arXiv:1807.07073]
 Beccari, G., Petr-Gotzens, M. G., Boffin, H. M. J., et al. 2017, *A&A*, 604, A22
 Bekki, K. 2009, *MNRAS*, 398, L36
 Belikov, A. N., Kharchenko, N. V., Piskunov, A. E., Schilbach, E., & Scholz, R.-D. 2002a, *A&A*, 387, 117
 Belikov, A. N., Kharchenko, N. V., Piskunov, A. E., et al. 2002b, *A&A*, 384, 145
 Bensby, T., Feltzing, S., & Oey, M. S. 2014, *A&A*, 562, A71
 Bouy, H. & Alves, J. 2015, *A&A*, 584, A26
 Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
 Cantat-Gaudin, T., Mapelli, M., Balaguer-Núñez, L., et al. 2018, *ArXiv e-prints* [arXiv:1808.00573]
 Chen, W. P. & Lee, H. T. 2008, The Lacerta OB1 Association, ed. B. Reipurth, 124
 Chen, Y., Bressan, A., Girardi, L., et al. 2015, *MNRAS*, 452, 1068
 Chen, Y., Girardi, L., Bressan, A., et al. 2014, *MNRAS*, 444, 2525
 Comeron, F. & Torra, J. 1992, *A&A*, 261, 94
 Comeron, F., Torra, J., & Gomez, A. E. 1998, *A&A*, 330, 975
 Da Rio, N., Robberto, M., Hillenbrand, L. A., Henning, T., & Stassun, K. G. 2012, *ApJ*, 748, 14
 Dame, T. M. 1993, in *American Institute of Physics Conference Series*, Vol. 278, Back to the Galaxy, ed. S. S. Holt & F. Verter, 267–278
 Damiani, F., Prisinzano, L., Jeffries, R. D., et al. 2017, *A&A*, 602, L1
 Damiani, F., Prisinzano, L., Pillitteri, I., Micela, G., & Sciortino, S. 2018, *ArXiv e-prints* [arXiv:1807.11884]
 de Bruijne, J. H. J. 1999a, *MNRAS*, 306, 381
 de Bruijne, J. H. J. 1999b, *MNRAS*, 310, 585
 de Geus, E. J. 1992, *A&A*, 262, 258
 de Zeeuw, P. T., Hoogerwerf, R., de Bruijne, J. H. J., Brown, A. G. A., & Blaauw, A. 1999, *AJ*, 117, 354
 Elias, F., Alfaro, E. J., & Cabrera-Caño, J. 2006a, *AJ*, 132, 1052
 Elias, F., Alfaro, E. J., & Cabrera-Caño, J. 2009, *MNRAS*, 397, 2
 Elias, F., Cabrera-Caño, J., & Alfaro, E. J. 2006b, *AJ*, 131, 2700
 Elmegreen, B. G. 1982, *ApJ*, 253, 655
 Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, *A&A*, 616, A10
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018b, *A&A*, 616, A1
 Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1
 Hoogerwerf, R. & Aguilar, L. A. 1999, *MNRAS*, 306, 394
 Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
 Jeffries, R. D., Jackson, R. J., Cottaar, M., et al. 2014, *A&A*, 563, A94
 Jeffries, R. D., Naylor, T., Walter, F. M., Pozzo, M. P., & Devey, C. R. 2009, *MNRAS*, 393, 538
 Katz, D., Sartoretti, P., Cropper, M., et al. 2018, *ArXiv e-prints* [arXiv:1804.09372]
 Kenyon, S. J., Dobrzycka, D., & Hartmann, L. 1994, *AJ*, 108, 1872
 Kounkel, M., Covey, K., Suárez, G., et al. 2018, *AJ*, 156, 84
 Krause, M. G. H., Burkert, A., Diehl, R., et al. 2018, *ArXiv e-prints* [arXiv:1808.04788]
 Kroupa, P., Jeřábková, T., Dinnbier, F., Beccari, G., & Yan, Z. 2018, *A&A*, 612, A74
 Kun, M., Kiss, Z. T., & Balog, Z. 2008, *Star Forming Regions in Cepheus*, ed. B. Reipurth, 136
 Kutner, M. L., Tucker, K. D., Chin, G., & Thaddeus, P. 1977, *ApJ*, 215, 521
 Lallement, R., Capitanio, L., Ruiz-Dern, L., et al. 2018, *A&A*, 616, A132
 Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, *A&A*, 561, A91
 Lindblad, P. O. 1967, *Bull. Astron. Inst. Netherlands*, 19, 34
 Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, 616, A2
 Olano, C. A. 1982, *A&A*, 112, 195
 Olano, C. A. 2001, *AJ*, 121, 295
 Osborn, W., Sano, Y., & Spalding, R. 2002, *PASP*, 114, 1382

- Pecaut, M. J. & Mamajek, E. E. 2016, MNRAS, 461, 794
- Pecaut, M. J., Mamajek, E. E., & Bubar, E. J. 2012, ApJ, 746, 154
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
- Pérez, F. & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21
- Poggio, E., Drimmel, R., Lattanzi, M. G., et al. 2018, MNRAS[arXiv:1805.03171]
- Poppel, W. 1997, Fund. Cosmic Phys., 18, 1
- Puspitarini, L., Lallement, R., Vergely, J.-L., & Snowden, S. L. 2014, A&A, 566, A13
- Rizzuto, A. C., Ireland, M. J., & Robertson, J. G. 2011, MNRAS, 416, 3108
- Röser, S., Schilbach, E., Goldman, B., et al. 2018, A&A, 614, A81
- Roslund, C. 1960, PASP, 72, 205
- Sacco, G. G., Jeffries, R. D., Randich, S., et al. 2015, A&A, 574, L7
- Sancisi, R., Goss, W. M., Anderson, C., Johansson, L. E. B., & Winnberg, A. 1974, A&A, 35, 445
- Scelsi, L., Maggio, A., Micela, G., et al. 2007, A&A, 468, 405
- Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2015, ApJ, 799, 116
- Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829
- Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287
- Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47–54
- Villa Vélez, J. A., Brown, A. G. A., & Kenworthy, M. A. 2018, Research Notes of the American Astronomical Society, 2, 58
- Wright, N. J. & Mamajek, E. E. 2018, MNRAS, 476, 381
- Zari, E., Brown, A. G. A., de Bruijne, J., Manara, C. F., & de Zeeuw, P. T. 2017, A&A, 608, A148

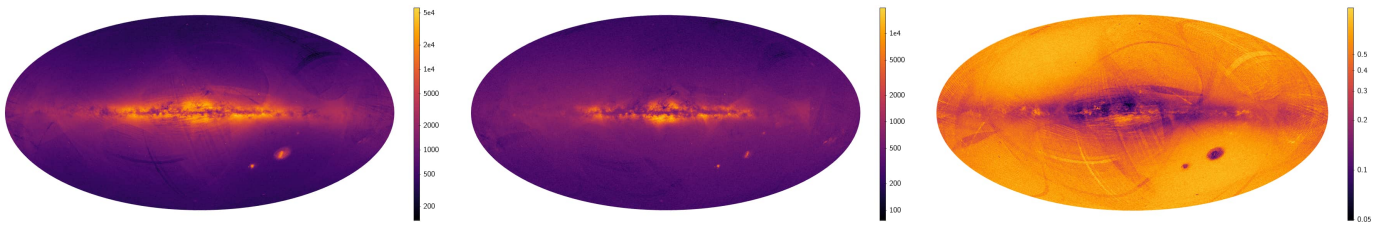


Fig. B.1. Left: Distribution in the sky of the sources within $d = 500$ pc. Centre: Distribution in the sky of the sources within $d = 500$ pc and $\sigma_w/\varpi > 5$. Right: Ratio between the distributions shown in the central and left panels.

Appendix A: ADQL queries

We report here an example of the queries used to select the sources in our field and to perform simple cross-matches.

UMS sample:

```
SELECT *
FROM gaiadr2.gaia_source AS g
WHERE g.parallax_over_error >= 5
AND g.phot_g_mean_mag + 5 * log10(g.parallax) - 10 <= 4.4
AND g.phot_bp_mean_mag - g.phot_rp_mean_mag <= 1.7
AND g.parallax >= 2.
```

PMS sample:

It is impossible to download all the entries of the catalogue for sources with $\varpi > 2$ mas, therefore it is necessary to use multiple queries (for example like the one below) and join the tables afterwards. We also recommend to create an account on the *Gaia* archive.

```
SELECT source_id, l, b, parallax, parallax_error, pmra, pmdec, radial_velocity, pmra_error,
pmdec_error, radial_velocity_error, phot_g_mean_mag, phot_bp_mean_mag, phot_rp_mean_mag
FROM gaiadr2.gaia_source
WHERE parallax >= 2.0 AND parallax <= 2.1
```

Appendix B: Source selection based on the relative parallax uncertainty

In Section 3.4 we mention that by selecting sources basing on their relative parallax errors we might introduce un-physical over-densities in the data due to the fact that *Gaia*'s scanning law favours the ecliptic poles ($|b| > 45^\circ$). This effect is well visible when studying the distribution in the sky of *all* the sources within $d = 500$ pc before and after applying the condition $\sigma_w/\varpi > 5$. Fig. B.1 (right) shows the ratio between the histograms of the distribution in the sky of the sources before and after the relative parallax uncertainty selection is applied. The ecliptic poles are the regions where the values of the map are close to unity, and without any artefacts due to the scanning law³. The region where we observe the lowest values of completeness is towards the galactic plane for small positive b values.

Appendix C: New cluster at $l, b \sim (218.5^\circ, -2^\circ)$

As mentioned in the main text of the paper, we report the discovery of a candidate young cluster centred roughly at $l, b = (218.5^\circ, -2^\circ)$. Fig. C.1 shows the proper motion diagram (left), the parallax distribution (centre), and the colour-magnitude diagram (right) of the sources within $215^\circ \leq l \leq 222^\circ$ and $-5^\circ \leq b \leq 0^\circ$. Except for a few outliers, visible in particular in the proper motion diagram and in the parallax distribution, the cluster prominently stands out as an over-density in the proper motion diagram and as a peak in the parallax distribution.

Appendix D: Age maps

In this section we separately show the 3D density maps of the sources younger than 20 Myr and older than 10 Myr (blue, Fig. B.1, right), younger than 10 Myr and older than 5 Myr (green, Fig. B.1, centre), and younger than 5 Myr (red, Fig. B.1, left).

Appendix E: Density maps corresponding to the top and central panel of Fig. 6

The conclusion that most of the sources tracing the dust features in the top panel of Fig. 6 correspond to extincted and reddened main sequence stars, and the subsequent decision to further select pre-main sequence candidates according to their extinction and tangential velocity, comes from a preliminary inspection of the 3D density maps. Fig. E.1 (left) shows the density map corresponding

³ Other artefacts are present, such as spuriously high parallaxes: these are taken into account in the text by applying the conditions C.1 and C.2 from Lindegren et al. (2018).

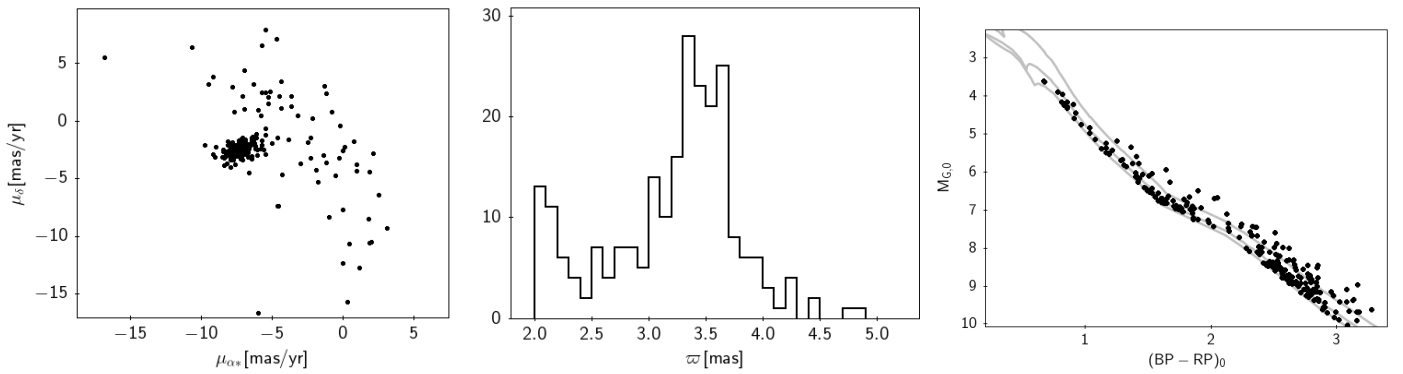


Fig. C.1. Left: Proper motion diagram of the sources selected in the region defined in the text. Proper motions cluster at $\mu_{\alpha*}, \mu_{\delta} \sim (-7., -2.5)$ mas yr⁻¹, with few, scattered outliers. Centre: Parallax histogram of the candidate cluster members. The histogram peaks at $\varpi \sim 3.4$ mas, indicating a distance to the cluster of ~ 295 pc. Right: Corrected colour-magnitude diagram of the candidate cluster members. The 10, 15, and 20 Myr PARSEC isochrones with solar metallicity and $A_V = 0$ mag are also plotted as grey solid lines.

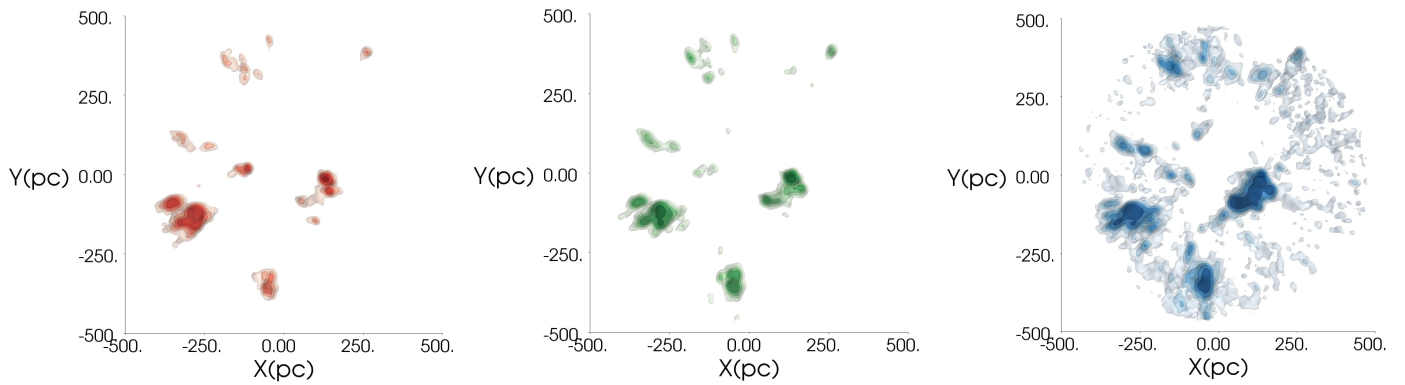


Fig. D.1. 3D density map of sources with age $10 < \tau < 20$ Myr (right), $5 < \tau < 10$ Myr (centre), $\tau < 5$ Myr (left).

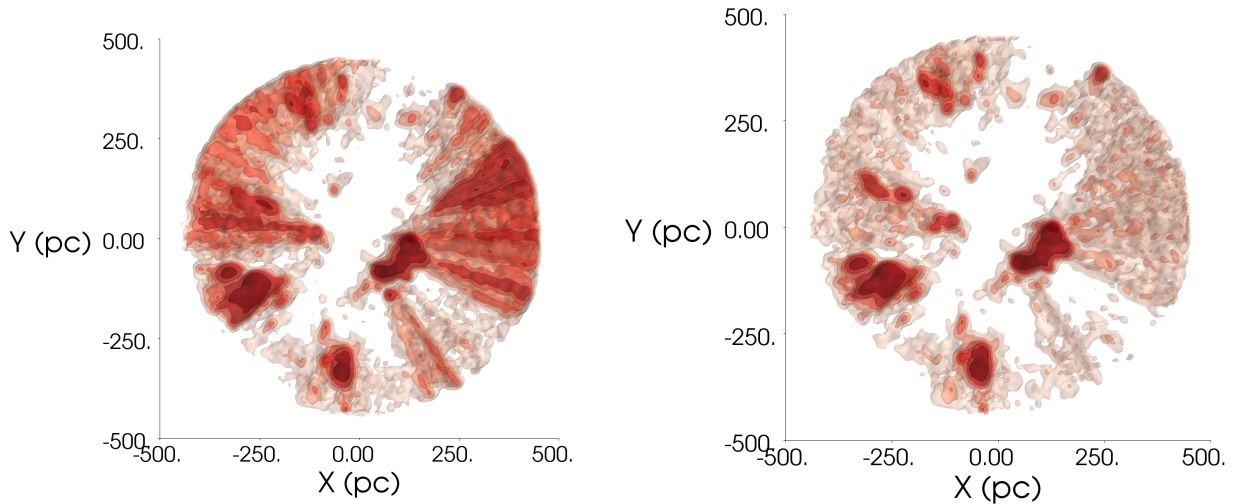


Fig. E.1. Left: 3D density map of the sources in the top panel of Fig. 6. Right: 3D density map of the sources in the central panel of Fig. 6.

to the top panel of Fig. 6, while Fig. E.1 (right) shows the density map corresponding to the central panel of Fig. 6. Fig. E.1 (left) does not show any additional clustering with respect to Fig. E.1 (right), except for dense 'stripes'. These features are located behind molecular clouds (see e.g. Lallement et al. 2018), and they are removed with the condition $A_G < 0.92$ mag, as shown in Fig. E.1 (left). Additional contaminants are removed by selecting stars according to their tangential velocity (compare Fig. E.1 (right) with Fig. 8).

Appendix F: UMS and PMS catalogues

Here we shortly describe the contents of the PMS and UMS catalogues. A detailed description of the column contents and format can be found in the *Gaia* DR2 documentation. Note that the proper motions are in galactic coordinates, thus we provide here the

correlation term between proper motion in galactic longitude and proper motion in galactic latitude: we stress however that for a proper use of the *Gaia* DR2 astrometry in galactic coordinates, users should transform the full covariance matrix of the ICRS astrometric parameters.

- `source_id`: unique source identifier (unique within a single release);
- `l`: galactic longitude [deg];
- `b`: galactic latitude [deg];
- `parallax`, `parallax` [mas];
- `parallax_error`, standard error of parallax [mas];
- `pm_l_cosb`: proper motion in galactic longitude [mas/yr];
- `pm_l_error`, standard error of proper motion in galactic longitude [mas/yr];
- `pm_b`: proper motion in galactic latitude [mas/yr] ;
- `pm_b_error`: standard error of proper motion in galactic latitude [mas/yr];
- `pm_l_pmb_corr`: correlation between proper motion in galactic longitude and proper motion in galactic latitude;
- `radial_velocity`: radial velocity [km/s];
- `radial_velocity_error`: radial velocity error [km/s];
- `phot_g_mean_mag`: G-band mean magnitude [mag];
- `phot_bp_mean_mag`: BP band mean magnitude [mag];
- `phot_rp_mean_mag`: RP band mean magnitude [mag];
- `phot_bp_rp_excess_factor`: BP/RP excess factor;
- `astrometric_chi2_al`: AL chi-square value;
- `astrometric_n_good_obs_al`: number of good observation AL;
- `A_G`: extinction in G-band [mag];
- `E_BPminRP`: colour excess in BP-RP [mag];
- `UWE`: Unit Weight Error, as defined in Lindegren et al. (2018).