Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/79821 holds various files of this Leiden University dissertation.

Author: Zari, E.M.
Title: Surveying young stars with Gaia: Orion and the Solar neighbourhood
Issue Date: 2019-10-22

Surveying young stars with Gaia: Orion and the Solar neighbourhood

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 22 oktober 2019
klokke 15:00 uur

door

Eleonora Zari
geboren te Milano, Italy
in 1989

Promotiecommissie

Promotor: Prof. dr. P.T. de Zeeuw
Co-promotor: Dr. A.G.A. Brown

Promotiecommissie: Prof. dr. J. Alves (University of Vienna)
                     Dr. J. de Bruijne (ESA/ESTEC)
                     Prof. dr. K.H. Kuijken
                     Dr. E. M. Rossi
                     Prof. dr. H.J.A. Rottgering

E quindi uscimmo a riveder le stelle.[1]

Dante,
Inferno, Canto XXXIV

---

[1]Thence we came forth to rebehold the stars.

Contents

1

Introduction

1.1 Star formation in the *Gaia* era

Studying how stars form is at the core of contemporary astrophysics research. It is not only interesting in itself, but it is also essential in understanding the formation and early evolution of planetary systems, and the structure and the evolution of galaxies.

The last stage of the massive star formation process, and the context in which new stars are formed, are the so-called *OB associations*, groups of young and massive stars of spectral type O and B (Ambartsumian 1947). By noting that the spatial densities of stars in OB associations are well below the threshold necessary to prevent their disruption by Galactic tidal forces, Ambartsumian calculated that associations must be young ($< 25\,\mathrm{Myr}$), a conclusion that was supported by ages derived by colour-magnitude diagrams and by theory of stellar structure and evolution. This agrees well with the fact that these groups are usually located in or near star-forming regions, and hence are prime sites for the study of star formation processes and of the interaction of early-type stars with the interstellar medium (see Blaauw 1964; de Zeeuw et al. 1999). Although O and B stars are mostly found in associations, some of them do not seem to be associated with any group or cluster. A fraction of those moves at high velocity: these are the so-called *runaway stars* (Blaauw 1952; Blaauw & Morgan 1954; Ambartsumian 1955).

Since the work of Ambartsumian, much progress has been made in our knowledge of OB associations. At the end of the 20th century, the *Hipparcos* mission allowed for an extensive census of stellar content of the nearby OB associations (de Zeeuw et al. 1999). This was complemented, in the past two decades, by an unprecedented stream of new observational information and a parallel renaissance in theoretical investigation and numerical modelling of the star-formation process (see the reviews by McKee & Ostriker 2007; Kennicutt & Evans 2012). Yet, some questions remain unanswered. How are associations formed and how do they disperse in the field? What causes the distinction between the formation of bound open clusters and unbound associations? What are the characteristics of the stellar populations within single associations in terms of age sequences and kinematics? What are the properties of the ensemble of OB associations? What is their disposition in space and how does it compare with what is observed in other galaxies?

The data of the *Gaia* satellite are crucial to address these questions, as they allow to study the spatial structure, kinematics, and ages of OB associations with unprecedented precision. In this thesis we obtained a detailed census of the young stellar populations in the solar neighbourhood, focusing in particular on the Orion OB association. We found that both single associations and the ensemble of OB associations in

the solar vicinity present a high degree of sub-structure in physical space, kinematics, and ages. The star formation history of the solar neighbourhood is complex, and it does not quite follow sequential star formation scenarios. This calls for a revision of our theories of the propagation and triggering of star formation. Data from the future releases of the *Gaia* satellite and from upcoming spectroscopic surveys will also contribute in exploring in more detail the kinematic and physical sub-structure of large star formation complexes.

In the remainder of this introduction, I will discuss the main features of OB associations. I will focus in particular on their spatial arrangement in the solar neighbourhood and on the properties of the Orion OB association, and I will describe the characteristics of O- and B-type runaway stars. I will give a short overview of the data products of the *Gaia* satellite and I will finally summarise the Chapters of this thesis and present some prospects for future research.

1.2 OB associations

OB associations were first recognised as loose groups of O- and B-type stars, but they contain members across the mass spectrum, including intermediate-mass A/F stars and lower-mass G/K/M stars, which are still in the pre-main sequence (PMS) phase of stellar evolution. Though the lower mass ($< 1.5\,\mathrm{M_\odot}$) stars blend in with the Galactic field population and are therefore much more difficult to identify than the OB stars, they comprise the dominant stellar component of OB associations (Briceño et al. 2007b).

The members of OB associations can be singled out using a combination of instruments and techniques, summarised for example in Brown et al. (1999) and Briceño et al. (2007a). Methods based on single-epoch photometry and on proper motions (and on their combination) were applied in this thesis. Low-mass PMS stars are located in the colour-magnitude diagram (CMD) above the zero-age main sequence (ZAMS). For this reason it is relatively straightforward to separate them from main sequence sources located at similar distances (see for instance Sherry et al. 2004; Kenyon et al. 2005; Bouy et al. 2014). Proper motion surveys allow to identify members of OB associations based on their kinematics (see for example de Zeeuw et al. 1999; de Bruijne 1999a; Hoogerwerf & Aguilar 1999). Indeed, associations are gravitationally unbound, however they have small internal velocity dispersion (a few kilometres per second), and thus they form coherent structures in velocity space. The streaming motion of the association as a whole, as well as the Solar motion, is reflected as a motion of the members towards a convergent point on the sky. An example of this is shown in Fig. 1.1, for the nearest OB association, Scorpius-Centaurus.

Precise proper motions allow to study the internal kinematic properties of OB associations, which provide clues for the understanding their formation. To explain the origin of OB associations two main competing models have been proposed. According to the first model (Lada & Lada 2003), OB associations are expanding remnants of star clusters. Star clusters are formed embedded within molecular clouds, where the gravitational potential of both the stars and the gas holds them together. When feedback disperses the gas left over from star formation, the cluster becomes super-virial and will expand and disperse, thus being visible for a short time as an OB association.
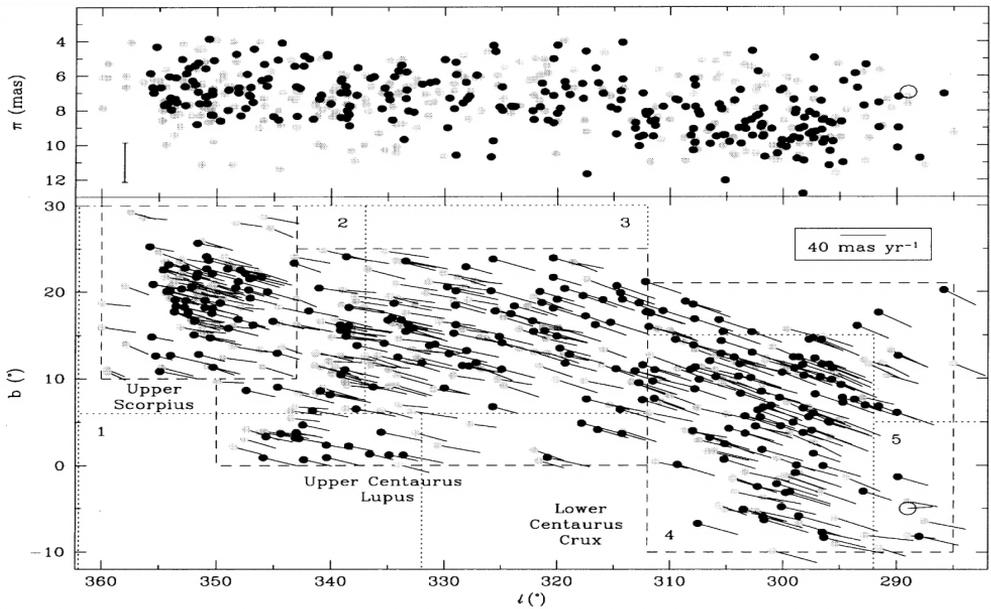
Figure 1.1: Positions and proper motions (bottom), and parallaxes (top), for 521 members of the Scorpius-Centaurus association (Sco OB2) selected from 7974 stars in the *Hipparcos* catalogue in the area bounded by the dashed lines (de Zeeuw et al. 1999). The vertical bar in the top panel corresponds to the average $\pm 1\sigma$ parallax range for the stars shown. The dotted lines are the schematic boundaries of the classical subgroups Upper Scorpius (2, US), Upper Centaurus Lupus (3, UCL), Lower Centaurus Crux (4, LCC), and the candidate subgroups (1 and 5) defined by (Blaauw 1964). The large open circle represents the open cluster IC 2602. The figure and the caption are from de Zeeuw et al. (1999).

The second model (Clark et al. 2005) instead predicts that OB associations are born in highly sub-structured, multiple small-scale star formation events that take place in long and filamentary molecular clouds. The kinematics of OB associations would keep memory of the parental gas sub-structure where they originated. The results reported by Wright et al. (2016), Wright & Mamajek (2018), and in this thesis (Chapter 3) seem to confirm the latter view.

A problem that both models need to explain is the star formation history of OB associations. Indeed, although OB associations as a whole occupy large regions in physical space ($\sim 100\,\mathrm{pc}$), they can be divided in smaller sub-groups, that can be distinguished on the basis of the ages of their members, their degree of association with interstellar matter (Blaauw 1964), and on the basis of their kinematics (see for example Cantat-Gaudin et al. 2019; Kounkel et al. 2018). Simple triggered star formation scenarios (see Preibisch & Zinnecker 2007, and references therein) struggle in explaining the lack of regular age sequences and the apparent coordination of star formation on large spatial scales, and more complex models are required to explain the observations (see for instance Krause et al. 2018).

1.3 The Gould Belt

OB associations in the solar vicinity seem to be arranged in a ring-like structure, inclined by $\sim 20°$ with respect to the plane of the Milky Way. This structure was recognised by by Herschel (1847) and Gould (1874) and became known as the *Gould Belt*. This huge ring of bright stars and gas, up to 700 pc in diameter, seems to link a number of the closest associations, some of which fit a coherent pattern of expansion and rotation (Lindblad et al. 1997; Torra et al. 1997). The Gould Belt was also found to be associated with young stars (Guillout et al. 1998) and interstellar material (Lindblad 1967), the latter interpreted as an expanding ring of gas (Olano 1982; Elmegreen 1982). Various scenarios have been proposed to explain the formation of the Belt, which include the passage of the Carina spiral arm near the Sun (Elmegreen 1993; Elmegreen & Efremov 1998), the impact of an high velocity cloud on the stellar disc (Comeron et al. 1998), a cascade of supernova explosions (Olano 2001), and the collision between a dark matter clump and a gas cloud (Bekki 2009). Elmegreen (1993) in particular proposed that the passage of the Carina spiral arm ~ 60 Myr ago triggered the formation of the Cas-Tau association. The Lindblad's ring could have been then generated by feedback and supernova explosions from high-mass stars in Cas-Tau. The Scorpius-Centaurus, Orion, Perseus, and Lacerta OB associations would have formed around 20 Myr ago from Lindblad's ring and constituted a second generation of star formation. The present star formation seen in Taurus and in Ophiuchus is regarded as the third generation. Figure 1.2 shows locations of the OB associations studied in de Zeeuw et al. (1999) projected onto the Galactic plane. de Zeeuw et al. (1999) concluded that the physical arrangement of the ensemble of OB associations was in qualitative agreement with Elmegreen (1993) picture, but called for a reassessment of the star formation history of the solar neighbourhood, as they observed that there was not a clear difference between bound open clusters and unbound expanding associations and that the total mass of young stellar groups might have been underestimated.

Such reassessment came ten years later, when Elias et al. (2009) studied the distribution of young open clusters in the solar neighbourhood, using again the *Hipparcos* catalogue. They proposed that the position with respect to the galactic plane and the kinematics of the two associations dominating the inclination of the Gould Belt, Orion and Scorpius-Centaurus, can be explained in terms of their relative position to the density maximum of the Local Arm in the solar neighbourhood. They therefore concluded that the Gould Belt could be explained by the result of the internal dynamics of the Galactic disc.

This conclusion has been further corroborated by Bouy & Alves (2015). Bouy & Alves (2015) re-analysed the distribution of O- and B-type stars in the solar neighbourhood, and by making use of a three-dimensional kernel estimation, they studied their spatial density and produced the three-dimensional density map shown in Fig. 1.3. They suggested that the distribution of O and B stars in the solar neighbourhood would be better described by stream-like structures, similarly to what is observed in other spiral galaxies, and concluded that there is no evidence of a ring-like structure such as the Gould Belt in the three dimensional configuration of young, bright stars in the solar neighbourhood. Bouy & Alves (2015) results were based on the *Hipparcos* data, and motivated us to perform the study presented in Chapter 4.

1.4   Orion

The figure of Orion the Hunter is a familiar sight in the winter sky of the Northern hemisphere (see Fig. 1.4). The area is an extraordinarily active site of star formation. Over the years, no similar region has received such intense astronomical scrutiny, or has been studied with such a variety of observational tools (see the reviews by Stahler & Palla 2005; Bally 2008).

The *Hipparcos* census of nearby OB associations (de Zeeuw et al. 1999) represented a major step forward in terms of determining the membership of OB associations, however the data was not accurate enough to make significant progress in Orion. The main reasons for this are that a) the distance to Orion is $\sim 400\,\mathrm{pc}$, thus the *Hipparcos* parallax uncertainties were large, and b) Orion's motion is mainly directed away from the Sun, thus the observed proper motions are small. Thus, a detailed characterisation of the stellar population of Orion in terms of kinematics, ages, and spatial structure was still missing: this constitutes one of the main topics of this thesis. In the following, we will describe the features of the Orion region relevant to this thesis.

The Orion OB association (Ori OB1) is divided in several groups and clusters, partially super-imposed along the line of sight (Blaauw 1964; Brown et al. 1994). Blaauw (1964) suggested that star formation sequentially propagated in the association. The members of the oldest sub-group, located north-west of the Belt stars (Ori OB1a, 8-12 Myr, Bally 2008) may have triggered the formation of the Ori OB1b sub-group (3-6 Myr) towards the Orion's Belt, from which star formation seemed to have propagated further south in the Ori OB1c region (2-6 Myr). The youngest sub-group is the Orion Nebula Cluster (ONC, see for instance Da Rio et al. 2014), located at the northern tip of the Orion A molecular cloud. Within these four groups, many clusters have been identified, such as 25 Ori (Briceño et al. 2007b), σ Ori (Walter et al. 2008) and λ Ori (Mathieu 2008). Spectroscopic data allowed to analyse the kinematic properties of
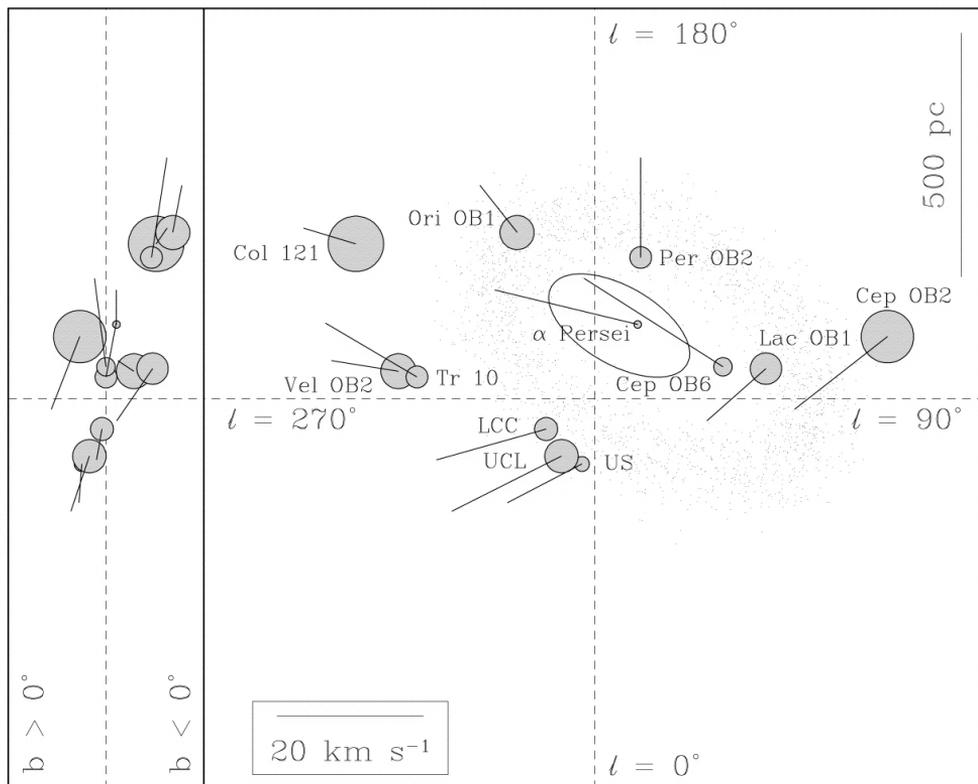
Figure 1.2: Locations of the OB associations studied in de Zeeuw et al. (1999) projected onto the Galactic plane. The gray circles indicate the physical dimensions as obtained from the angular dimensions and mean distances, on the same scale. The lines represent the streaming motions, derived from the average proper motions, mean distances and median radial velocities of the secure members, corrected for "standard" solar motion and Galactic rotation. The ellipse around the α Persei cluster indicates the Cas-Tau association. The small dots schematically represent the Olano (1982) model of the Gould Belt. The figure and caption are from de Zeeuw et al. (1999).
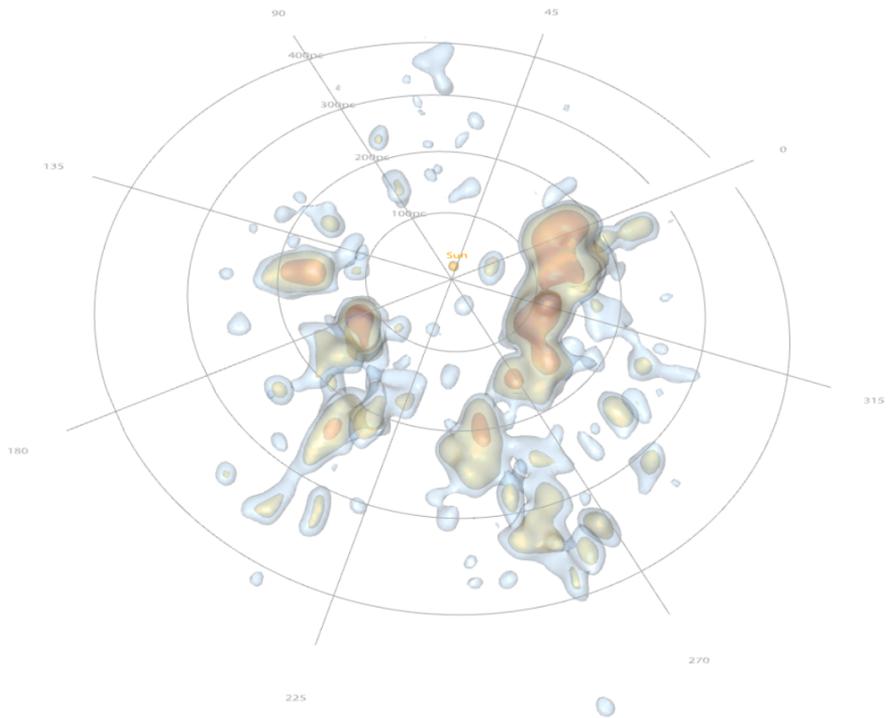
Figure 1.3: 3D map of OB star density iso-surface (1.0, 1.38 and 2.76 $\times 10^{-4}$ OB star per pc^3 (Bouy & Alves 2015). The circles have 100, 200, 300, and 400 pc respectively. The radii represent longitude values. The figure and the caption are from Bouy & Alves (2015).
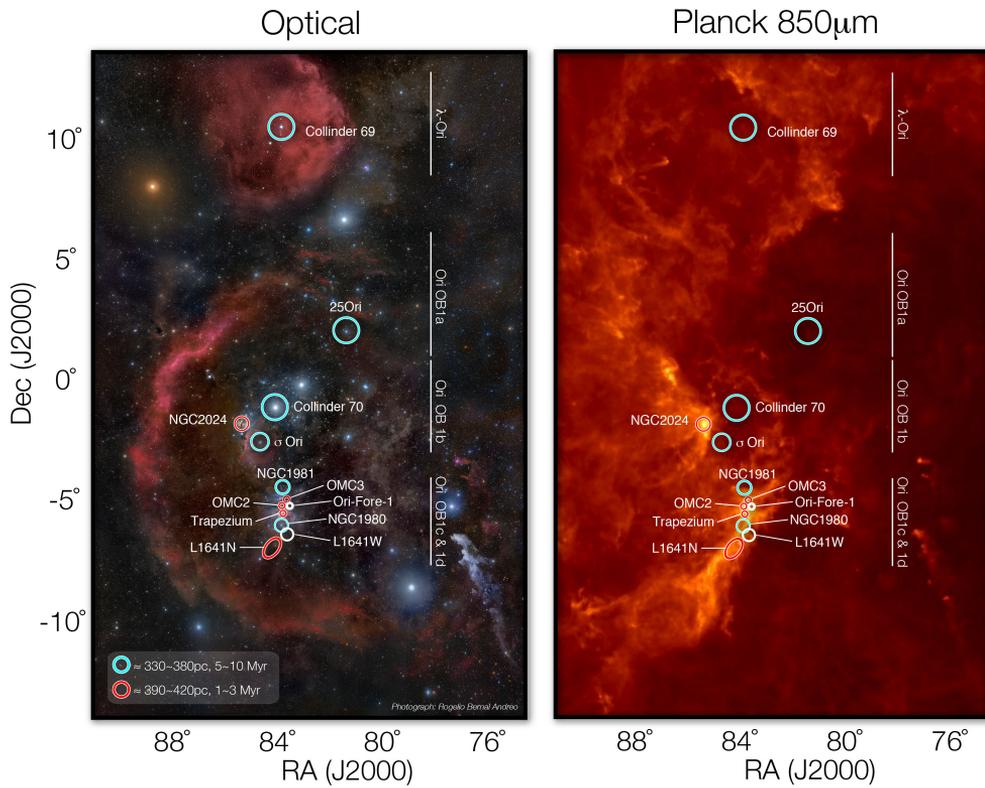
Figure 1.4: Left panel: distribution of groups over-plotted on an optical photograph of the Orion constellation (courtesy of Rogelio Bernal Andreo - *DeepSkyColors.com*). Right panel: same as left panel, but over-plotted on a far-infrared (850 μm) *Planck* map. The figure and the caption are from Bouy et al. (2014).
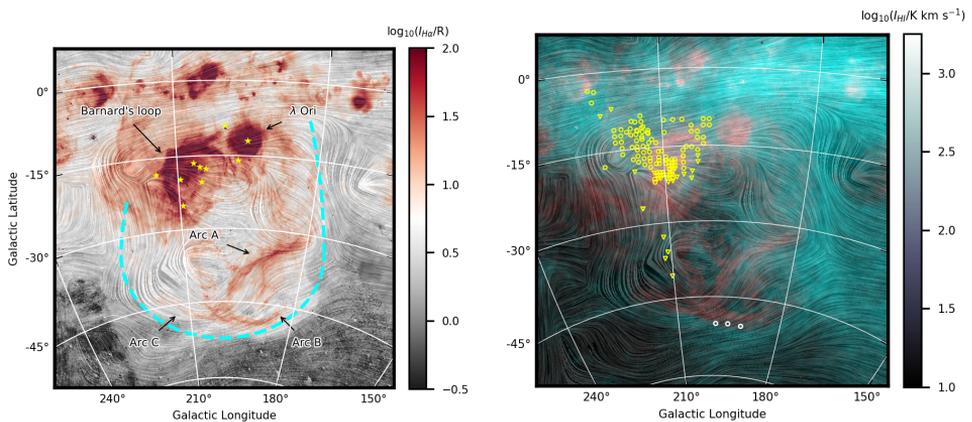
Figure 1.5: The drapery pattern corresponds to the plane-of-the-sky magnetic field orientation inferred from the Planck 353 GHz polarisation observations. Left. Total integrated Hα emission map. The dashed line indicates the approximate location of the edge of the super-bubble. The yellow symbols correspond to the main stars in the Orion constellation . Right. Total integrated Hα emission and HI 21 cm emission integrated between - 20 and 20 $\mathrm{km\,s^{-1}}$ shown in red and teal colours, respectively. The yellow symbols correspond to the line-of-sight magnetic field directions derived from the HI emission-line Zeeman splitting observations. The circles and triangles correspond to magnetic fields pointing toward and away from the observer, respectively. The three white circles in the bottom are the regions analysed in Soler et al. (2018), from which these figure and caption are taken.

these groups. Briceño et al. (2007b) and Jeffries et al. (2006) found in particular that the 25 Ori and σ Ori clusters have different kinematic properties than the sub-groups in which they are located (Ori OB1a and OB1b, respectively). Alves & Bouy (2012) and Bouy et al. (2014) recently reported the discovery of a young population of stars in the foreground of the ONC, which was however questioned by Da Rio et al. (2016), Fang et al. (2017) and Kounkel et al. (2017a), while Kubiak et al. (2016) identified a rich and young population surrounding ϵ Ori.

The combined effects of UV radiation, stellar winds, and supernova explosions from the Orion OB1 association have created a bubble that spans $\sim 40°$ in the sky (or 300 pc at a distance of 400 pc): the Orion-Eridanus super-bubble. Ochsendorf et al. (2015) studied in detail the structure and the evolution of the Orion-Eridanus super-bubble, concluding that it consists of a series of nested shells. They also found that Barnard's Loop is a part of a complete bubble, probably associated with a supernova remnant. Both the Barnard's Loop bubble and the λ Ori Bubble are expanding within the Orion-Eridanus super-bubble. By using polarization observations by the *Planck* satellite, Soler et al. (2018) characterised the magnetic field in the Orion-Eridanus super-bubble, finding that the large-scale magnetic field in the region was primarily shaped by the expanding super-bubble (see Fig. 1.5).

Orion contains two giant molecular clouds ($M \sim 10^5\,\mathrm{M_\odot}$): the Orion A molecular cloud, located in the southern portion of the constellation, and the Orion B cloud, that lies at the east of the Orion's Belt (Bally 2008). Both the clouds are thought to be located within the walls of the Orion-Eridanus super-bubble. Schlafly et al. (2015) presented 3D maps of dust reddening, tracing the total column density towards the Orion clouds (reported in Fig. 1.6). They found that the Orion A and B clouds are
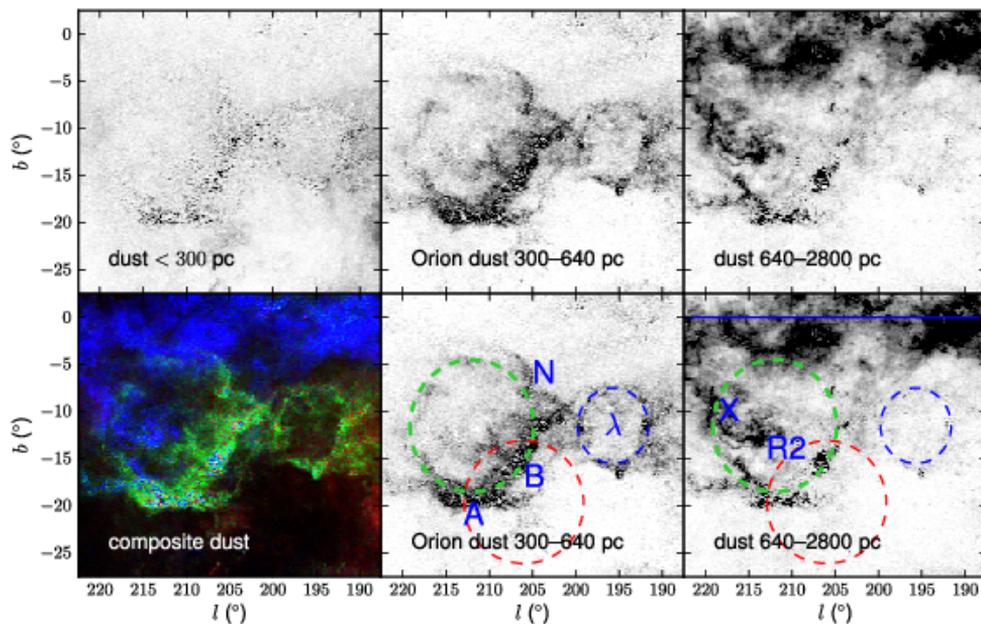
Figure 1.6: The 3D distribution of dust towards the Orion Molecular Complex. The top panels show the column density of dust with distance < 300 pc, 300-640 pc, and 640-2800 pc, respectively. The fourth panel (bottom left) shows a 3-colour composite image of these three slices, illustrating the 3D distribution of dust in the region. Finally, the fifth and sixth panels again show the Orion and more distant dust, this time over-plotting circles tracing the various bubble-like structures in the region. The green dashed circle shows the Orion dust ring; the blue dashed circle shows the λ Ori molecular ring; and the red dashed circle approximately aligns with Barnard's Loop (see Figs. 1.4 and 1.5). The last two panels also label the Orion A (A) and Orion B (B) molecular clouds, the Northern Filament (N), the star λ Ori, Monoceros R2 (R2), the Crossbones (X), and the Galactic plane (horizontal line). Differential extinction and an insufficient number of well-observed stars lead to artefacts in the far distance slice through particularly dense clouds in Orion A and B. White to black corresponds to 0-0.7 mag E(B-V). This same scale is used for each of the colour planes in the lower left panel. The figure and the caption are from Schlafly et al. (2015).

part of a "dust ring", which may have implications on the triggering of star formation in the region.

1.5 OB stars on the run

As mentioned in Section 1 of this introduction, not all O- and B-type stars are found in OB associations and clusters. A large fraction of these field objects moves at very high velocities: these are referred to as "runaway" stars (Blaauw 1952; Ambartsumian 1955). Orion has been the source of several well known runaway stars, including the $150 \, \text{km s}^{-1}$ runaway star AE Auriga, and the $117 \, \text{km s}^{-1}$ μ Columbae which is moving exactly in the opposite direction (Blaauw 1991). Hoogerwerf et al. (2001) used new *Hipparcos* proper motion data to show that these two stars, and the colliding wind X-ray binary ι Ori were at the same location in the sky ≈ 2.6 Myr ago. Gualandris et al. (2004) argue that the two runaway stars and ι Ori suffered a four-body interaction in

which two binaries in the same cluster underwent an exchange. The two most-massive members became the tight ι Ori binary; the gravitational energy released kicked the two less massive stars out of the region at high velocity. The process explained above describes one of the two runaway production channels, and it is usually referred to as *dynamical ejection scenario* (DES, Poveda et al. 1967; Leonard 1991). The second scenario, the *binary supernova scenario* (BSS, Blaauw 1961; Zwicky 1957; Boersma 1961), predicts that a runaway star might originally have been a member of a close binary pair consisting of two massive stars. If the companion exploded as a supernova, the star of interest could escape with a speed equal to the orbital value. Runaway stars have been identified in the *Hipparcos* catalogue by Hoogerwerf et al. (2001) and Tetzlaff et al. (2011). Hoogerwerf et al. (2001) selected 56 sources of spectral type from O to B5 with total peculiar velocities higher than $30 \, \mathrm{km \, s^{-1}}$, and, by studying their orbit identified the parent associations for a sub-set of them. Tetzlaff et al. (2011) identified young stars (< 50 Myr) of any spectral type, and selected those with large peculiar velocities, finding in total 2547 candidate runaway stars. The present and upcoming *Gaia* data releases are expected to drastically increase the available sample of stars with precisely known velocities, allowing for the construction of more complete samples of runaway star candidates. This is the goal of Chapter 5. Such samples will be then compared with the results of numerical simulations that predict the fraction of runaway stars produced by the BSS or the DES, such as those by Renzo et al. (2019b) and Ryu et al. (2017). This will be useful to determine the relative importance of the two formation mechanisms, which in turn will provide more clues on massive star formation and evolution (see Renzo et al. 2019b; Gvaramadze et al. 2009; Portegies Zwart et al. 2007).

1.6 Gaia

Gaia is an ESA mission, launched at the end of 2013 (Perryman et al. 2001; Gaia Collaboration et al. 2016a). The main aim of *Gaia* is to measure the three-dimensional spatial and the three-dimensional velocity distribution of stars and to determine their astrophysical properties, such as surface gravity and effective temperature, to map and understand the formation, structure, and past and future evolution of our Galaxy. *Gaia*'s astrometry delivers absolute parallaxes and proper motions. Complementary photometry and radial velocities are also provided by *Gaia* so that astrophysical parameters and six dimensional phase space information can be derived.

Two years and half after the launch, the first release of data was presented (hereafter *Gaia* DR1). *Gaia* DR1 (Gaia Collaboration et al. 2016a,b) is based on the first 14 months of mission and consists of three components. The first component consists of a primary astrometric data set which contains the positions, parallaxes, and mean proper motions for about 2 million of the brightest stars in common with the *Hipparcos* and Tycho-2 catalogues, the Tycho-Gaia Astrometric Solution (TGAS), and a secondary astrometric data set containing the positions for an additional 1.1 billion sources. The second component is the photometric data set, consisting of mean G-band magnitudes for all sources. The third component is formed by the G-band light curves and the characteristics of ~ 3000 Cepheid and RR Lyrae stars, observed at high cadence around the south ecliptic pole. For the primary astrometric data set the typ-

ical uncertainty is about 0.3 mas for the positions and parallaxes, and about 1 mas/yr for the proper motions. A systematic component of ~ 0.3 mas should be added to the parallax uncertainties. For the subset of $\sim 94\,000$ Hipparcos stars in the primary data set, the proper motions are much more precise at about 0.06 mas/yr. For the secondary astrometric data set, the typical uncertainty of the positions is ~ 10 mas. The median uncertainties on the mean G-band magnitudes range from the mmag level to ~ 0.03 mag over the magnitude range 5 to 20.7 mag.

   The second *Gaia* data release (*Gaia* DR2, Gaia Collaboration et al. 2018a), which is based on the data collected during the first 22 months of the nominal mission lifetime, was made public on the 25th of April 2018. *Gaia* DR2 represents a major advance with respect to *Gaia* DR1, making the leap to a high-precision parallax and proper motion catalogue for over 1 billion sources, supplemented by precise and homogeneous multi-band all- sky photometry and a large radial velocity survey at the bright ($G \sim 13$ mag) end. *Gaia* DR2 contains celestial positions and the apparent brightness in G-band for approximately 1.7 billion sources. For 1.3 billion of those sources, parallaxes and proper motions are in addition available. This data release contains four new elements: broad-band colour information in the form of the apparent brightness in the G_{BP} (330-680 nm) and G_{RP} (630-1050 nm) bands for 1.4 billion sources; median radial velocities for ≈ 7 million stars; for between 77 and 161 million sources estimates of the stellar effective temperature, extinction, reddening, and radius and luminosity; variability information for 0.5 million stars; epoch astrometry and photometry for a pre-selected list of 14 000 minor planets in the solar system.

1.7   This thesis

In Chapter 2, we use *Gaia* DR1 to explore the three-dimensional arrangement and age ordering of the many stellar groups toward the Orion OB association, aiming at a new classification and characterisation of the stellar population not embedded in the Orion A and B molecular clouds. We find evidence for the presence of a young population at a parallax $\varpi \approx 2.65\,\mathrm{mas}$, which is loosely distributed around the following known clusters: 25 Ori, ϵ Ori, and σ Ori, and NGC 1980 (ι Ori) and the Orion Nebula Cluster (ONC). The low mass counterpart of this population is visible in the colour-magnitude diagrams constructed by combining *Gaia* DR1 G-band photometry and 2MASS, and in the density distribution of the sources on the sky. We estimate the ages of this population using a Bayesian isochronal fitting procedure assuming a unique parallax value for all the sources, and we infer the presence of an age gradient going from 25 Ori (13-15 Myr) to the ONC (1-2 Myr). Finally, we provisionally relate the stellar groups to the gas and dust features in Orion. These results represent the first step toward using Gaia data to unravel the complex star formation history of the Orion region in terms of the various star formation episodes, their duration, and their effects on the surrounding interstellar medium.

   In Chapter 3, we present a study of the three dimensional structure, kinematics, and age distribution of the Orion OB association, based on *Gaia* DR2. The goal of this Chapter is to obtain a complete picture of the star formation history of the Orion complex and to relate our findings to theories of sequential and triggered star formation. We select the Orion population with simple photometric criteria, and we explored

its physical arrangement by using a three dimensional density map. The map shows structures that extend for roughly $150\,\mathrm{pc}$ along the line of sight, divided in multiple sub-clusters. We separate the different groups by using a density-based clustering algorithm, and we studied their kinematic properties first by inspecting their proper motion distribution, and then by applying a kinematic modelling code based on an iterative maximum likelihood approach, which we use to derive their mean velocity, velocity dispersion and isotropic expansion. By using an isochrone fitting procedure we provide ages and extinction values for all the groups. We confirm the presence of an old population (~ 15 Myr) towards the 25 Ori region, and we find that groups with ages of $12 - 15\,\mathrm{Myr}$ are present also towards the Belt region. We notice the presence of a population of ~ 10 Myr also in front of the Orion A molecular cloud. Our findings suggest that star formation in Orion does not follow a simple sequential scenario, but instead consists of multiple events, which caused kinematic and physical sub-structure. To fully explain the detailed sequence of events, specific simulations and further radial velocity data are needed.

In Chapter 4, we study the three dimensional arrangement of young stars in the solar neighbourhood using *Gaia* DR2 and we provide a new, original view of the spatial configuration of the star-forming regions within 500 pc of 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 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. We conclude that the 3D density maps show no evidence for the existence of the ring-like structure which is usually referred to as the Gould Belt.

In Chapter 5, we search for early type runaway stars within 1 kpc from the Sun by using *Gaia* DR2 and the stellar parameters provided in the `StarHorse` catalogue (Anders et al. 2019). We select upper main sequence (UMS) sources by applying simple photometric cuts. Our sample consists of O-, B- and early A-type sources. We study the tangential velocity, and, when possible, the total velocity distribution of our sample, and we classify as candidate runaway stars those sources that have tangential velocities significantly different from the rest of the population (2σ) or total velocities higher than $30\,\mathrm{km\,s^{-1}}$. We study the orbits of the candidate runaway stars with literature radial velocities, and we find that around half of our sources were produced further than 1 kpc. We focus on the runaway star candidates in the Orion and Scorpius-Centaurus (Sco-Cen) regions. In Orion, we confirm previously known runaway stars and we enlarge the sample by adding 6 new runaway candidates. In

Sco-Cen we identify two runaway star candidates that likely share the same origin. The analysis of the entire sample is on-going. Finally, we discuss our findings in the context of other studies, end we estimate the completeness of our sample. To further study the candidate runaway stars, more radial velocities are needed. These could be obtained from planned surveys, such as SDSS-V, WEAVE, and 4MOST, but also from dedicated proposals.

1.8 Outlook

In this thesis we have found many clues indicating that OB associations are complex entities, and we provided a description of their properties in terms physical structure, kinematics, and ages. We used early-type massive stars and pre-main sequence sources to trace the structure of the solar neighbourhood within 500 pc from the Sun and we studied the kinematics and dynamics of some of the fastest young stars in the Milky Way.

We did not however answer many questions, starting from: how are OB associations formed? The scenarios proposed to model the formation of OB associations assume that radiation and winds from massive stars disperse the gas surrounding them, locally terminating the star formation process and driving shocks in other regions, which cause cloud collapse and new star formation episodes. The models make different predictions for the observations, however none of them seem to completely explain the data. Recent studies on the Scorpius-Centaurus (Sco-Cen) association by Pecaut & Mamajek (2016) and Krause et al. (2018) show complex star formation histories, indicative of multi-stage formation processes, and not consistent with simple triggered star formation scenarios. Future *Gaia* data releases complemented with spectroscopic surveys, such as SDSS-V, WEAVE, and 4MOST, will enormously increase our knowledge of the formation and evolution of OB associations, both in the solar vicinity and in distant regions of the Milky Way. At the same time, detailed simulations of large scale star formation events will be needed to interpret the data. Another way to test theories of triggered star formation is to compare the kinematics of past and present star formation episodes. For this purpose it will be possible to combine the data from future *Gaia* releases with proper motions data in the infra-red such as those of the VISIONS survey. VISIONS, the VISTA star formation atlas, is a survey that aims to construct a sub-arcsec near-infrared atlas of all nearby ($< 500\,\mathrm{pc}$) star formation complexes from the southern hemisphere. The survey will provide multi-epoch, H-band observations that will be used to derive proper motions for the sources observed, with precision of $1 - 2\,\mathrm{mas/yr}$. By using VISIONS, it will be possible to relate the motions of embedded sources, invisible to *Gaia*, with those of evolved young stars that have already dispersed the gas surrounding them.

Going beyond the solar neighbourhood, O and B-type stars can be used to trace the structure of the spiral arms, and can probe the spiral arm features in remote regions of the Milky Way. Indeed, our position in the disc of the Milky Way does not allow to capture the global picture easily. For example, the number of spiral arms is still somewhat debated, although it is considered to be either 2 or 4. This has implications on the structure of our Galaxy: a large number of arms would support the view that the Galaxy better resembles a flocculent, rather than grand design spiral. An alter-

native interpretation is that the Galaxy has two main spiral arms, with the other two arms perhaps only present in gas and young stars (Drimmel 2000) By combining future *Gaia* data releases with, for instance CO and dust extinction maps we will be able to study in detail the connection between the spatial configuration and the different kinematic properties of gas and stars in the disc of the Galaxy. By doing so we will address two fundamental questions in large scale star formation studies:

1) which are the mechanisms triggering and propagating star formation in the Galaxy?

2) how is the interstellar medium shaped and transformed under the influence of young massive stars?

2

Mapping young stellar populations towards Orion with Gaia DR1

We use the first data release of the *Gaia* mission to explore the three dimensional arrangement and the age ordering of the many stellar groups towards the Orion OB association, aiming at a new classification and characterization of the stellar population not embedded in the Orion A and B molecular clouds. We make use of the parallaxes and proper motions provided in the *Tycho Gaia Astrometric Solution* (TGAS) sub-set of the *Gaia* catalogue, and of the combination of *Gaia* and 2MASS photometry. In TGAS, we find evidence for the presence of a young population, at a parallax $\varpi \sim 2.65\,\mathrm{mas}$, loosely distributed around some known clusters: 25 Ori, ϵ Ori and σ Ori, and NGC 1980 (ι Ori) and the Orion Nebula Cluster (ONC). The low mass counterpart of this population is visible in the color-magnitude diagrams constructed by combining *Gaia* G photometry and 2MASS. We study the density distribution of the young sources in the sky, using a Kernel Density Estimation (KDE). We find the same groups as in TGAS, and also some other density enhancements that might be related to the recently discovered Orion X group, the Orion dust ring, and to the λ Ori complex. The maps also suggest that the 25 Ori group presents a northern elongation. We estimate the ages of this population using a Bayesian isochronal fitting procedure, assuming a unique parallax value for all the sources, and we infer the presence of an age gradient going from 25 Ori (13-15 Myr) to the ONC (1-2 Myr). We confirm this age ordering by repeating the Bayesian fit using the Pan-STARRS1 data. Intriguingly, the estimated ages towards the NGC 1980 cluster span a broad range of values. This can either be due to the presence of two populations coming from two different episodes of star formation or to a large spread along the line of sight of the same population. Some confusion might arise from the presence of unresolved binaries, which are not modelled in the fit, and usually mimic a younger population. Finally, we provisionally relate the stellar groups to the gas and dust features in Orion. Our results form the first step towards using the *Gaia* data to unravel the complex star formation history of the Orion region in terms of the different star formation episodes, their duration, and their effects on the surrounding interstellar medium.

2.1   Introduction

OB stars are not distributed randomly in the sky, but cluster in loose, unbound groups, which are usually referred to as OB associations (Blaauw 1964). In the solar vicinity, OB associations are located near star-forming regions (Bally 2008), hence they are prime sites for large scale studies of star formation processes and of the effects of early-type stars on the interstellar medium.

At the end of the last century, the data of the *Hipparcos* satellite (Perryman 1997) allowed to characterize the stellar content and the kinematic properties of nearby OB associations, deeply changing our knowledge and understanding of the solar vicinity and the entire Gould's Belt (de Zeeuw et al. 1999). The canonical methods used for OB association member identification rely on the fact that stars belonging to the same OB association share the same mean velocity (plus a small random velocity dispersion). The common space velocity is perceived as a motion of the members towards a convergent point in the sky (for more details see e.g. de Bruijne 1999a; Hoogerwerf & Aguilar 1999). Unfortunately, the motion of the Orion OB association is directed primarily radially away from the Sun. For this reason the methods of membership determination using the *Hipparcos* proper motions did not perform well in Orion.

The Orion star forming region is the nearest ($d \sim 400\,\mathrm{pc}$) giant molecular cloud complex and it is a site of active star formation, including high mass stars. All stages of star formation can be found here, from deeply embedded protoclusters, to fully exposed OB associations (e.g. Brown et al. 1994; Bally 2008; Briceno 2008; Muench et al. 2008; Da Rio et al. 2014; Getman et al. 2014). The different modes of star formation occurring here (isolated, distributed, and clustered) allow us to study the effect of the environment on star formation processes in great detail. Moreover, the Orion region is an excellent nearby example of the effects that young, massive stars have on the surrounding interstellar medium. The Orion-Eridanus superbubble is an expanding structure, probably driven by the combined effects of ionizing UV radiation, stellar winds, and supernova explosions from the OB association (Ochsendorf et al. 2015; Schlafly et al. 2015).

The Orion OB association consists of several groups, with different ages, partially superimposed along our line of sight (Bally 2008) and extending over an area of $\sim$ $30° \times 25°$ (corresponding to roughly $200\,\mathrm{pc} \times 170\,\mathrm{pc}$). Blaauw (1964) divided the Orion OB association into four subgroups. Orion OB1a is located Northwest of the Belt stars and has an age of about 8 to 12 Myr (Brown et al. 1994). Orion OB1b contains the Belt stars and has an age estimate ranging from 1.7 to 8 Myr (Brown et al. 1994; Bally 2008). Orion OB1c (Bally 2008, estimated age from 2 to 6 Myr) includes the Sword stars and is located directly in front of the Orion Nebula, M43, and NGC 1977. Hence, it is very hard to separate the stellar populations of OB1c and OB1d, the latter corresponding to the Orion Nebula Cluster (ONC, see e.g. Da Rio et al. 2014). It is not clear whether the entire region is a single continuous star forming event, where Ori OB1c is the more evolved stellar population emerging from the cloud where group 1d still resides, or whether 1c and 1d represent two different star formation events (see e.g. Muench et al. 2008). In subsequent studies, many more sub-groups have been identified, such as 25 Ori (Briceño et al. 2007b), σ Ori (Walter et al. 2008) and λ Ori (Mathieu 2008). Though located in the direction of the Orion OB1a and OB1b subgroups, the σ Ori and 25 Ori sub-groups have different kinematic properties with

respect to the traditional association members (Briceño et al. 2007b; Jeffries et al. 2006); the λ Ori group (Mathieu 2008) formation could have been triggered by the expansion of the bubble created by Orion OB1a. Its age and distance from the center of OB1a are also similar to those of OB1c. More recently, Alves & Bouy (2012) and Bouy et al. (2014) reported the discovery of a young population of stars in the foreground of the ONC, which was however questioned by Da Rio et al. (2016), Fang et al. (2017) and Kounkel et al. (2017a). Finally, Kubiak et al. (2016) identified a rich and young population surrounding ϵ Ori.

In this study, we use the first *Gaia* data release (Gaia Collaboration et al. 2016b,a), hereafter *Gaia* DR1, to explore the three dimensional arrangement and the age ordering of the many stellar groups between the Sun and the Orion molecular clouds, with the overall goal to construct a new classification and characterization of the young, non-embedded stellar population in the region. Our approach is based on the parallaxes provided for stars brighter than $G \sim 12\,\mathrm{mag}$ in the *Tycho-Gaia Astrometric Solution* (TGAS Michalik et al. 2015; Lindegren et al. 2016) sub-set of the *Gaia* DR1 catalogue, and on the combination of *Gaia* DR1 and 2MASS photometry. These data are briefly described in Section 2. We find evidence for the presence of a young (age < 20 Myr) population, loosely clustered around some known groups: 25 Ori, ϵ Ori and σ Ori, and NGC 1980 and the ONC. We derive distances to these sub-groups and (relative) ages in Section 3. In Section 4 we use the Pan-STARRS1 photometric catalogue (Chambers et al. 2016) to confirm our age ranking. Our results, which we discuss in Section 5 and summarize in Section 6, are the first step in utilising *Gaia* data to unveil the complex star formation history of Orion and give a general overview of the episodes and the duration of the star formation processes in the entire region.

2.2 Data

The analysis presented in this study is based on the content of *Gaia* DR1 (Gaia Collaboration et al. 2016b; van Leeuwen et al. 2017), complemented with the photometric data from the 2MASS catalogue (Skrutskie et al. 2006) and the Pan-STARRS1 photometric catalogue (Chambers et al. 2016). Fig 2.1 shows the field selected for this study:

$$190° <= l <= 220°,$$
$$-30° <= b <= -5°. \tag{2.1}$$

We chose this field by slightly enlarging the region considered in de Zeeuw et al. (1999). We performed the cross-match using the *Gaia* archive (Marrese et al., in preparation). The query is reported in Appendix 2.B. In the cross-match with 2MASS, we included only the sources with photometry flag 'ph_qual = AAA' and we requested the angular distance of the cross-matched sources to be < 1". We decided to exclude from our analysis the sources that are either young stars inside the cloud or background galaxies. We performed this filtering with a $(J - K)$ vs $(H - K_s)$ color-magnitude diagram, where extincted sources are easily identified along the reddening

Table 2.1: Coordinates of the stars and clusters shown in Fig. 2.2.

Name	(l, b) [deg]
λ Ori	195, -12.0
25 Ori	201, -18.3
ϵ Ori	205.2 -17.2
σ Ori	206.8, -17.3
NGC 1980	209.5, -19.6
NGC 1981	208, -19.0
NGC 1977	208.4, -19.1

band. Following Alves & Bouy (2012), we required that:

$$J - H < -1.05\,(H - K_s) + 0.97\,\text{mag},$$
$$J < 15\,\text{mag},$$
$$H - K_s > -0.2\,\text{mag}, \ J - H < 0.74\,\text{mag}, \ H - K_s < 0.43\,\text{mag}. \qquad (2.2)$$

The first condition is taken as the border between non-extincted and extincted sources. The second is meant to reject faint sources to make the selection more robust against photometric errors. The third condition excludes sources with dubious infra-red colours (either bluer or redder than main sequence stars). The total number of *Gaia* sources in the field is $N = 9,926,756$. The number of stars resulting from the cross-match with 2MASS is $N = 5,059,068$, which further decreases to only $N = 1,450,911$ after applying the photometric selection. Fig. 2.2 shows a schematic representation of the field. The stellar groups relevant for this study are indicated as black empty circles and red stars. The coordinates of the stars and clusters shown are reported in Table 2.1. H_α emission (Finkbeiner 2003) is shown with blue contours , while dust structures (Planck Collaboration et al. 2014) are plotted in black.

2.3 Orion in *Gaia* DR1

In this section we identify and characterize the stellar population towards Orion. At first, we focus on the TGAS sub-sample and, after making a preliminary selection based on proper motions, we study the source distribution in parallax intervals. We notice the presence of an interesting concentration of sources towards the centre of the field, peaking roughly at parallax $\varpi = 2.65\,\text{mas}$ (Sec. 2.3.1). The sources belonging to this concentration also create a sequence in the color-magnitude diagrams made combining *Gaia* DR1 and 2MASS photometry (Sec. 2.3.2). These findings prompt us to look at the entire *Gaia* DR1. In the same color magnitude diagrams, we notice the presence of a young sequence, well visible between $G = 14\,\text{mag}$ and $G = 18\,\text{mag}$, which we interpret as the faint counterpart of the TGAS sequence. We make a preliminary selection of the sources belonging to the sequence, and we study their distribution in the sky, finding that they corresponded to the TGAS concentrations (Sec. 2.3.3). We refine our selection, and finally we determine the ages of the groups we identify (Sec. 2.3.4).
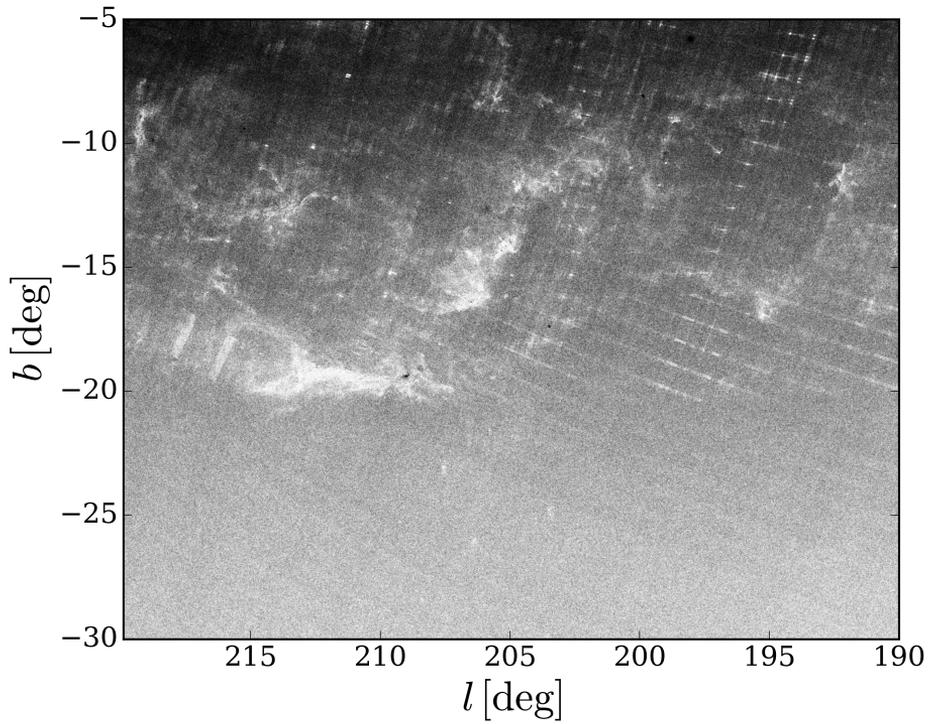
Figure 2.1: Sky area around the Orion constellation with the *Gaia* DR1 sources selected for this study. The number of stars shown in the figure is $N = 9926756$. The white areas correspond to the Orion A and B molecular clouds, centred respectively at $(l, b) \sim (212, -19)$ and $(l, b) = (206, -16)$. Well visible are also the λ Ori ring at $(l, b) \sim (196, -12)$ and Monoceros R2, at $(l, b) \sim (214, -13)$. The inclined stripes reflect the *Gaia* scanning law and correspond to patches in the sky where *Gaia* DR1 is highly incomplete (see Gaia Collaboration et al. 2016b).

Figure 2.2: Schematic representation of the field. The black contours correspond to the regions where $A_V >$ 2.5 mag (Planck Collaboration et al. 2014), while the blue contours show the H_α structures (Finkbeiner 2003): Barnard's loop and the λ Ori bubble. The positions of some known groups and stars are indicated with black circles and red stars, respectively.
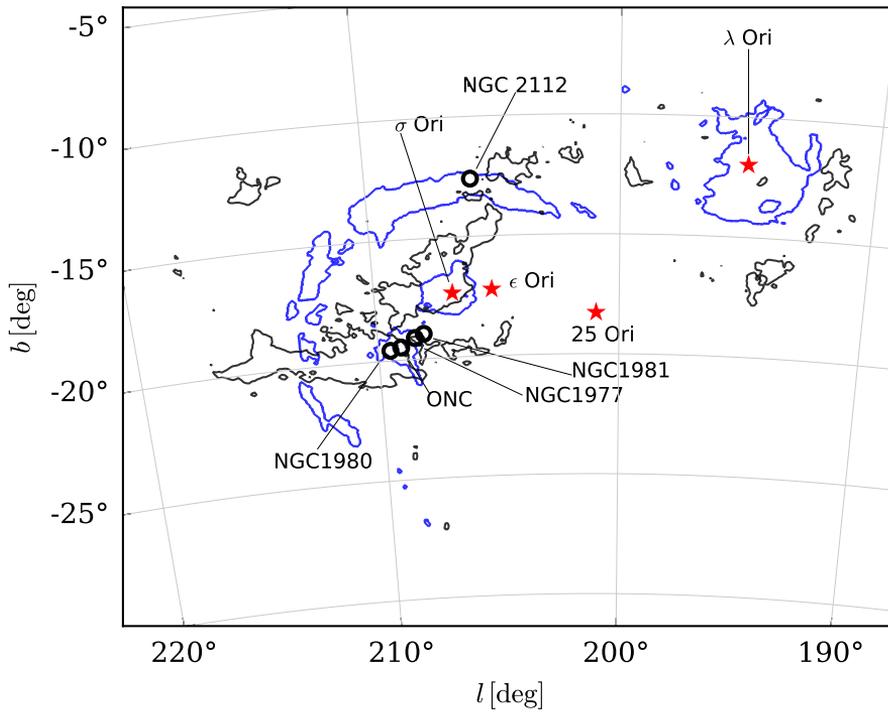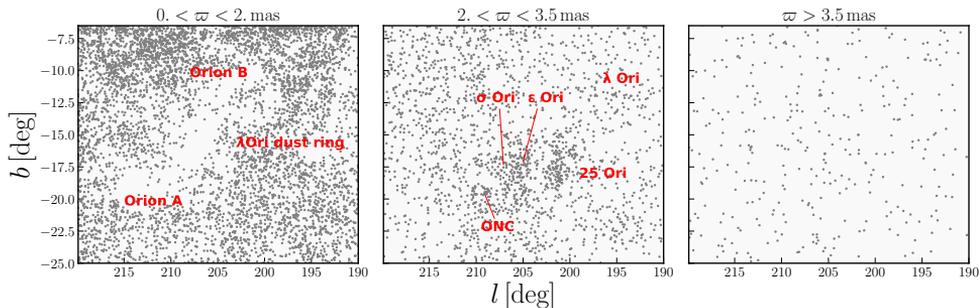
Figure 2.3: Positions in the sky of the TGAS sources selected with Eq. (2.3) in three different parallax intervals. The first panel shows stars with $0 < \varpi < 2.$ mas: the outlines of the Orion A and B molecular clouds and the λ Ori dust ring are visible as regions with a lack of sources. The second panel shows the stars with parallax $2. < \varpi < 3.5$ mas. Some density enhancements are visible towards the center of the field, $(l, b) \sim (205, -18)$. The third panel shows foreground sources, with $\varpi > 3.5$ mas.

2.3.1   Distances: the *Tycho-Gaia* sub-sample

Parallaxes and proper motions are available only for a sub-sample of *Gaia* DR1, namely the *Tycho-Gaia Astrometric Solution* (TGAS Michalik et al. 2015; Lindegren et al. 2016). We consider all the TGAS sources in the field. Since the motion of Orion OB1 is mostly directed radially away from the Sun, the observed proper motions are small. For this reason, a rough selection of the TGAS sources can be made requiring:

$$(\mu_{\alpha*} - 0.5)^2 + (\mu_\delta + 1)^2 < 25 \, \text{mas}^2 \, \text{yr}^{-2}, \tag{2.3}$$

where $\mu_{\alpha*}$ and μ_δ are the proper motions in right ascension and declination. The selection above follows roughly de Zeeuw et al. (1999). Fig. 2.3 shows the distribution in the sky of the sources selected with Eq. (2.3) as a function of their parallax ϖ, from small ($\varpi = 0$ mas) to large parallaxes up until $\varpi = 5$ mas (therefore until $d = 200$ pc). The outline of the Orion A and B clouds and of the λ Ori dust ring is visible (compare with Fig. 2.1) in the first panel, which show sources further away than $d = 500$ pc. This makes us confident that the sorting of sources in distance (through parallax) is correct. The second panel in Fig. 2.3 shows stars with parallax $2 < \varpi < 3.5$ mas, which corresponds to a distance $285 < d < 500$ pc. Some source over-densities towards the center of the field, $(l, b) \sim (205°, -18°)$, are clearly visible, and they are not due to projection effects but are indicative of real clustering in three dimensional space. We studied the distribution in the sky of the sources with parallaxes $2 < \varpi < 3.5$ mas using a Kernel Density Estimation (KDE). The KDE is a non-parametric way to estimate the probability density function of the distribution of the sources in the sky without any assumption on their distribution. Furthermore, it smooths the contribution of each data point over a local neighbourhood and it should therefore deliver a more robust estimate of the structure of the data and its density function. We used a multivariate normal kernel, with isotropic bandwidth $= 0.4°$. This value was chosen empirically as a good compromise between over- and under-smoothing physical density enhancements among random density fluctuations. To avoid projection distortions, we used a metric where the distance between two points on a curved surface
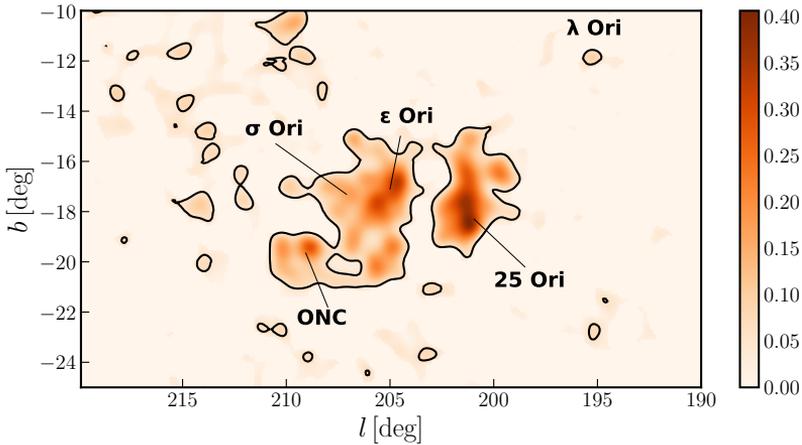
Figure 2.4: Kernel density estimation (Gaussian Kernel with bandwidth $0.4°$) of the TGAS sources with parallax $2 < \varpi < 3.5$ mas. The contours represent the $S = 3$ density levels.
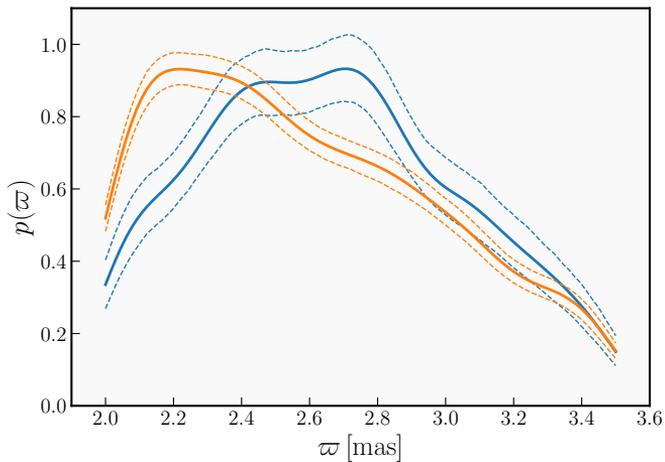
Figure 2.5: KDE of the parallax distribution of TGAS sources with $2 < \varpi < 3.5$ mas (orange thick dashed line) and of the sources belonging to the density enhancements defined in the text (blue thick solid line). The fine lines represent the 5th and 95th percentiles, and where computed with the bootstrapping procedure described in the text. The median value of the distribution is $\varpi \sim 2.65$ mas.

Figure 2.6: Median parallax of the sources within the TGAS $S = 3$ levels over bins of 1×1 degrees. Along $200° < l < 212°$ a gradient in the parallaxes is visible, suggesting that the density enhancements visible in Fig. 2.4 have different distances, with the one associated with 25 Ori being closer than the one towards NGC 1980. The λ Ori group is visible at $l \sim 195°$.

is determined by the haversine formula. The details of the procedure are described in Appendix C.

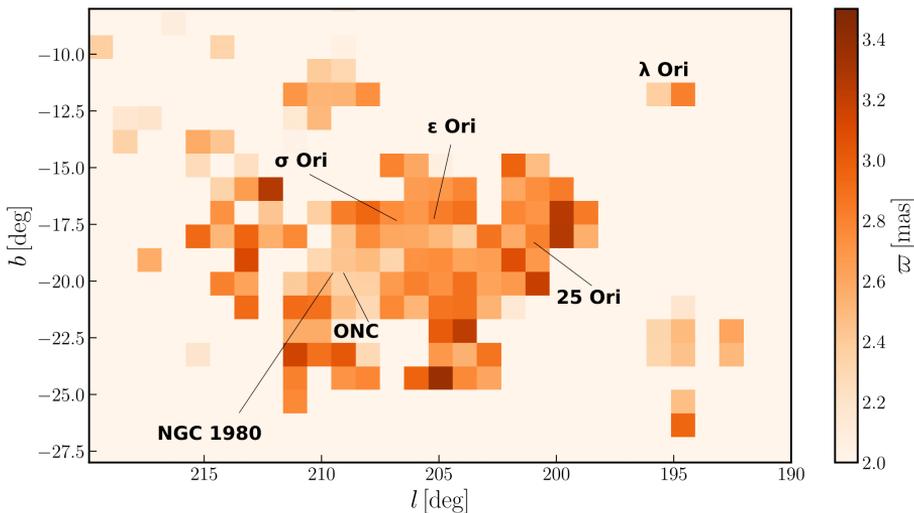To assess the significance of the density enhancements we assume that the field stars are distributed uniformly in longitude, while the source density varies in latitude. We thus average the source density over longitude along fixed latitude bins and we estimate the variance in source density using the same binning. The significance of the density enhancements is:

$$S(l,b) = \frac{D(l,b) - \langle D(b) \rangle}{\sqrt{\text{Var}\,(D(b))}} \tag{2.4}$$

where $D(l,b)$ is the density estimate obtained with the KDE, $\langle D(b) \rangle$ is the average density as a function of latitude, and $\text{Var}\,(D(b))$ is the variance per latitude. Fig. 2.4 shows the source probability density function, and the black contours represent the $S = 3$ levels. Fig. 2.5 shows the KDE of the parallax distribution of all the sources with $2 < \varpi < 3.5\,\text{mas}$ and of those within the $S = 3$ contour levels (solid blue and orange dashed line, respectively). We used a Gaussian Kernel with bandwidth = 0.1 mas, which is comparable to the average parallax error ($\sim 0.3\text{mas}$). The distribution of the sources within the $S = 3$ contour levels peaks at $\varpi \sim 2.65\,\text{mas}$. This supports the notion that the stars within the density enhancements are concentrated in space. To confirm the significance of the difference between the parallax distribution of the two samples, we performed $N = 1000$ realizations of the parallax density distribution (of both samples) by randomly sampling the single stellar parallaxes, then we computed the 5th and the 95th percentiles, which are shown as fine lines in 2.5. Finally, we noticed that the spread in the parallax distribution ($\sim 0.5\,\text{mas}$) is larger than the

typical parallax error, therefore we can hypothesize that it is due to an actual distance spread of $\sim 150\,\mathrm{pc}$, and not only to the dispersion induced by the errors.

Fig. 2.6 shows the median parallax over bins of $1° \times 1°$ for the sources within the $S = 3$ levels. The stars associated with 25 Ori have slightly larger parallaxes than those in the direction towards the ONC, which implies smaller distances from the Sun. We computed the median parallaxes in $2° \times 2°$ boxes centred in 25 Ori, ϵ Ori and the ONC. We obtained:

- 25 Ori: $\varpi = 2.81^{+0.46}_{-0.46}\,\mathrm{mas}\ (d \sim 355\,\mathrm{pc})$;

- ϵ Ori: $\varpi = 2.76^{+0.33}_{-0.35}\,\mathrm{mas}\ (d \sim 362\,\mathrm{pc})$;

- ONC: $\varpi = 2.42^{+0.2}_{-0.22}\,\mathrm{mas}\ (d \sim 413)$,

where the quoted errors correspond to the 16th and 84th percentiles.

These values are consistent with the photometric distances determined by Brown et al. (1994): $380 \pm 90\,\mathrm{pc}$ for Ori1a; $360 \pm 70\,\mathrm{pc}$ for Ori OB1b; and $400 \pm 90\mathrm{pc}$ for OB1c. Using the *Hipparcos* parallaxes de Zeeuw et al. (1999) reported the mean distances to be: $336 \pm 16\,\mathrm{pc}$ for Ori OB1a; $473 \pm 33\,\mathrm{pc}$ for Ori OB1b; and $506 \pm 37\mathrm{pc}$ for Ori OB1c. Distances to the Orion Nebula Cluster have been determined by, among others: Stassun et al. (2004); Hirota et al. (2007); Jeffries (2007); Menten et al. (2007); Sandstrom et al. (2007); Kim et al. (2008) and Kraus et al. (2009). These distance estimates range from $389^{+24}_{-21}\,\mathrm{pc}$ to $437 \pm 19\,\mathrm{pc}$. The latest distance estimate was obtained by Kounkel et al. (2017b), who found a distance of $388 \pm 5\,\mathrm{pc}$ using radio VLBA observations of Young Stellar Objects (YSOs). Thus the TGAS distances are quite in agreement with the estimates above.

2.3.2   Color magnitude diagrams

We combine *Gaia* and 2MASS photometry to make color-magnitude diagrams of the sources within the $S = 3$ levels defined in Fig. 2.4. These sources define a sequence at the bright end of the color-magnitude diagram (black big dots in Fig. 2.7, left). The spread of the sequence does not significantly change using apparent or absolute magnitudes. This prompts us to look further at the entire field, using the entire *Gaia* DR1 catalogue to find evidence of the faint counterpart of the concentration reported in Sec. 2.3.1. Fig.2.7 (left) shows a G vs. $G - J$ color magnitude diagram of the central region of the field, with coordinates:

$$195° < l < 212°,$$
$$-22° < b < -12°.$$

Fig. 2.7 (right) shows the same color magnitude diagram after unsharp masking. A dense, red sequence is visible between $G = 14\,\mathrm{mag}$ and $G = 18\,\mathrm{mag}$. This kind of sequence (also reported for example by Alves & Bouy 2012) indicates the presence of a population of young stars. Indeed, the locus of the sequence is situated above the main sequence at the distance of Orion. Several basic characteristics can be inferred from the diagram:

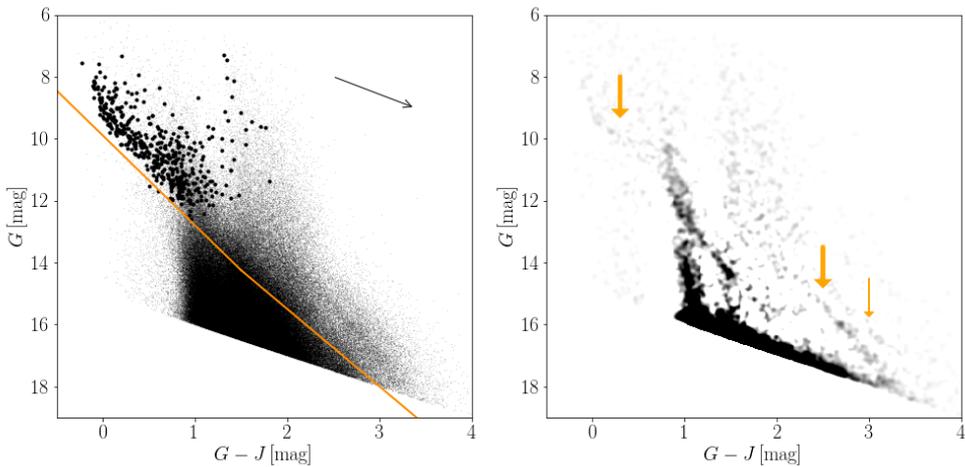1. The density of the sequence suggests that the population is rich;

Figure 2.7: Left: colour magnitude diagram of the *Gaia* sources cross matched with 2MASS. The sources we focused on are those responsible for the dense, red sequence in the lower part of the diagram. The orange line is defined in Eq. (2.5), and was used to separate the bulk of the field stars from the population we intended to study. The big black points represent the sources within the TGAS $S = 3$ contour levels of Fig. 2.4. The arrow shows the reddening vector corresponding to $A_V = 1\,\mathrm{mag}$. Right: same color magnitude diagram as on the left, after unsharp masking. The most interesting features (bright, TGAS sequence; faint *Gaia* DR1 sequence; binary sequence) are highlighted with the orange arrows.

2. The sequence appears not to be significantly affected by reddening, indicating that the sources are in front of or at the edges of the clouds;

3. The dispersion of the sequence is $\sim 0.5\,\mathrm{mag}$. This can be due to multiple reasons, such as: the presence of unresolved binaries, the presence of groups of different ages or distances, or of field contaminants.

Since our field is large, the number of contaminants is high. Therefore, we decided to eliminate the bulk of the field stars by requiring the following conditions to hold (orange line in Fig. 2.7 left):

$$G < 2.5\,(G - J) + 10.5 \ \text{ for } \ G > 14.25\,\mathrm{mag}$$
$$G < 2.9\,(G - J) + 9.9 \ \text{ for } \ G < 14.25\,\mathrm{mag}. \tag{2.5}$$

2.3.3 Source distribution

We choose to study the distribution in the sky of the sources selected with Eq. (2.5) repeating the procedure explained in Sec. 2.3.1. We analyse the source density using again a multivariate normal kernel, with isotropic bandwidth $= 0.3°$ and haversine metric. Fig. 2.8 shows the normalized probability density function of the source distribution on the sky. The dashed contours represent the $S = 3$ levels of the TGAS density map. The density enhancements towards the centre of the field are in the same direction as the groups shown in Fig. 2.2 and reported in Table 2.1. The density peak in $(l, b) \sim (206°, -12.5°)$ is associated to the old open cluster NGC 2112 (age $\sim$ 1.8 Gyr and distance $\sim 940\,\mathrm{pc}$, see e.g. Carraro et al. (2008) and references therein).
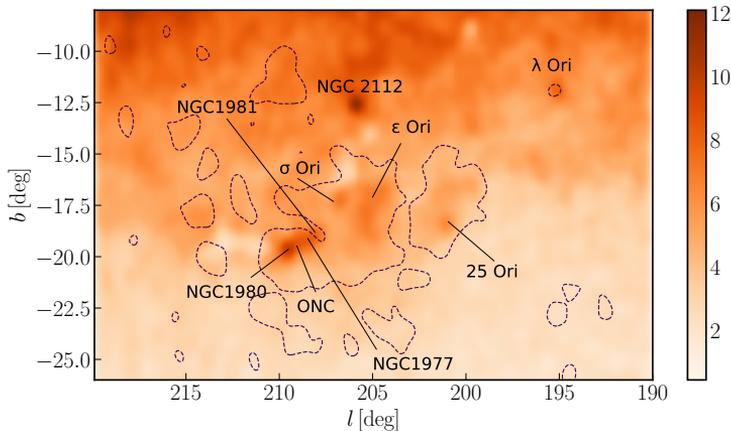
Figure 2.8: Normalized probability density function of the stars selected with with Eq. (2.5) (Gaussian kernel with bandwidth = 0.03°). The density enhancements visible in the centre of the field (Galactic longitude between 200° and 210°, Galactic latitude −20° and −15°) are related to the TGAS density enhancements (the black dashed contours correspond to the $S = 3$ levels of the TGAS density map of Fig. 2.4). The peak at $(l, b) \sim (206, -12.5)$ deg corresponds to the open cluster NGC 2112.

Fig. 2.9 shows $D(l, b) - \langle D \rangle$ (same notation as in Sec. 2.3.2), and the contours represent the $S = 1$ (gray) and $S = 2$ (black) significance levels. A certain degree of contamination is present, however the groups clearly separate from the field stars. Aside from the structures already highlighted in the TGAS map of Fig. 2.4, some other features are visible in the KDE of Fig. 2.9.

- The density enhancements towards λ Ori include not only the central cluster (Collinder 69, $\sim (195°, -12°)$) but also some structures probably related to Barnard 30 ($\sim 192°, -11.5°$) and LDN 1588 ($\sim 194.5°, -15.8°$). Some small over-densities are located on the Hα bubble to the left of LDN 1588 and they do not correspond to any previously known group.

- The shape of 25 Ori is elongated, and presents a northern and a southern 'extension', which are also present in the TGAS KDE of Fig. 2.4.

- South of ϵ Ori, a significant over-density is present, possibly related to the Orion X group, discovered by Bouy & Alves (2015).

- Around the centre of the Orion dust ring ($\sim 214°, -13°$) discovered by Schlafly et al. (2015) a number of densities enhancements are present. These over-densities are visible also in the TGAS map of Fig. 2.4, but here they are more evident.

For the following analysis steps, we selected all the sources related to the most significant density enhancements, i.e. those within the $S = 2$ contour levels shown in Fig. 2.9.
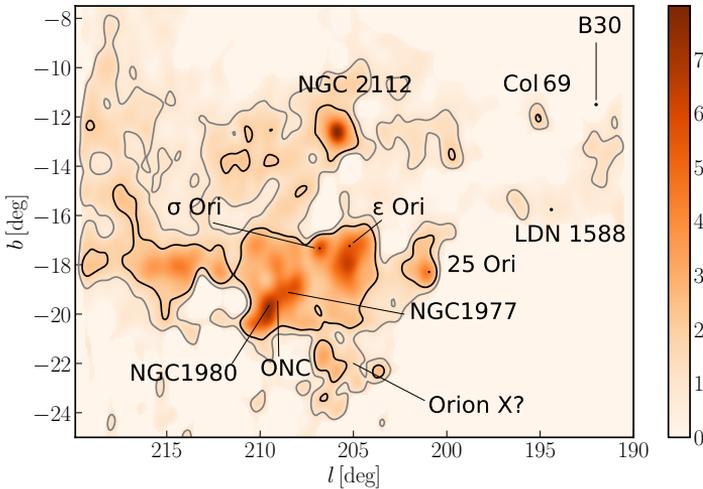
Figure 2.9: Background subtracted kernel density estimate of the sources selected through Eq. (2.5). The subtraction procedure is explained in Sec. 2.3.2. The density enhancements are highlighted by the contour levels, corresponding to $S = 1$ (gray) and $S = 2$ (black).

2.3.4 Age estimates

To determine the age(s) of the population(s) we identified, we perform a Bayesian isochrone fit using a method similar to the one described in Jørgensen & Lindegren (2005) and, more recently, in Valls-Gabaud (2014). These authors used Bayesian theory to derive stellar ages based on a comparison of observed data with theoretical isochrones. Age (t) is one free parameter of the problem, but not the only one: the initial stellar mass (m) and the chemical composition (Z) are also considered as model parameters. We simplify the problem assuming a fixed value for Z. Using the same notation as Jørgensen & Lindegren (2005), the posterior probability $f(t, m)$ for the age and mass is given by:

$$f(t, m) = f_0(t, m)L(t, m), \tag{2.6}$$

where $f_0(t, m)$ is the prior probability density and L the likelihood function. Integrating with respect to m gives the posterior probability function of the age of the star, $f(t)$. We assume independent Gaussian errors on all the observed quantities, with standard errors σ_i. The likelihood function is then:

$$L(t, m) = \prod_{i=1}^{n} \left(\frac{1}{(2\pi)^{1/2}\sigma_i} \right) \times \exp\left(-\chi^2/2 \right),$$

with:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{q_i^{\text{obs}} - q_i(t, m)}{\sigma_i} \right)^2,$$

where n is the number of observed quantities, and q^{obs} and $q(t, m)$ are the vectors of observed and modelled quantities. Following Jørgensen & Lindegren (2005), we write the prior as:

$$f_0(t, m) = \psi(t)\xi(m),$$

where $\psi(t)$ is the prior on the star formation history and $\xi(m)$ is the prior on the initial mass function. We assume a flat prior on the star formation history, and a power law for the initial mass function (IMF)

$$\xi(m) \propto m^{-a},$$

with $a = 2.7$. We choose a power law following Jørgensen & Lindegren (2005). We also test other IMFs, and find that the final results are not strongly dependent on the chosen IMF. We adopt the maximum of $f(t)$ as our best estimate of the stellar age. We compute the confidence interval following the procedure explained in detail in Jørgensen & Lindegren (2005). It might happen that the maximum of $f(t)$ coincides exactly with one of the extreme ages considered. In this case only an upper or a lower bound to the age can be set and we call our age estimate *ill defined*. On the other case, if the maximum of $f(t)$ falls within the age range considered, we call our age estimate *well defined*.

To perform the fit we compare the observed G magnitude and $G - J$ color to those predicted by the PARSEC (PAdova and TRieste Stellar Evolution Code Bressan et al. 2012; Chen et al. 2014; Tang et al. 2014) library of stellar evolutionary tracks. We used isochronal tracks from $\log(\text{age/yr}) = 6.0$ (1 Myr) to $\log(\text{age/yr}) = 8.5$ (200 Myr), with a step of $\log(\text{age/yr}) = 0.01$. We choose the range above since we are mainly interested in young (age < 20 Myr) sources. As mentioned above, we fixed the metallicity to $Z = 0.02$, following Brown et al. (1994). The isochronal tracks have an extinction correction of $A_V = 0.25$ mag. The correction was derived computing the average extinction towards the stars in Brown et al. (1994). We decided to fix the extinction to a single value mainly to keep the problem simple. Besides, we have excluded mostly of the extincted sources when we applied the criteria of Eq. 2.2.

We applied the fitting procedure to all the stars resulting from the selection procedure in Sec. 2.3.3, fixing the parallax to the mean value derived in the Sec. 2.3.1, i.e. $\varpi = 2.65$ mas. This choice is motivated primarily by the fact that with the current data quality is not possible to precisely disentangle the spatial structure of the region. More sophisticated choices for the parallax values are described in Appendix, however, even if they lead to different single age estimates, they do not change the general conclusions of the analysis. In particular the age ranking of the groups does not change.

Fig. 2.10 shows the color magnitude diagram of the sources with estimated age younger than 20 Myr. The gray crosses are the sources whose age is ill defined, the black dots represent the sources with well defined ages. Noteworthy, the sources with ill-defined age consist mainly of galactic contaminants, which we could then remove from our sample.

Fig. 2.11 shows the density (obtained with a Gaussian kernel, with bandwidth $= 0.05°$) of the source sky distribution as a function of their age, t. The densities are normalized to their individual maximum, so that their color scale is the same. The coordinates of the density enhancements change with time. This means that the groups we identified have different relative ages:

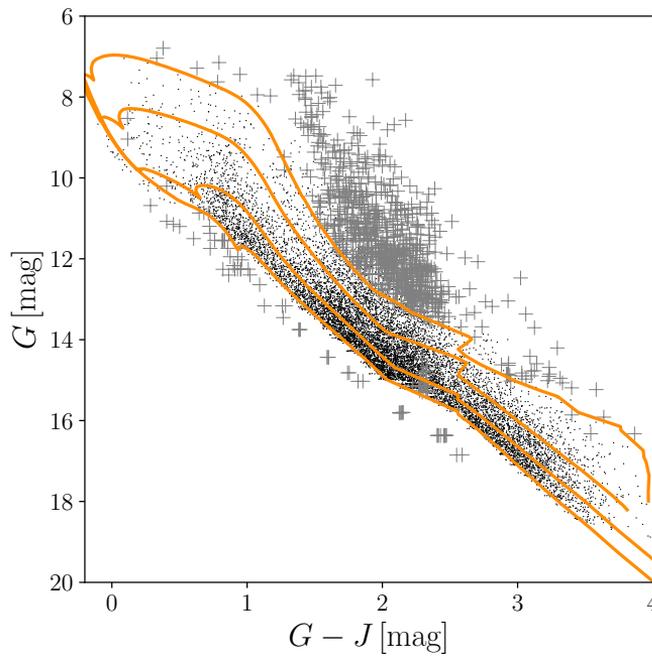**Figure 2.10:** Color magnitude diagrams of the sources with estimated age younger than 20 Myr. Black dots represent sources with well defined age estimate, gray crosses represent sources with ill-defined age estimate. The sources with ill-defined age estimates most likely belong to the Galactic disc. The orange lines are the PARSEC isochrones at 1, 3, 10 and 20 Myr at a distance of $\sim 380\,\mathrm{pc}$.

- *σ Ori.* The peak associated σ Ori $((l, b) = (207, -17.5)$ deg) is in the first panel $(1 < t < 3$ Myr), and some residuals are present also in the second panel $(3 < t < 5$ Myr) and in the fourth $(7 < t < 9$ Myr). Hernández et al. (2007), Sherry et al. (2008), and Zapatero Osorio et al. (2002) all estimate an age of 2- 4 Myr, which is compatible with what we find. Instead, Bell et al. (2013) puts the cluster at 6 Myr.

- *25 Ori.* The 25 Ori group $((l, b) = (20.1, -18.3)$ deg) appears in the third panel $(5 < t < 7$ Myr), peaks in the sixth panel $(9 < t < 11$ Myr) and then fades away. Briceño et al. (2007b) found that the age of 25 Ori is $\sim 7 - 10$ Myr. Our age estimate is slightly older, but still fits the picture of 25 Ori being the oldest group in the region.

- *Belt population.* The population towards ϵ Ori $((l, b) \sim (205.2, -17.2)$ deg) becomes prominent for $t > 9$ Myr. Here, Kubiak et al. (2016) estimated the age to be older than ~ 5 Myr, without any other constraint.

- *ONC, NGC 1980, NGC 1981, and NGC 1977.* The over-densities associated with NGC 1980, NGC 1981, NGC 1977 and the ONC (centred in $(l, b) \sim (209, -19.5)$ deg) are very prominent until the eighth panel of Fig. 2.11. In this last case it is difficult to disentangle exactly which group is younger, especially because the underlying data point distribution is smoothed by the Kernel. The density enhancement in the first panel $(1 < t < 3\,\mathrm{Myr})$ is most likely related to the ONC and L1641 (Reggiani et al. 2011; Da Rio et al. 2014, 2016). The density enhancement associated with NGC 1977 peaks in the same age ranges $(7 < t < 9\,\mathrm{Myr})$ as the one associated with NGC 1980, which however remains visible until later ages $(15 < t < 20\,\mathrm{Myr})$ and fades away only for $t > 20\,\mathrm{Myr}$. Finally, the density enhancement associated with NGC 1981 does not clearly stand out in any panel, excluding perhaps the ones with age $11 < t < 13\,\mathrm{Myr}$ and $13 < t < 15\,\mathrm{Myr}$. An interesting feature of the maps is the fact that the shape and position of the density enhancements related to NGC 1980 change with time. In particular, for early ages only one peak is present, while from ∼7 Myr two peaks are visible. This is a further confirmation that the density enhancements in the first three age panels include L1641 and the ONC, which are indeed younger than the other groups. Bouy et al. (2014) derived an age $\sim 5 - 10$ Myr for NGC 1980 and NGC 1981.

The last panel shows the stars with estimated ages > 20 Myr. The source distribution is uniform. These are field stars, with estimated ages ranging from 20 to 200 Myr.

Our fitting procedure does not take into account the presence of unresolved binaries among our data. Since the sample includes pre-main sequence stars, the binary population could be mistaken for a younger population at the same distance. For example, the binary counterpart of a population with age $t \sim 12$ Myr falls in the same locus of the $G - J$ vs G color magnitude diagram as a population with age $t \sim 7$ Myr. This means that the fit could mistake the unresolved binaries for a younger population, therefore the interpretation of Fig. 2.11 requires some care. Another caveat is related to the definition of the *Gaia G* band in the PARSEC libraries. Indeed, the nominal *Gaia G* passband (Jordi et al. 2010) implemented in the PARSEC libraries is different from the actual one (cfr. Carrasco et al. 2016). This affects the values of *G* and

G-J predicted by the PARSEC libraries and therefore our absolute age estimates, but does not influence the age ordering. The same can be said for the extinction. Choosing a different (constant) extinction value shifts the isochronal tracks, and therefore the estimated age is different, but does not modify the age ranking. In conclusion, the age ranking we obtain is robust, and, even with all the aforementioned cautions, Fig. 2.11 shows the potential of producing age maps for the Orion region.
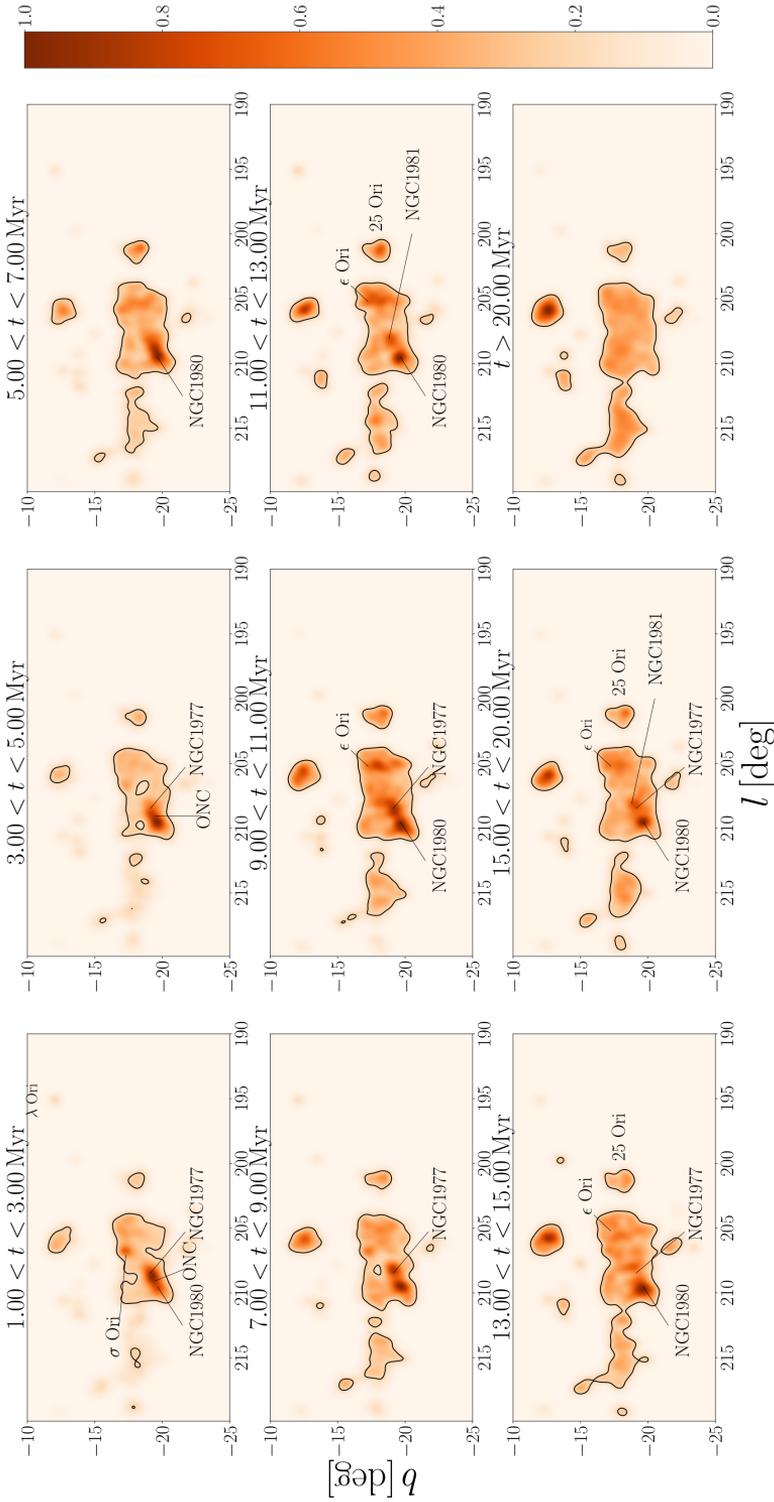
Figure 2.11: Distribution on the sky of the sources selected in Sec. 2.3.2 for different age intervals. The ages are computed using the isochrone fitting procedure described in Sec. 2.3.4. The contours represent the 0.05 density level and are shown only for visualization purposes. Note how the position of the density enhancements changes depending on the age. The first eight panels show stars with estimated ages < 20 Myr, while the last one shows older sources. The young stars are not coeval, in particular the age distribution shows a gradient, going from 25 Ori and ε Ori towards the ONC and NGC 1980. The last panel shows the field stars, whose estimated age is older than 20 Myr.

2.4 Orion in Pan-STARRS1

To confirm the age ordering we obtain with *Gaia* DR1, we apply the analysis described in Sec. 2.3 to the recently published Pan-STARRS1 photometric catalogue (Chambers et al. 2016; Magnier et al. 2016).

Pan-STARRS1 has carried out a set of distinct synoptic imaging sky surveys including the 3π Steradian Survey and the Medium Deep Survey in 5 bands ($grizy$). The mean 5σ point source limiting sensitivities in the stacked 3π Steradian Survey in $grizy$ are $(23.3, 23.2, 23.1, 22.3, 21.4)$ magnitudes respectively. For stars fainter than $r \sim 12\,\mathrm{mag}$, Pan-STARRS1 and *Gaia* DR1 photometric accuracies are comparable. Stars brighter than $r \sim 12\,\mathrm{mag}$ have large photometric errors in the PanSTARRS filters, therefore we decide to exclude them from our sample. We consider the same field defined in Eq. (2.1) and we perform a cross-match of the sources with *Gaia* DR1 and 2MASS, using a cross-match radius of 1″. We do not account for proper motions, since the mean epoch of the Pan-STARRS1 observations goes from 2008 to 2014 for the cross-matched stars and therefore the cross-match radius is larger than the distance covered in the sky by any star moving with an average proper motion of a few $\mathrm{mas\,yr^{-1}}$. We obtain $N = 88\,607$ cross-matched sources, and we analyse this sample with the same procedure explained in Sec. 3. Briefly, we first exclude the bulk of the field stars making a cut in the $r - i$ vs. r color-magnitude diagram:

$$r < 5 \times (r - i) + 12\,\mathrm{mag}. \tag{2.7}$$

Then we perform the same JHK photometric selection as in Eq. 2.5, and we study the on-sky distribution of the sources. We find some density enhancements, corresponding to those already investigated with the *Gaia* DR1 only. We then smooth the data point distribution in Galactic coordinates using a multivariate Gaussian kernel with bandwidth $0.3°$. We select all the sources within the $S = 2$ density levels and we estimate the single stellar ages with the same Bayesian fitting procedure described above. In this case however we do not use the *Gaia* and 2MASS photometry, but the r and i Pan-STARRS1 bands.

Fig. 2.12 shows the on-sky distribution of the sources with similar ages. The age intervals used are the same as in Fig. 2.11. The density enhancements corresponding to known groups are visible. Moreover, by comparing Figs. 2.11 and 2.12, one can immediately notice that the same groups appear in the same age intervals except for the ϵ Ori group, that appears slightly older than with *Gaia* DR1 photometry. Indeed the ϵ Ori density enhancement peaks in $15 < t < 20\,\mathrm{Myr}$ with PanSTARRS photometry, while it is spread between $11 < t < 20\,\mathrm{Myr}$ with *Gaia* DR1. Another interesting feature of the Pan-STARRS1 age maps are the density enhancements below ϵ Ori. These structures appear prominently in the oldest age panels, and might be related to the Orion X population (Bouy & Alves 2015).

These results strengthen our confidence in the age estimates obtained with *Gaia* photometry, in particular regarding the age ordering.
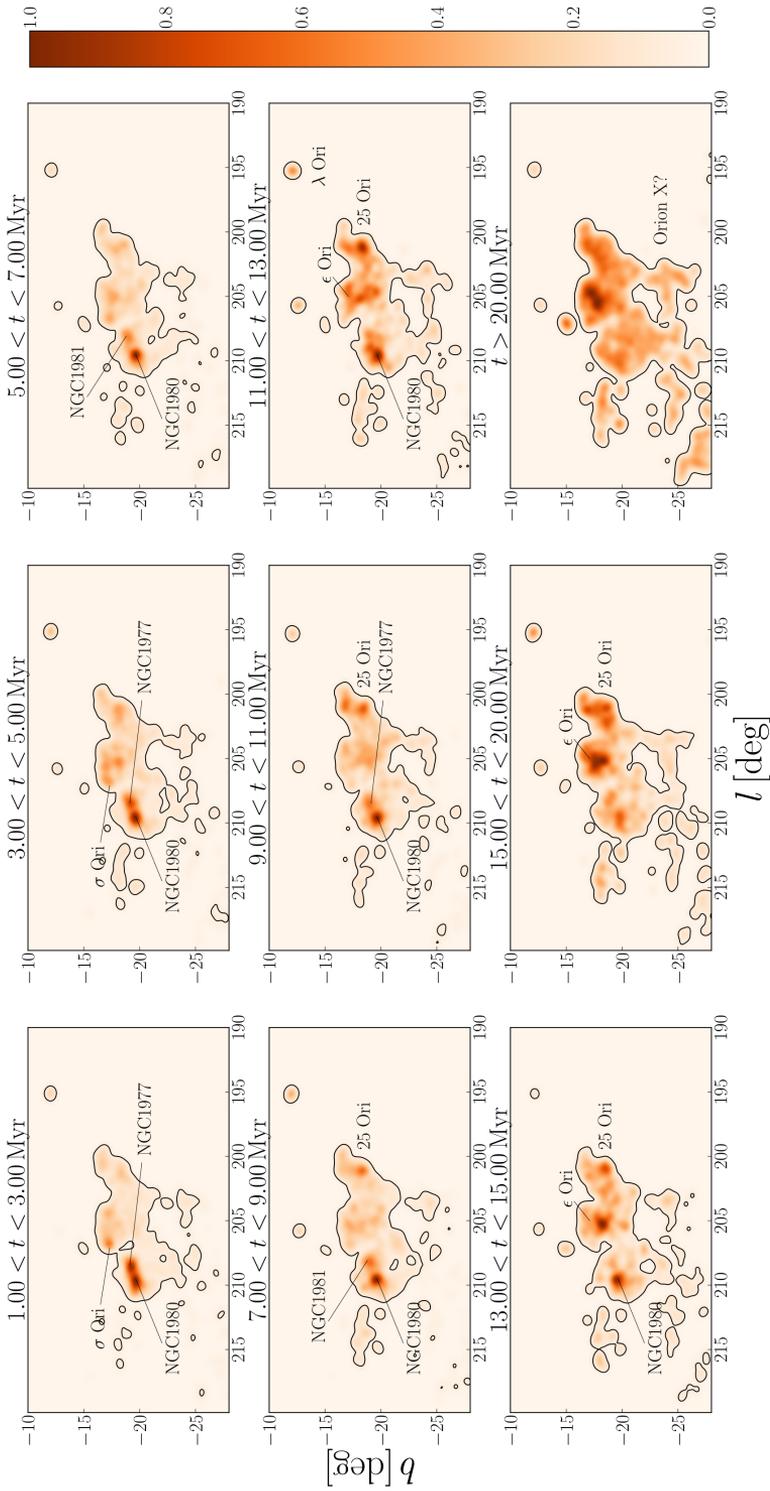
Figure 2.12: Same as Fig. 2.8 but using the Pan-STARRS1 r and i band to derive ages. The contours represent the 0.05 density levels and are shown only for visualization purposes.

2.5 Discussion

The present analysis confirms the presence of a large and diffuse young population towards Orion, whose average distance is $d \sim 380\,\mathrm{pc}$. The ages determined in Sec. 2.3.4 show that the groups are young (age < 20 Myr) and not coeval. The age ranking determined using *Gaia* and 2MASS photometry (Fig. 2.8) is consistent with that determined using Pan-STARRS1 (Fig. 2.12).
Figs. 2.9, 2.11, and 2.12 show some important features, which can potentially give new insights on our understanding of the Orion region.

The Orion dust ring. As already mentioned in Sec. 3.3, a number of over-densities are present towards the Orion dust ring discovered by Schlafly et al. (2015). The age analysis is not conclusive since many over-densities are not within $S = 2$. Unfortunately, there are no proper motions and/or parallaxes available for these sources (nor in *Gaia* DR1 nor in other surveys), and their distribution in the color-magnitude diagram is not very informative. Additional clues about their origin will be hopefully provided by *Gaia* DR2.

The Orion Blue-stream. Bouy & Alves (2015) studied the 3D spatial density of OB stars in the Solar neighbourhood and found three large stream-like structures, one of which is located towards $l \sim 200°$ in the Orion constellation (Orion X). Fig. 2.13 shows the position of the candidate members of the Orion X group as blue stars. Even though the candidate member centre looks slightly shifted with respect to the density enhancements shown in the map, it is difficult to argue that these stars are not related to the young population we analysed in this study. Bouy & Alves (2015) report that the parallax distribution of the Orion X sources goes from $\varpi \sim 3\,\mathrm{mas}$ to $\varpi \sim 6$ mas ($150 < d < 300\,\mathrm{pc}$), which indicates that Orion X is in the foreground of the Orion complex. Bouy & Alves (2015) also propose that the newly discovered complex could be older than Orion OB1 and therefore constitute the front edge of a stream of star formation propagating further away from the Sun.

To test this scenario we proceeded as follows. First we complemented the bright end of TGAS with Hipparcos data, then we selected the stars using the proper motion criterion of Eq. (2.3) and with $3 < \varpi < 7$ mas. In this way we restricted our sample to the stars probably kinematically related to the Orion OB association, but on average closer to the Sun. The density of the distribution of theses sources in the sky is shown in Fig. 2.13, together with the Orion X candidate members. We selected the sources within the S = 2 levels (with S defined in Section 3), and we used the Bayesian isochronal fitting procedure to estimate the age of this population. Note that out of the 48 Orion X candidate members listed in Bouy & Alves (2015), only 22 are included in TGAS (the others are probably too bright). To perform the isochronal fit, we could actually use the measured parallax, instead of one single value. The age distribution for the foreground sources is shown in Fig. 2.14 (orange histogram). As a comparison, the age distribution of the sources within the density enhancements and with $2 < \varpi < 3.5\,\mathrm{mas}$ is also shown (blue histogram). On average, the foreground population looks older, which is consistent with the picture that Bouy & Alves (2015) proposed. There are however two caveats:

- the age distributions are broad;

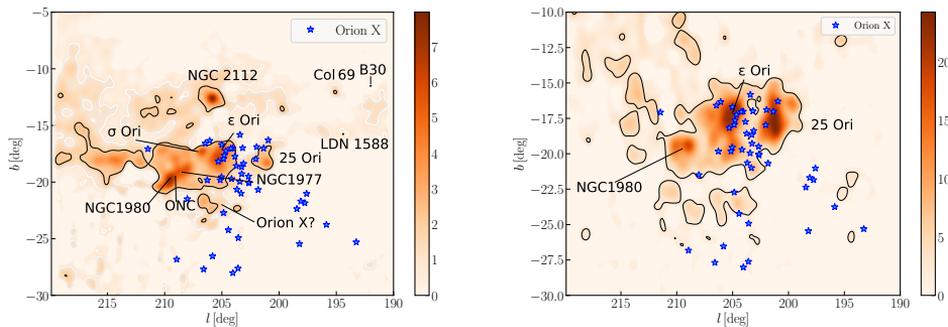- the parallax errors are large and dominate the age estimate.

Figure 2.13: Left: The Orion X candidate members from Bouy & Alves (2015) are plotted over the kernel density estimation of Fig. 2.9 as blue stars. Right: The Orion X candidate members are plotted over the kernel density estimation of the TGAS sources with $3 < \varpi < 7$ mas.

With future *Gaia* releases we will be able to further study the Orion X population and more precisely characterize it.

25 Ori. As pointed out in Sec. 3.3 the 25 Ori group presents a northern extension ($\sim 200°, -17°$) visible in the TGAS, Gaia DR1 and Pan-STARRS1 density maps. The northern extension parallax is only slightly larger than that of the 25 Ori group, and the age analysis suggests that the groups are coeval. With a different approach, Lombardi et al. (2017) find evidence of the same kind of structure (see their Fig. 15). *Gaia DR2* will be fundamental in discerning the properties of this new substructure of the 25 Ori group.

The λ Ori group. In Sec. 3.3 we pointed out some over-densities located on the Hα bubble surrounding λ Ori, which are not related to known groups (to our knowledge). We further investigated the stars belonging to these over-densities, however there are no parallaxes nor proper motions available for these sources and it is difficult to draw firm conclusions from the photometry only (also combining *Gaia* DR1 and Pan-STARRS1). In this case as well, we have to conclude that hopefully *Gaia* DR2 will clarify if this groups are real or not.

NGC 1980 and the ONC. One of the most interesting features of the maps of Fig. 2.11 and Fig. 2.12 is the prominent density enhancement towards NGC 1980, NGC 1977 and the ONC. The density enhancement is not concentrated in only one panel, but persists in all of them and disappears in the last one. This can be explained in at least two ways:

- there are multiple populations at roughly the same distance, with different ages;

- there is only one population with a single age, however its spread along the line of sight is so large that using only one parallax value for the fit is not accurate enough.

Both explanations have supporters. Alves & Bouy (2012) suggested that NGC 1980 is not directly related to the ONC, i.e. they are not the same population emerging from its parental cloud but are instead distinct overlapping populations. On the other hand, based on the fact that the kinematic properties of NGC 1980 are indistinguishable from those of the rest of the population at the same position in the sky, Da Rio
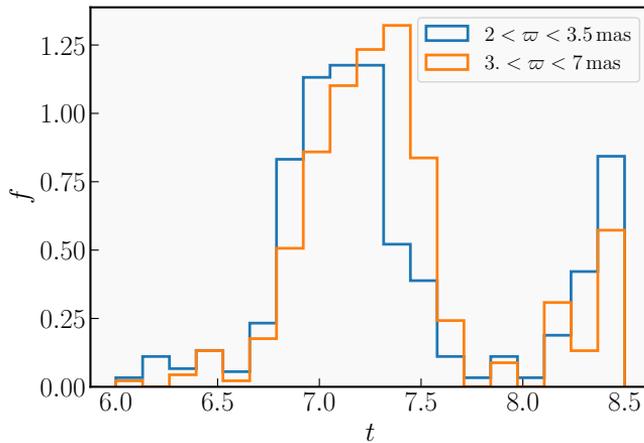
Figure 2.14: Age distribution of the TGAS sources with $2 < \varpi < 3.5\,\text{mas}$ (blue) and $3. < \varpi < 7\,\text{mas}$ (orange). The median of the distributions is respectively $t = 7.19\log(age/yr)$ ($\sim 15\,\text{Myr}$) and $t = 7.27\log(age/yr)$ ($\sim 19\,\text{Myr}$).

et al. (2016) argued that NGC 1980 simply represents the older tail of the age distribution around the ONC, in the context of an extended star formation event. Using isochronal ages, Fang et al. (2017) find that the foreground population has a median age of 1-2 Myr, which is similar to that of the other young stars in Orion A. Furthermore they confirm that the kinematics of the foreground population is similar to that of the molecular clouds and of other young stars in the region. They therefore argue against the presence of a large foreground cluster in front of Orion A. Kounkel et al. (2017a) estimate that the age of NGC 1980 is $\sim 3\,\text{Myr}$, which is comparable with the study by Fang et al. (2017), however they are not able to confirm or disprove whether NGC 1980 is in the foreground on the ONC. Finally, Beccari et al. (2017) discovered three well-separated pre-main sequences in the $r - i$ vs r color-magnitude diagram obtained with the data of the wide field optical camera OmegaCAM on the VLT Survey Telescope (VST) in a region around the ONC. These sequences can be explained as a population of unresolved binaries or as three populations with different ages. The populations studied by Beccari et al. are unlikely to be related to NGC 1980, however, if confirmed, they would constitute an example of non-coeval populations in the same cluster. Fig. 2.11 shows that the group corresponding to NGC 1980 is well defined not only at very young ages ($1 < t < 3\,\text{Myr}$), but at least until $t \sim 15\text{Myr}$. We will discuss below the influence that unresolved binaries have on our age determination (indeed our fit does not account for them), the main point being that unresolved binaries influence the youngest age intervals, not the oldest. This would point towards the actual existence of two populations, the first related to the ONC, the second to the Alves & Bouy (2012) foreground population.

In conclusion, the ages of the stellar populations towards Orion show a gradient, which goes from 25 Ori and ϵ Ori towards the ONC and the Orion A and B clouds. The age gradient is also associated to a parallax gradient: indeed the older population towards 25 Ori and ϵ Ori is also closer to the Sun than the younger one towards the

ONC (see also Fig. 2.6). *Gaia* DR2 will provide distances to the individual stars of each different group, and we will therefore be able to obtain also more precise ages for them.

To study whether or not the parallax gradient influences the age determination, we performed the same Bayesian isochrone fit changing each star's parallax according to its position, following Fig. 2.6. We also performed the analysis including a uniform prior on the parallax distribution, and then marginalizing over the parallax. In both cases, the estimated ages for the single groups have some small variations, however our conclusions do not significantly change.

To test how our result depend on the set of isochrones we chose, we performed the fit again, using the MESA Isochrones and Stellar Tracks (MIST Dotter 2016; Choi et al. 2016). We fixed the metallicity to $Z_\odot$ and we applied the usual extinction correction of $A_V = 0.25\,\mathrm{mag}$. Whilst in this case the single ages are in general estimated to be younger than with the PARSEC models (e.g., the 25 Ori group peaks between $9 < t < 11\,\mathrm{Myr}$), the age ordering does not change significantly.

Finally, we studied the distribution in the sky of the coeval sources fainter than 14 magnitudes. In this magnitude range we can remove the sources that are most likely Galactic contaminants. We found again the same groups and the same age ordering.

As mentioned above and in Sec. 2.3.4, the unresolved binary sequence could stand out as a separate, seemingly younger population, which would add further complications to the age determination of the group. The Bayesian fitting procedure does not take into account the presence of unresolved binaries. The net effect of this is that the unresolved binaries population is mistaken for a younger population. For example, the difference in magnitudes between the 5-7 Myr and the 13-20 Myr isochrones correspond almost exactly to the 0.75 mag separating the primary sequence from the unresolved binary sequence. This is a major cause of age spread and it could greatly affect our age estimates, thus it appears even clearer that great care needs to be used when analysing them. On the other hand however, binary should affect all populations in the same way. This further support the robustness of our relative age estimates.

Another intriguing problem is related to the relation between the the density enhancements, the diffusely distributed massive stars, and the gas distribution. Fig. 2.15 shows the $S = 3, 6$ and 9 contour levels of the over-densities on top of an extinction map obtained with *Planck* data (Planck Collaboration et al. 2014) probing the dark clouds. The older group 25 Ori is located far away from the gas, while the younger groups of λ Ori, σ Ori, ϵ Ori, NGC 1977 and NGC 1980 closely follow the clouds. Orion A and B are behind the density enhancements. The three dimensional structure of the region is still unclear, and the current data accuracy is not yet good enough to draw definite conclusions, especially at the distance and direction of Orion. The data quality however will improve in future *Gaia* releases, and likewise our understanding of the region. In particular, precise parallaxes, proper motions, and radial velocities will allow us to address directly the recent discovery that the Orion clouds might be part of an ancient dust ring (Schlafly et al. 2015), the blue streams scenario proposed by Bouy & Alves (2015), and the complex nested shell picture unveiled by Ochsendorf et al. (2015).
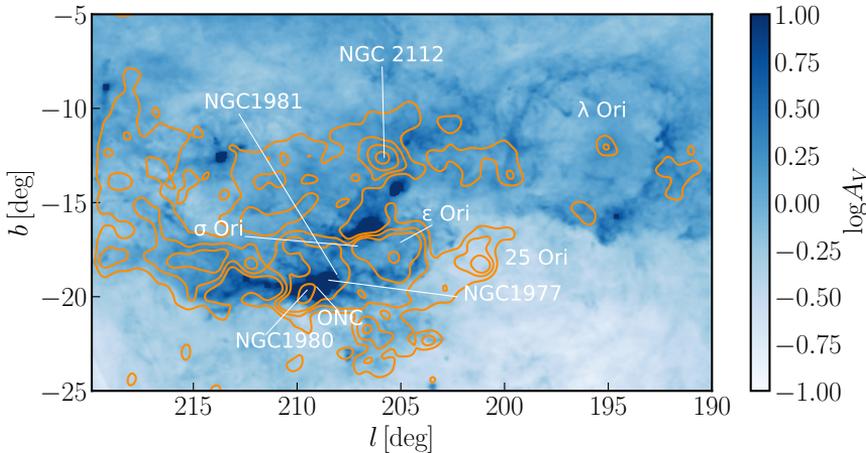
Figure 2.15: Planck extinction map of the Orion field (Planck Collaboration et al. 2014). The contour levels represent the $S = 1, 2, 3$ and 6 levels of the density distribution shown in Fig. 2.8.

2.6 Conclusions

In this paper we made use of *Gaia* DR1 (Gaia Collaboration et al. 2016b,a; van Leeuwen et al. 2017) to study the stellar populations towards Orion. Our results are as follows:

- Using TGAS (Michalik et al. 2015; Lindegren et al. 2016) we found evidence for the presence of a young population, at a parallax $\varpi \sim 2.65\,\mathrm{mas}$ ($d \sim 377\,\mathrm{pc}$), loosely distributed around some known clusters: 25 Ori, ϵ Ori and σ Ori, and NGC 1980 and ONC. The stars belonging to these groupings define a sequence in all the color magnitude diagrams constructed by combining *Gaia* DR1 and 2MASS photometry.

- We considered the entire *Gaia* DR1, again realizing color magnitude diagrams combining *Gaia* and 2MASS photometry for the entire field. Well visible between $G = 14\,\mathrm{mag}$ and $G = 18\,\mathrm{mag}$, we found the low mass counterpart of the sources isolated with TGAS.

- After a preliminary selection to exclude field stars, we studied the distribution in the sky of the sources belonging to this sequence using a Kernel Density Estimation (KDE). We found density enhancements in the sky distribution comparable to those in the TGAS sample.

- We estimated the ages of the sources within the density enhancements, using a Bayesian isochrone fitting procedure described in detail in Jørgensen & Lindegren (2005). We assumed all the stars to be at the same parallax, $\varpi = 2.65$. We found that the groupings have different ages. In particular, there is an age gradient going from 25 Ori (13-15 Myr) to the ONC (1 Myr).

- To consolidate our findings, we repeated the fitting procedure using the sources

in common with Pan-STARRS1 (Chambers et al. 2016; Magnier et al. 2016) r and i filters, finding the same age ordering as with *Gaia* DR1.

- We studied the distribution in the sky of the groups we found. In particular:

  1. The 25 Ori cluster presents a northern extension, reported also by Lombardi et al. (2017).

  2. Some of the density enhancements towards the λ Ori complex are related to known clusters (Col 69, B30, and LDN 1588), but some other over-densities on the left of the ring are new. Unfortunately it was not possible to investigate them further since we have neither precise proper motions, nor parallaxes.

  3. Some over-densities are also present within the Orion dust ring discovered by Schlafly et al. (2015), and they might be related to the star formation process out of which the ring was formed. In this case as well however, more data are needed to confirm our speculations.

  4. The Orion X candidate members (Bouy & Alves 2015) are related to some of the density enhancements shown in Fig. 2.11. We studied the sky and age distribution of the TGAS sources with proper motions as in Eq. (2.3) and parallax $3 < \varpi < 7\,\mathrm{mas}$, and we found that the stars with $2 < \varpi < 3.5\,\mathrm{mas}$ are on average younger than those with $3 < \varpi < 7\,\mathrm{mas}$.

- We discussed the implications of the age ranking we obtained. We found that the estimated ages towards the NGC 1980 cluster span a broad range of values. This can either be due to the presence of two populations coming from two different episodes of star formation or to a large spread along the line of sight of the same population. Some confusion might arise also from the presence of unresolved binaries, which are not modelled in the fit, and usually stand out as a younger population. We related our findings to previous works by Bouy et al. (2014); Da Rio et al. (2016) and Fang et al. (2017).

- Finally, we link the stellar groups to the gas and dust features in Orion, albeit in a qualitative and preliminary fashion. Future *Gaia* releases will allow to address these questions in unparalleled detail.

2.A   Color-magnitude and color-color diagrams

In this Appendix we show the color-color and color-magnitude diagrams constructed combining *Gaia* DR1 and 2MASS photometry. The sources in the first panel are those remaining after applying the 2MASS photometry quality selection cut ('ph_qual = AAA'). The other panels show the cuts of Eq. (2.2). Note that we did not apply exactly the same photometric criteria as in Alves & Bouy (2012) because there is probably a typo in their Eq. (1) that causes 0 sources to be selected. However, Fig. 2.16 looks similar to their Fig. 4.
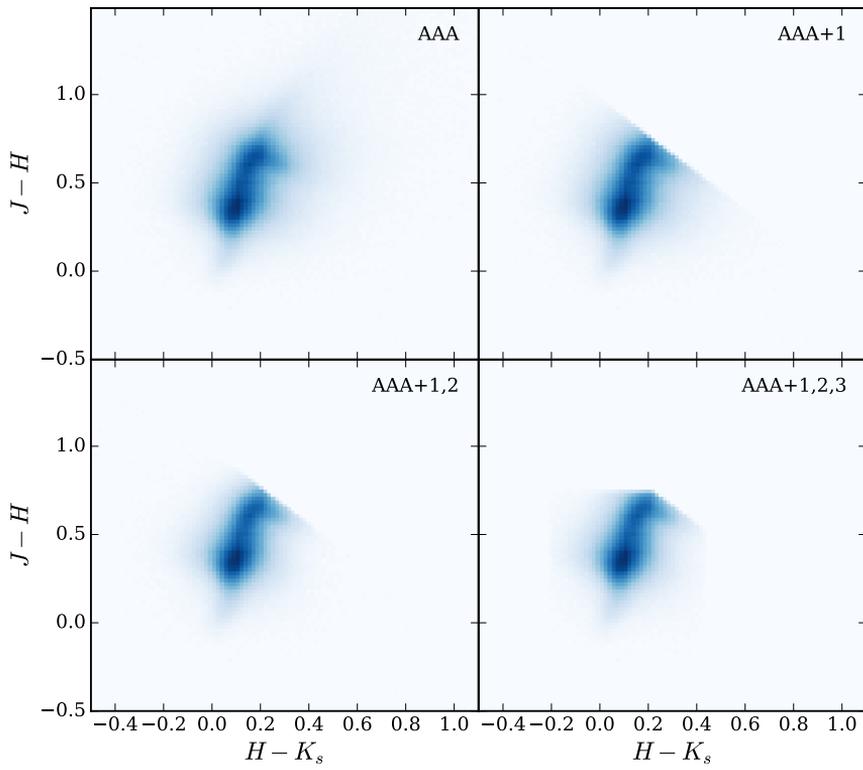
Figure 2.16: Color-color diagrams of the sources resulting from the selection criteria in Section 2.
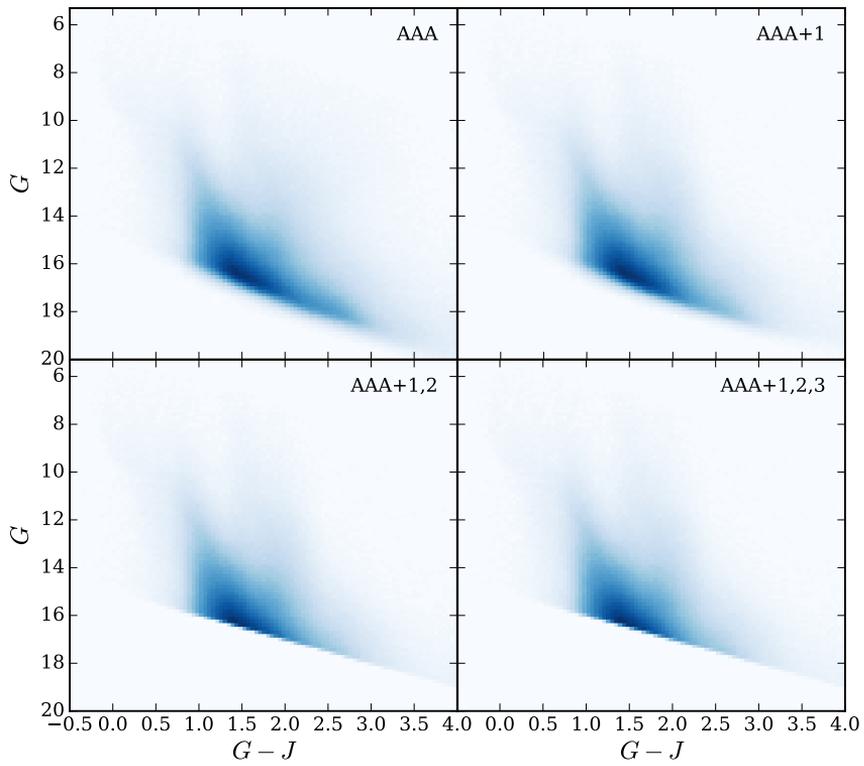
Figure 2.17: Color-magnitude diagrams of the sources resulting from the selection criteria in Section 2.

2.B   ADQL queries

We report here the queries used to a) select the sources in our field and b) perform the cross-match with 2MASS.

Field selection:
```
select gaia.source_id, gaia.ra, gaia.dec, gaia.l, gaia.b,
gaia.phot_g_mean_mag, gaia.pmra, gaia.pmdec, gaia.parallax,
gaia.pmra_error, gaia.pmdec_error, gaia.parallax_error
from gaiadr1.gaia_source as gaia
where gaia.l>=190.0 and gaia.l<=220.0 and
gaia.b>=-30.0 and gaia.b<=-5.0
```

Cross Match with 2MASS:
```
select gaia.source_id, gaia.l, gaia.b, gaia.phot_g_mean_mag,
gaia.phot_g_mean_flux, gaia.phot_g_mean_flux_error,
gaia.parallax,gaia.parallax_error, gaia.pmra,gaia.pmdec,
gaia.pmra_error, gaia.pmdec_error,
tmass.j_m, tmass.j_msigcom, tmass.h_m, tmass.h_msigcom,
tmass.ks_m, tmass.ks_msigcom, tmass.ph_qual
from gaiadr1.gaia_source as gaia
inner join gaiadr1.tmass_best_neighbour as xmatch
on gaia.source_id = xmatch.source_id
inner join gaiadr1.tmass_original_valid as tmass
on tmass.tmass_oid = xmatch.tmass_oid
where gaia.l > 190.0 and gaia.l < 220.0
and gaia.b < -5.0 and gaia.b > -30.0 and
xmatch.angular_distance < 1.0
```

We run the queries using the *Gaia* archive. On the archive, we suggest the user to create a personal account. This indeed allows to save queries and store data (up to 1 GB).

2.C   Kernel Density Estimation on the sphere

The referee pointed out that the kernel density estimation carried out on flat projections of the Orion sky field will suffer from area distortions, and suggested the use of the von Mises-Fisher (vMF) kernel, which is intended for analyses on the unit sphere. This kernel is given by the following equation for a two-dimensional unit sphere (i.e., for three-dimensional unit vectors x):

$$f(\mathrm{x}|\mathrm{m}, \kappa) = \frac{\kappa^{1/2}}{(2\pi)^{3/2} I_{\frac{1}{2}}(\kappa)} e^{\kappa \mathrm{m}^\top \mathrm{x}}, \tag{2.8}$$

where the unit vector m represents the mean direction for the kernel and $I_{\frac{1}{2}}$ is the modified Bessel function of the first kind and order $1/2$, and $\mathrm{m}^{\mathsf{T}}\mathrm{x}$ indicates the inner product of m and x. See Banerjee et al. (2005) for details. The parameter $\kappa \geqslant 0$ indicates the concentration of the kernel around m. The normalization constant can be re-written by considering that:

$$I_{\frac{1}{2}}(z) = \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \sinh z \qquad (2.9)$$

(eq. 10.39.1 in Olver et al. 2010), which leads to:

$$f(\mathrm{x}|\mathrm{m},\kappa) = \frac{\kappa}{4\pi \sinh \kappa} e^{\kappa \mathrm{m}^{\mathsf{T}}\mathrm{x}} = \frac{\kappa}{2\pi(e^{\kappa} - e^{-\kappa})} e^{\kappa \mathrm{m}^{\mathsf{T}}\mathrm{x}} . \qquad (2.10)$$

The exponent in the kernel contains the inner product of m and x and this can also be written as:

$$\mathrm{m}^{\mathsf{T}}\mathrm{x} = \cos \rho = \sin \delta_m \sin \delta + \cos(\alpha - \alpha_m) \cos \delta_m \cos \delta , \qquad (2.11)$$

where (α, δ) represent the ICRS coordinates of the points on the sky, and ρ is the angle between m and x. On the unit sphere this angle also represents the distance along a great circle between the points m and x, also known as the 'haversine distance'. The value of ρ can also be calculated using the haversine function (hav) given by:

$$\mathrm{hav}(\theta) = \sin^2 \left(\frac{\theta}{2}\right) = \frac{1 - \cos \theta}{2} . \qquad (2.12)$$

The formula for ρ then becomes:

$$\mathrm{hav}(\rho) = \mathrm{hav}(\delta - \delta_m) + \cos \delta_m \cos \delta \, \mathrm{hav}(\alpha - \alpha_m) . \qquad (2.13)$$

This can be verified by writing out both sides of the equation in terms of $(1 - \cos \theta)/2$.

To continue, we note that the half-width at half maximum of the vMF kernel expressed in terms of ρ (ρ_{HWHM}) is given by:

$$\rho_{\mathrm{HWHM}} = \arccos \left(1 - \frac{\ln 2}{\kappa}\right) , \qquad (2.14)$$

where Eq. 2.14 follows from:

$$e^{\kappa \cos \rho_{\mathrm{HWHM}}} = \frac{e^{\kappa}}{2}, \qquad (2.15)$$

as the maximum of $f(\mathrm{x}|\mathrm{m},\kappa)$ occurs when $\cos \rho = 1$.

Equivalently, for a given ρ_{HWHM} the corresponding value of κ is:

$$\kappa = \frac{\ln 2}{1 - \cos \rho_{HWHM}} . \qquad (2.16)$$

In our kernel density estimates of source distributions on the sky the kernel sizes are of order 1 degree (0.017 radians) or less. This is already in the regime where to good accuracy $\cos \rho \approx 1 - \rho^2/2$. At the same time the value of κ becomes very large (~ 4550,

see Eq. 2.16 for $\rho_{\mathrm{HWHM}} = 0.017$), such that $\sinh \kappa \to \exp(\kappa)/2$. Hence the vMF kernel becomes approximately:

$$f(\mathrm{x}|\mathrm{m}, \kappa) \approx \frac{\kappa}{2\pi} e^{-\frac{\kappa}{2}\rho^2} . \tag{2.17}$$

This is in fact a 2D Normal distribution with standard deviations $\sigma = 1/\sqrt{\kappa}$ along the two principal axes, where in the small angle regime one can write $\rho^2 = (\Delta\alpha \cos \delta)^2 + \Delta\delta^2$, with $\Delta\alpha = \alpha - \alpha_m$ and $\Delta\delta = \delta - \delta_m$. This shows that in our case (with kernel sizes of a degree or less), the vMF kernel can be approximated as a 2D Gaussian in terms of the haversine distance.

Our implementation of the kernel density estimate is in Python and makes use of the the the `sklearn.neighbors.KernelDensity` module in the `scikit-learn` package by specifying that the 'haversine' metric should be used during the fitting stage of the density estimate (using the parameters `kernel='gaussian'` and `metric='haversine'`).

3

Structure, kinematics, and ages of the young stellar populations in the Orion region

We present a study of the three dimensional structure, kinematics, and age distribution of the Orion OB association, based on the second data release of the *Gaia* satellite (*Gaia* DR2). Our goal is to obtain a complete picture of the star formation history of the Orion complex and to relate our findings to theories of sequential and triggered star formation. We selected the Orion population with simple photometric criteria, and we constructed a three dimensional map in galactic Cartesian co-ordinates to study the physical arrangement of the stellar clusters in the Orion region. The map shows structures that extend for roughly $150\,\mathrm{pc}$ along the line of sight, divided in multiple sub-clusters. We separated different groups by using the density-based clustering algorithm DBSCAN. We studied the kinematic properties of all the groups found by DBSCAN first by inspecting their proper motion distribution, and then by applying a kinematic modeling code based on an iterative maximum likelihood approach, which we used to derive their mean velocity, velocity dispersion, and isotropic expansion. We provide ages and extinction values for all the groups by using an isochrone fitting procedure. We confirm the presence of an old population ($\sim 15\,\mathrm{Myr}$) towards the 25 Ori region, and we find that groups with ages of $12 - 15\,\mathrm{Myr}$ are present also towards the Belt region. A population of $\sim 10\,\mathrm{Myr}$ is also present in front of the Orion A molecular cloud. Our findings suggest that star formation in Orion does not follow a simple sequential scenario, but instead consists of multiple events, which caused kinematic and physical sub-structure. To fully explain the detailed sequence of events, specific simulations and further radial velocity data are needed.

3.1   Introduction

The tendency of O and B type stars to loosely cluster in the sky was recognised at the beginning of the 20th century by the pioneering studies summarised in Blaauw (1964). At the end of the last century, the data of the *Hipparcos* satellite allowed de Zeeuw et al. (1999), de Bruijne (1999a), Hoogerwerf & Aguilar (1999), and many others to characterise the stellar content and the kinematic properties of nearby OB associations. OB associations have been long considered as expanding remnants of your star clusters as their members are widely dispersed over the sky (Brown et al. 1999; Lada & Lada 2003). The classical explanation for this is that star clusters are formed embedded within molecular clouds, where the gravitational potential of both the stars and the gas holds them together. When feedback disperses the gas left over from star formation, the cluster becomes supervirial and will expand and disperse, thus being visible for a short time as an OB association. While many observations support this model (Lada & Lada 2003, and references therein), it has been difficult to test whether OB associations are indeed expanding. Wright et al. (2016) and Wright & Mamajek (2018) studied the kinematics of the Cygnus OB2 and Scorpius-Centaurus associations respectively, and concluded that they were not formed by the disruption of individual star clusters. Wright & Mamajek (2018) further concluded that Sco-Cen was likely born highly sub-structured, with multiple small-scale star formation events contributing to the overall OB association, and not as a single, monolithic burst of clustered star formation. These conclusions can be related to the fact that the distribution of young stars within their parental molecular clouds is fractal, hierarchical, and follows the filamentary structures of the dense gas, both spatially (Gutermuth et al. 2008) and kinematically (Hacar et al. 2016). Clusters then form where filaments overlap (Myers 2009; Schneider et al. 2012; Hacar et al. 2016, 2017): their formation might be due to higher column densities or to the merging of filaments that have already formed stars. OB associations would therefore constitute the final stage of this star formation mechanism. They still keep memory of the parental gas sub-structure where they originated as they slowly disperse in the field.

   At a distance of $\sim 380\,\mathrm{pc}$ (Zari et al. 2017), the Orion star forming region is the nearest site of active high-mass star formation. It is a benchmark for studying all stages and modes of star formation (Brown et al. 1994; Jeffries et al. 2006; Bally 2008; Briceno 2008; Muench et al. 2008; Da Rio et al. 2014; Getman et al. 2014; Da Rio et al. 2016; Hacar et al. 2016; Kubiak et al. 2017; Fang et al. 2017; Kounkel et al. 2017a), in addition to the effect of star formation processes on the surrounding interstellar medium (Ochsendorf et al. 2015; Schlafly et al. 2015; Soler et al. 2018). Zari et al. (2017) used *Gaia* DR1 (Gaia Collaboration et al. 2016b,a) to study the density distribution of the young, non-embedded stellar population in the sky, and obtained a first picture of the star formation history of the Orion region in terms of the various star formation episodes, their duration, and their effects on the surrounding interstellar medium. Even though proper motions where available for the *Tycho-Gaia* astrometric solution (TGAS, Michalik et al. 2015) sub-set of *Gaia* DR1, they were not accurate enough to perform a precise kinematic analysis. Proper motions in Orion are indeed small as stars move on average radially away from the Sun. Furthermore, to derive the ages of the stellar populations, a single distance value was considered ($d \sim 380\,\mathrm{pc}$) as parallax uncertainties were too large to resolve the spatial configuration of the groups

that were identified. By combining the data of the second release of the *Gaia* satellite (hereafter *Gaia* DR2 Gaia Collaboration et al. 2018a) and APOGEE-2, Kounkel et al. (2018) study the entire Orion complex, providing a classification of the stellar population in five groups, and an analysis of their ages and kinematics. Kos et al. (2018) use *Gaia* DR2 parameters supplemented with radial velocities from the GALAH and APOGEE surveys to perform a clustering analysis towards the 25 Ori cluster region. They find that one cluster is significantly older ($21 \pm 2\,\mathrm{Myr}$) compared to the rest of the region. Großschedl et al. (2018) investigate the 3D shape and orientation of the Orion A molecular cloud by analysing the distances of mid-infrared selected young stellar objects, and find that the cloud is elongated and oriented towards the galactic plane, and presents two different components one dense and star forming and one $\sim 75\,\mathrm{pc}$ long, more diffuse and star-formation quiet.

In this work, we use *Gaia* DR2 to study the three dimensional (3D) structure of the Orion OB association, we model the kinematics of the sub-groups that constitute it and we give estimates of their ages, to obtain a complete picture of the star formation history of the region and to put it in the broader context of the theories of sequential and triggered star formation. In Section 2 we present the data and describe how we select the young stellar population in Orion. In Section 3 we study its 3D configuration in Cartesian galactic co-ordinates, and we isolate young groups by making use of the DBSCAN clustering algorithm. In Section 4 we perform the kinematic analysis by using a maximum likelihood approach. In Section 5 we derive ages and extinctions of all the groups resulting from the analysis of Section 4. In Section 6 we discuss our findings. The conclusions of this work are summarised in Section 7.

3.2 Data

Following Zari et al. (2017), we select the sources with co-ordinates

$$190° < l < 220°, \qquad -30° < b < -5°, \tag{3.1}$$

and we restrict our sample to the sources with $1.5 < \varpi < 5.0\,\mathrm{mas}$. Since the Orion population moves mostly radially away from the Sun, we consider only stars with small proper motions:

$$|\mu_{\alpha*}| < 10\,\mathrm{mas\,yr^{-1}}, \qquad |\mu_\delta| < 10\,\mathrm{mas\,yr^{-1}}. \tag{3.2}$$

We derive distances by inverting parallaxes, $d = 1000/\varpi\,\mathrm{pc}$ thus we restrict our sample to sources with $\varpi/\sigma_\varpi > 5.0$, following the recommendations in Bailer-Jones (2015). The effect of this cut is to exclude sources at faint magnitudes ($G > 20\,\mathrm{mag}$), but it does not introduce significant biases in the determination of distances to the clusters or the study of their 3D configuration.
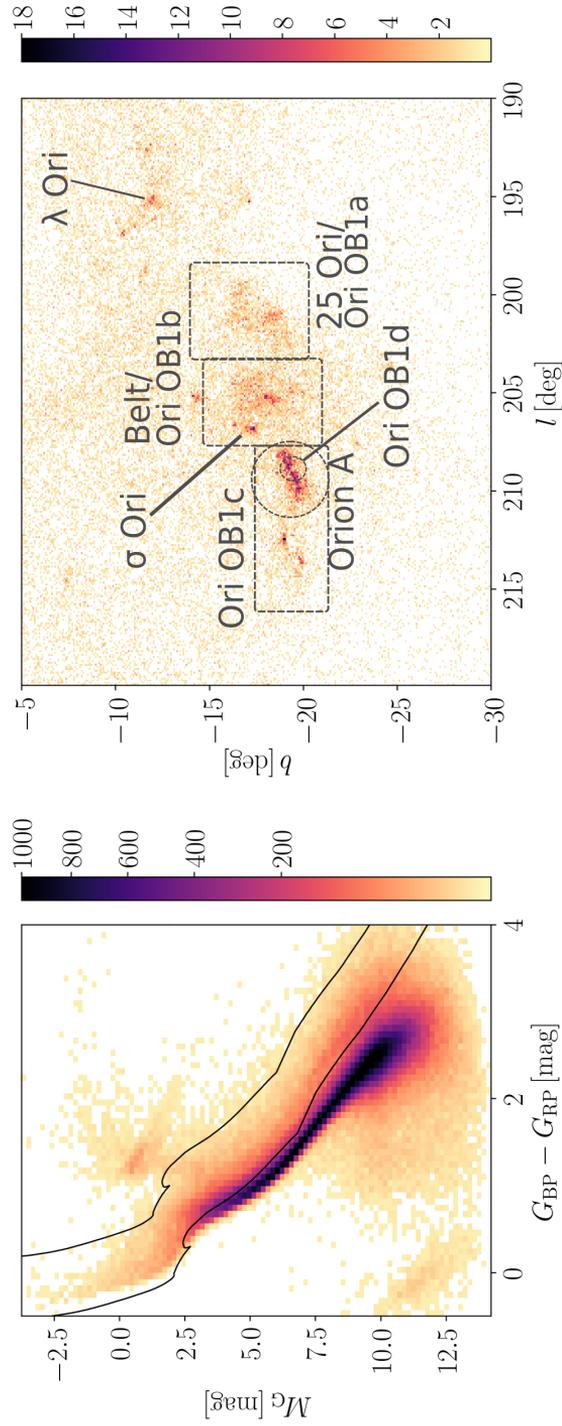
Figure 3.1: Observed colour-magnitude diagram (left) and sky distribution (right) of the sources selected in Section 2. The solid black lines in the left panel show the isochrones defined in Eq. 3.6, which are used to select the young stellar population in Orion.

3.2.1 Obtaining a 'clean' sample

We apply the following cuts on the photometric and astrometric quality, based on Lindegren et al. (2018) complemented by the information contained on the *Gaia* known issues page (`https://www.cosmos.esa.int/web/gaia/dr2-known-issues`). We select all the sources with RUWE < 1.4, following the slides by Lindegren et al. (see `https://www.cosmos.esa.int/web/gaia/dr2-known-issues`).
The renormalized unit weight error (RUWE) is defined as:

$$\text{RUWE} = \sqrt{\chi^2/(N-5)}/u_0(C,G) \tag{3.3}$$

where: χ^2 is the astrometric goodness-of-fit in the AL direction (`astrometric_chi2_al`); N is the number of good observations AL (`astrometric_n_good_obs_al`); $u_0(C,G)$ is an empirical normalization factor, which is a function of $C = G_{\text{BP}} - G_{\text{RP}}$ and G. This cut seeks to remove sources with spurious parallaxes or proper motions.
We use the flux excess ratio:

$$E = (I_{\text{BP}} + I_{\text{RP}})/I_G, \tag{3.4}$$

where I_{X} is the photometric flux in band X, to exclude sources with possible issues in the BP and RP photometry, affecting in particular faint sources in crowded areas. We apply Eq. C.2 in Lindegren et al. (2018), which we report here for clarity:

$$1.0 + 0.015(G_{\text{BP}} - G_{\text{RP}})^2 < E < 1.3 + 0.06(G_{\text{BP}} - G_{\text{RP}})^2. \tag{3.5}$$

Evans et al. (2018) and Arenou et al. (2018) mention that *Gaia* DR2 photometry is affected by some systematic errors. Evans et al. (2018) and Maíz Apellániz & Weiler (2018) propose corrections to mitigate these effects. We apply these corrections and we report them here for clarity:

- $2 \leqslant G \leqslant 6$ mag:
  $G_{\text{corr}} = -0.047344 + 1.16405G - 0.046799G^2 + 0.0035015G^3$

- $2 \leqslant G \leqslant 4$ mag:
  $G_{\text{BP,corr}} = G_{\text{BP}} - 2.0384 + 0.95282G - 0.11018G^2$

- $2 \leqslant G \leqslant 3.5$ mag:
  $G_{\text{RP,corr}} = G_{\text{RP}} - 13.946 + 14.239G_{\text{RP}} - 4.23G_{\text{RP}}^2 + 0.4532G_{\text{RP}}^3$

- $6 \leqslant G \leqslant 16$ mag:
  $G_{\text{corr}} = G - 0.0032(G - 6.0)$

- $G > 16$ mag:
  $G_{\text{corr}} = G - 0.032$

In the rest of the paper we use the corrected G, G_{BP}, and G_{RP} magnitudes without using the subscript "*corr*".

3.2.2   Selecting the young stellar population

Figure 4.2 (left) shows the M_G vs. $G_{BP} - G_{RP}$ colour-magnitude diagram of the 'clean' sample obtained in Section 2.1. Although faint, the pre-main sequence and the upper main sequence, indicating the presence of the young population in the region, are visible, and can be used to guide the selection of the young stellar populations towards Orion.

To select young stars, we use the PARSEC isochrones (Bressan et al. 2012; Tang et al. 2014; Chen et al. 2014) with $A_V = 0.3$ mag and age $\tau = 10$ Myr to define the following region in the M_G vs. $G_{BP} - G_{RP}$ colour-magnitude diagram (solid black lines in Fig. 4.2):

$$G_{BP} - G_{RP} - 0.2 \leqslant M_G \tag{3.6}$$
$$G_{BP} - G_{RP} + 0.5 \geqslant M_G - 0.8$$

We choose $A_V = 0.3$ mag following Zari et al. (2017). The distribution in the sky of the sources selected in this fashion is shown in Fig. 4.2 (right). The regions in which we divide the field are also indicated, together with the sub-groups in which the Orion OB1 association is classically split: Orion OB1a, OB1b, OB1c, and OB1d. The same groups identified in Zari et al. (2017) and Kounkel et al. (2018) are visible, which confirms the correctness of the selection.

In Section 4 we focus on the kinematics of the Orion population. To complement the *Gaia* DR2 radial velocities we cross-matched our sources with the APOGEE DR14 catalogue (Abolfathi et al. 2018). The APOGEE synthetic heliocentric velocities (SYNTH-VHELIO_AVG, an average of the individual measured RVs using spectra cross-correlations with single best-match synthetic spectrum) were used.

3.3   3D distribution and identification of clusters

We first study the three-dimensional (3D) distribution of sources using a similar approach as in Zari et al. (2018). In summary, we:

1. compute galactic Cartesian co-ordinates for all the sources, x_g, y_g, z_g;

2. define a volume, V = (800, 800, 350), centred in the Sun, and we divide it in $3 \times 3 \times 3$ pc cubes;

3. compute the number of sources in each cube;

4. compute the source density $D(x_g, y_g, z_g)$ by smoothing the distribution with a Gaussian filter, with width $w = 2$ pc;

5. normalise the density distribution from 0 to 1 by applying the sigmoidal logistic function:
$$f(D) = \frac{L}{1 + e^{-\kappa(D - D_0)}} - 1, \tag{3.7}$$
   with $L = 2$, $\kappa = 4$ pc, and $D_0 = 0$.

Fig. 4.10 shows the density distribution of sources $f(D)$ on the galactic plane for different values of z_g. Different density enhancements are visible, corresponding to well known-clusters. The first and second panel show stars in the Orion A molecular cloud. The Orion Nebula Cluster (ONC) corresponds to the most prominent density enhancement. The third panel is particularly interesting because it clearly shows the presence of a foreground population to the ONC, confirming the conclusions by Bouy et al. (2014). Some clusters corresponding to the Belt region also become visible, although the bulk of the population is located between $Z = -116\,\text{pc}$ and $Z = -101\,\text{pc}$. The last three panels mainly show the λ Ori cluster. At $Z = -92\,\text{pc}$ the northern elongation of the 25 Ori group is visible. The density distribution looks elongated towards the line of sight: this is an effect of the parallax errors. The parallax error distribution is peaked at $\sigma_\varpi = 0.046\,\text{mas}$, but presents a long tail towards larger values (the $84th$ percentile is $0.11\,\text{mas}$).

To isolate the members of each cluster, we first consider only the sources within the density level $f(D) = 0.5$ of the 3D map shown in Fig. 4.10. This value is arbitrary and aims at selecting the densest regions of the maps. The clusters are then separated by using the DBSCAN algorithm [1]. As described for instance by Price-Jones & Bovy (2019), DBSCAN is a density-based clustering algorithm that views clusters as areas of high density separated by areas of low density in space, without requiring any prior assumption on the number of groups present. There are two parameters to the algorithm, `min_samples` and `eps`, which define the density of the clusters. Higher `min_samples` or lower `eps` values indicate higher densities necessary to form a cluster. Clusters in Orion have different sizes and numbers of members, and therefore different densities: for this reason we need to apply the clustering algorithm twice. The first time we use `min_samples`= 50 and `eps`= 7 pc to isolate the main structures, shown in Figs. 4.9 and 3.4 (top), obtaining five groups. The group that encompasses 25 Ori, the Belt region and the Orion A foreground can be visibly divided in sub-groups. Thus we apply DBSCAN only to this group with different paramenters: we find that `min_samples`= 30 and `eps`= 5 pc are the best values to separate all the sub-clusters (see Figs. 4.9 and 3.4, bottom).

This method has the drawback of excluding stars that might be related to the star formation events in Orion, but are more dispersed than the rest of the population in 3D space (but could still be compact in proper motion space). This is further discussed in Section 6.

---

[1]We use the `scikit-learn` implementation of the algorithm (Pedregosa et al. 2011)
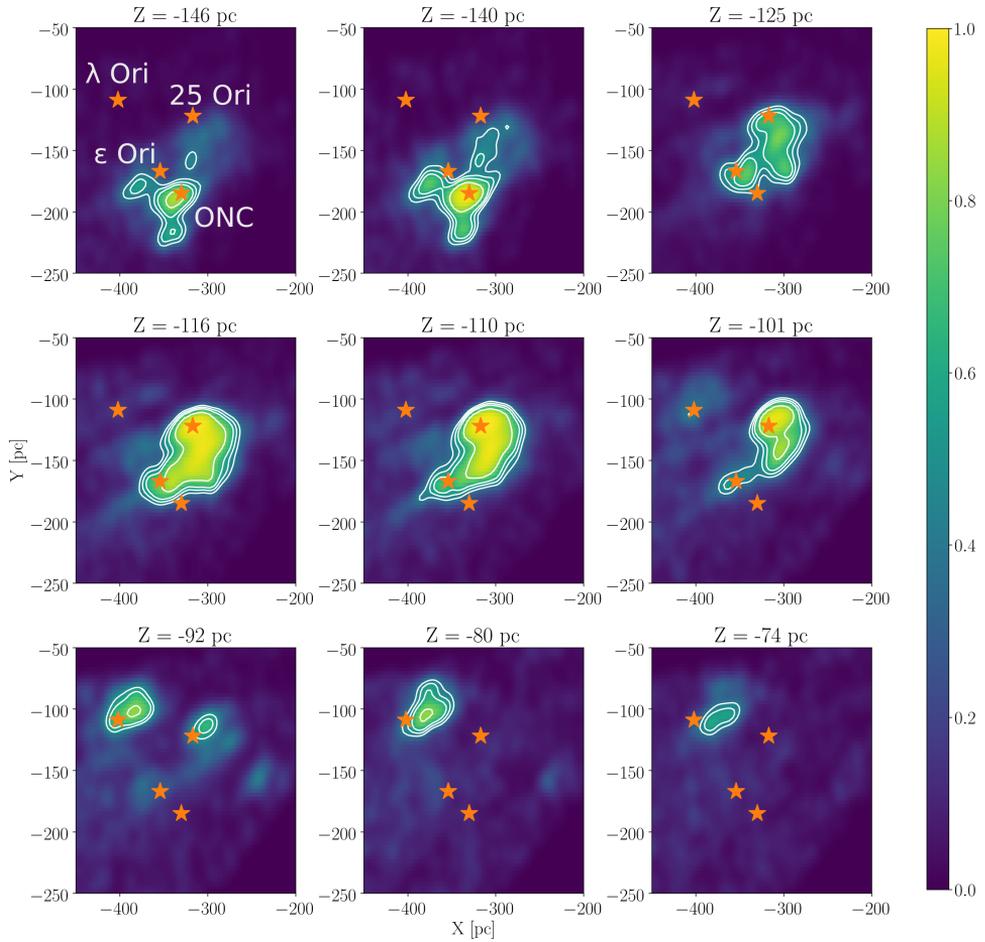
Figure 3.2: Density distribution of the sources in Orion for different Z values. The orange stars indicate the positions of (from top to bottom in each panel): λ Ori, 25 Ori, ϵ Ori, and the Orion Nebula cluster. The white solid contours represent the 0.4, 0.5, 0.6 and 0.8 density levels (the density is normalised to have values from 0 to 1). The Sun is located at $(X, Y) = (0, 0)$.
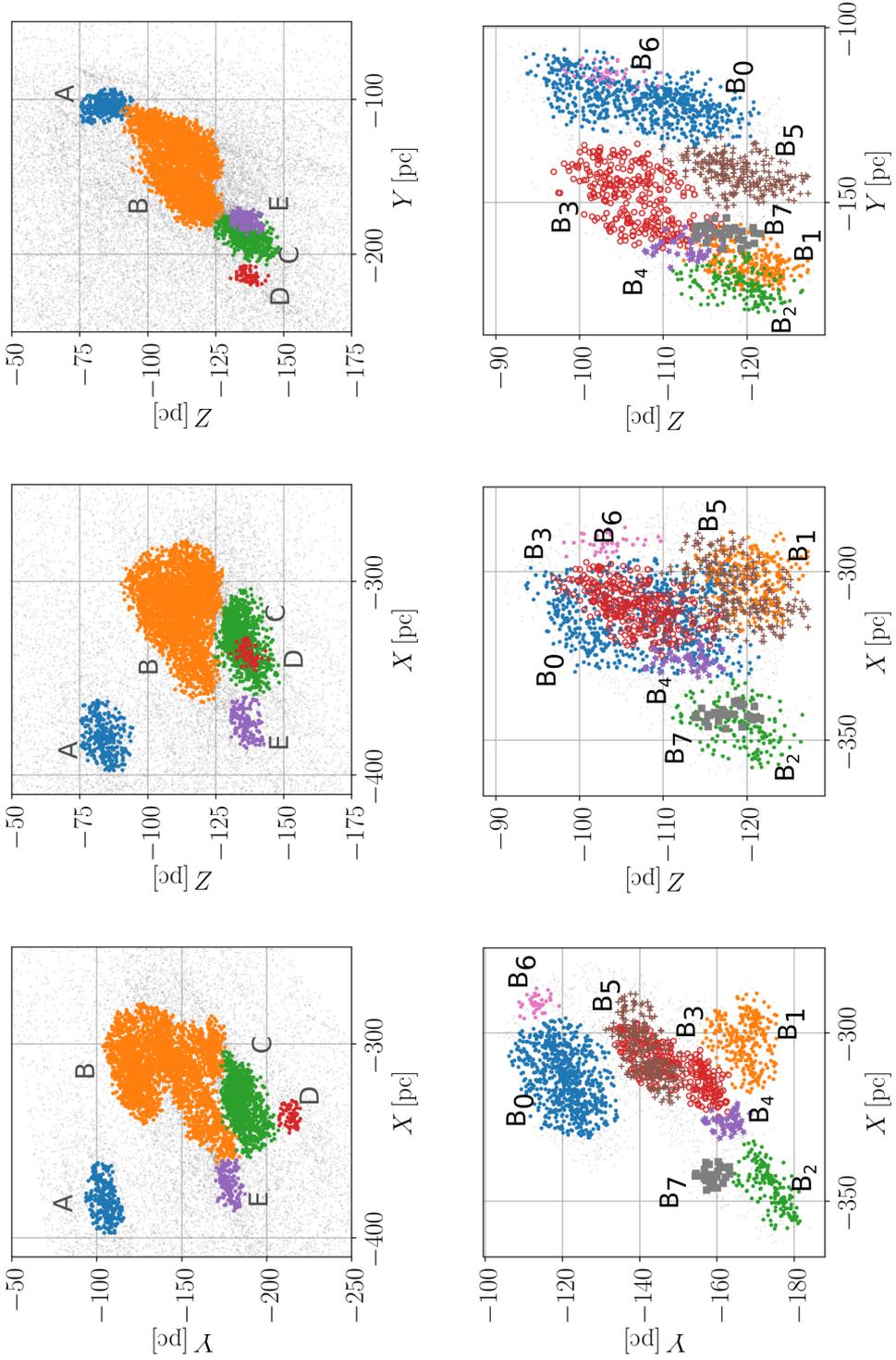
Figure 3.3: Distribution of the groups identified with the first (top) and second (bottom) iteration of DBSCAN in the planes (X, Y), (X, Z), and (Y, Z).

3.4 Kinematics

In this section we study the kinematics of the groups selected in the previous section. We use an iterative maximum likelihood approach to determine a) the average motion of the groups, b) their velocity dispersion, and c) (where possible) the presence of a linear expansion term. We use the method proposed by Lindegren et al. (2000) and applied in Reino et al. (2018) and Bravi et al. (2018), adding however a term to take into account a potential expansion of the cluster from its centre. The method is summarised in Section 3.4.1, tested in Appendix A, and the results are presented in Section 4.2. Here we use ICRS co-ordinates, which we differentiate from galactic co-ordinates by adding the subscript $'I'$ when needed.

3.4.1 Method

Our method extends the maximum-likelihood method developed by Lindegren et al. (2000, L00) by adding measured radial velocities (see Reino et al. 2018) and by including a linear expansion term in the cluster velocity model. Following L00, we assume that the members of a cluster share the same three-dimensional space motion with a small isotropic dispersion term. Reino et al. (2018) extended L00's method by:

- adding measured radial velocity, whenever available, as a fourth observable, besides trigonometric parallax and proper motion;

- making a transition from the χ^2 statistic used in L00, and denoted g, to a p value or $1 - \text{CDF(g, DOF)}$ as a goodness-of-fit statistic;

- using a mixed three- and four-dimensional likelihood function so that both stars with and without known radial velocity can be treated simultaneously.

Following L00, we include a linear expansion term in the cluster velocity model by writing the expected space velocity of a single star at position $\boldsymbol{b}_i$ as:

$$\boldsymbol{u}_i = \boldsymbol{v}_0 + \boldsymbol{T}(\boldsymbol{b}_i - \boldsymbol{b}_0), \tag{3.8}$$

where $\boldsymbol{b}_0$ is an arbitrary reference position, namely the point where the local velocity $\boldsymbol{u}(\boldsymbol{b})$ assumes the status of 'centroid' velocity $\boldsymbol{v}_0$. The co-ordinates of $\boldsymbol{b}_0$ are therefore fixed in advance. The matrix $\boldsymbol{T}$ is simply a diagonal matrix of the form:

$$\boldsymbol{T} = \begin{bmatrix} \kappa & 0 & 0 \\ 0 & \kappa & 0 \\ 0 & 0 & \kappa \end{bmatrix}$$

An expanding cluster will have $\kappa > 0$, from which an expansion age, $\tau = 1/(\gamma\kappa)\,[\text{Myr}]$ can be derived (γ is a conversion factor of $1.0227\,\text{pc}\,\text{Myr}^{-1}\,\text{km}^{-1}\,\text{s}$, see for example Wright & Mamajek 2018).

The method is applied to the members of the clusters identified in Section 3. These clusters still contain 'outliers', that is real non-members, or members which have (slightly) discrepant astrometry (and/or radial velocities) as a result of unrecognised multiplicity, them escaping from the cluster, etc. Such outliers can be found, after maximising the likelihood function, by computing the p value (associated with a particular g
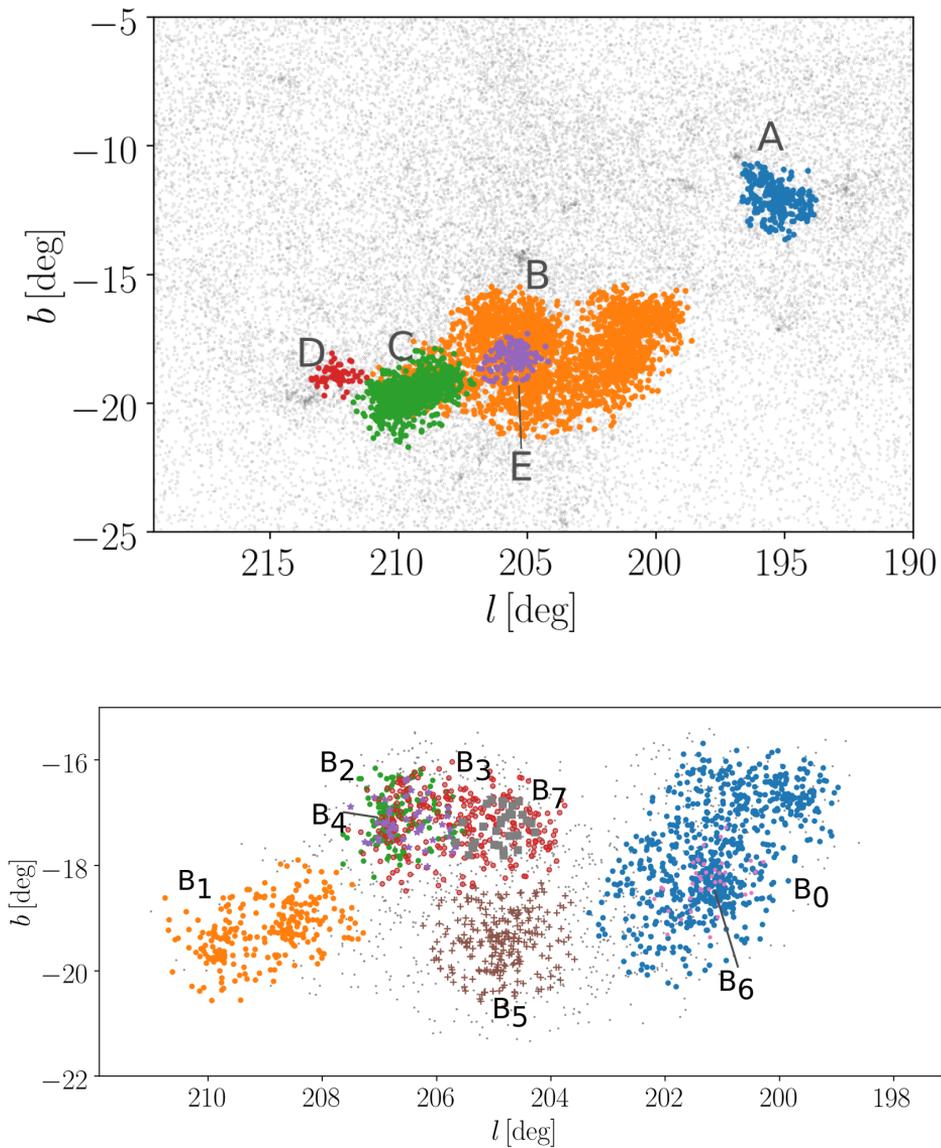
Figure 3.4: Sky distribution of the groups identified with the first (top) and second (bottom) iteration of DBSCAN. The colours correspond to those in Fig. 4.9.

value) for each star in the solution (Eq. 19 in L00). The largest outlier is removed from the sample and a new maximum likelihood solution is determined, until all g values are acceptably small ($g_i \leqslant g_{lim}$ or $p_i \geqslant p_{lim}$). The stopping criterion is the same as in Reino et al. (2018), and is associated to a significance level $p_{lim} = 0.0027$. As noted in Reino et al. (2018), if one stops too early, real outliers will be left and the best-fit velocity dispersion will remain too high. On the contrary, one can keep on iterating and removing outliers until just two stars with very similar three-dimensional motions are left, severely underestimating the velocity dispersion. Astrometric data only can not distinguish between expansion or contraction of a cluster from a change in $\boldsymbol{v}_0$ (see L00). Therefore when the fraction of measured radial velocities is lower than the 20% we do not estimate the expansion coefficient κ (implicitly assuming $\kappa = 0$). The threshold is conservative for certain groups, but the derived parameters are robust for all the groups.

3.4.2   Results

The results of the kinematic modelling code are give in Table 3.1. Being quite isolated with respect to the rest of the population, the λ Ori group (group A) is easy to identify and separate from the others, therefore the results do not require any specific clarification. This is not the case for the groups with $199° < l < 216°$. We comment on the results for these groups by dividing them in three 'regions' according to their sky distribution: the 25 Ori region, the Belt region, and the Orion A region.

25 Ori

We define the 25 Ori region as:

$$199° < l < 203° \qquad -20° < b < -15°, \tag{3.9}$$

which corresponds to the groups B_0 and B_6 identified by DBSCAN. The proper motions of the sources in the region (black dots in Fig. 3.5, left) separate in two clumps. This was shown also by Kos et al. (2018), who however apply a different classification scheme to separate the clusters in the region. The separation is also visible when considering the proper motion diagram of group B_0 (blue dots in Fig. 3.5, left). The number of sources is lower because the DBSCAN algorithm favours the high density groups (so when the density drops under a certain level the stars are considered as 'noise stars' and not classified as members of any cluster).

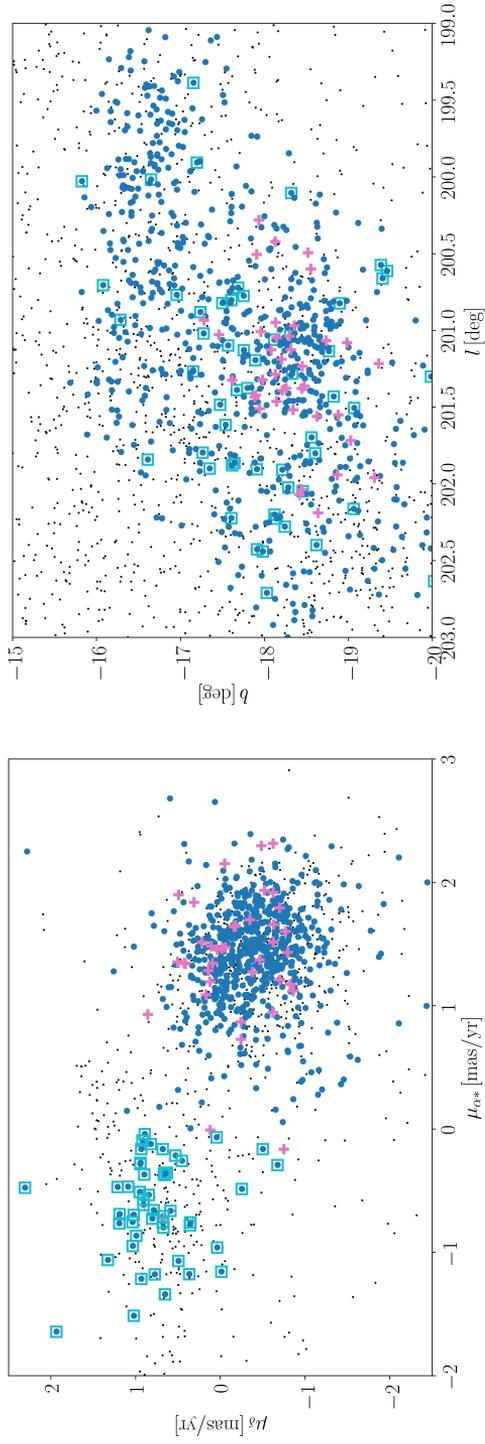**Figure 3.5:** Left: proper motion diagram of the stars in the 25 Ori region defined in the text (black dots), of the stars belonging to group B_0 (blue dots), and of the stars belonging to group $B_{0,b}$ (light blue empty squares) and B_6 (pink crosses). Right: sky distribution in galactic co-ordinates of group B_0, $B_{0,b}$, and B_6. The colours and symbols are the same as on the left.

We considered the sources selected by DBSCAN, and we isolated the second group ($B_{0,b}$, light blue squares in Fig. 3.5, left) by applying the following cuts in proper motion space:

$$\mu_{\alpha*} < 0 \,\mathrm{mas\,yr^{-1}} \qquad \mu_\delta > -1 \,\mathrm{mas\,yr^{-1}}. \tag{3.10}$$

We applied separately the kinematic modelling code to the two groups. The results are reported in Table 3.1. We also run the kinematic modelling code considering all the sources in the region, after separating the two groups using the same criteria of Eq. 3.10. The estimated parameters are consistent. The sky distribution of the sources of group B_0 and $B_{0,b}$ is shown in Fig. 3.5 (right panel). While group B_0's distribution shows a clump towards 25 Ori, and the northern elongation reported for instance by Lombardi et al. (2017) and Briceño et al. (2019), group $B_{0,b}$'s sources are scattered in the field and do not show any clear concentration. Together with the findings by Kos et al. (2018) in terms of ages (see also Section 5), this points to the conclusion that group $B_{0,b}$ is slowly dispersing in the galactic field. Here we are limiting our samples to the 25 Ori region, but in principle members of the $B_{0,b}$ group could be found spread over a larger area of the sky (and 3D space).

Group B_6 consists only of 30 members, none of which has a measured radial velocity, therefore we decided not apply the kinematic modelling code. The parallax distribution suggests that B_6 is closer to the Sun than group B_0, while the proper motion distribution does not show any difference with respect to group B_0. We suspect that group B_6 coincides with a small over-density of sources within group B_0, which gets classified as a separate group because of a local density drop. We ran the kinematic modelling code for groups B_0 and B_6 together: the estimated parameters are consistent with those found for group B_0 only, which supports our hypothesis.

Belt

Many of the clusters identified by DBSCAN (B_2, B_3, B_4, B_5, B_7 and E) are located in the Sky towards the Belt region. Fig. 3.6 shows the proper motion diagram for the Belt region defined as

$$203° < l < 207.5° \qquad -21° < b < -13°. \tag{3.11}$$

Proper motions in the Belt region present a high degree of sub-structure, indicating that the Belt hosts groups with different kinematic properties.

- Groups B_2 and B_4 are mostly located towards the σ Ori cluster (see Fig. 3.7) and ζ Ori. Group B_3's members are spread towards ϵ Ori and δ Ori. The parameters estimated by the kinematic modelling code suggest that B_2 and B_4 have compatible $v_{y,I}$ values, which are significantly different from those of group B_3. This is consistent with what is found by Jeffries et al. (2006), who already notice the presence of two kinematics components towards the cluster. The kinematic properties of group B_3 are similar to those of groups D, B_0 (not located in the Belt region, see Fig. 3.4), and B_5. We notice that group B_2's velocity dispersion is large ($\sim 1.6\,\mathrm{km\,s^{-1}}$) compared for instance to that of group B_3 ($0.41 \pm 0.02\,\mathrm{km\,s^{-1}}$). The proper motion distribution shows indeed some substructures, which cause the large value of the velocity dispersion. As mentioned above, the presence of kinematic substructure may indicate the co-existence of

groups with different kinematics in the same area. An inspection of group B_2's 3D configuration (see Fig. 4.9, in particular the $X-Y$ projection) shows that the source distribution is not uniform, and seems to be divided into (at least two) elongated structures.

- Group B_5 is located below the Belt, towards η Ori, and shares similar kinematics with group B_3, although they seem to be well separated in space (see Fig. 4.9 and 3.4). The proper motion distribution shows two clumps, similar to what is observed towards 25 Ori. We separate the the smaller clump, which we refer to as $B_{5,b}$ by using simple cuts in proper motion space:

$$0.3 \, \mathrm{mas \, yr}^{-1} < \mu_{\alpha*} < 2. \, \mathrm{mas \, yr}^{-1};$$
$$-0.8 \, \mathrm{mas \, yr}^{-1} < \mu_{\delta} < 0.3 \, \mathrm{mas \, yr}^{-1}. \tag{3.12}$$

In contrast to what we have done for group $B_{0,b}$, here we apply the conditions of Eq. 3.12 to all the sources in the Belt region, and not just those within the $f(D) = 0.5$ level of the 3D density map. This is the reason why the number of sources is higher than for group E (see Table 3.1). This choice is motivated by the fact that the visual inspection of the proper motion diagram suggests that the clump is more extended and the number of sources is larger than what found by DBSCAN. Further, the number of sources of the smaller clump is too small to retrieve the kinematic parameters accurately. The parameters estimated by the kinematic modelling code (see Table 3.1) show that group B_5 and group $B_{5,b}$ have different kinematic properties, while having similar parallaxes. Comparing Fig. 3.6 and Fig. 3.10 one can notice that the region defined in Eq. 3.12 also includes sources classified as members of group B_2. The sky distribution of sources belonging to group $B_{5,b}$ (see Fig. 3.8) shows indeed some sources clustering around σ Ori. Most of the sources however are located in the same region as group B_5, although they are spread throughout the entire longitudinal extent of the Belt region. This seems to suggest that group $B_{5,b}$ is more extended than the Belt region, especially to lower galactic latitudes and longitudes. Similar conclusions can be drawn after studying the 3D distribution of group $B_{5,b}$ (Fig. 3.8): some sources clump in the same area as group B_2 and B_4 (σ Ori), while others are located closer to group B_5. This explains why DBSCAN does not separate successfully groups B_5 and $B_{5,b}$: their members show different kinematics but are mixed in space.
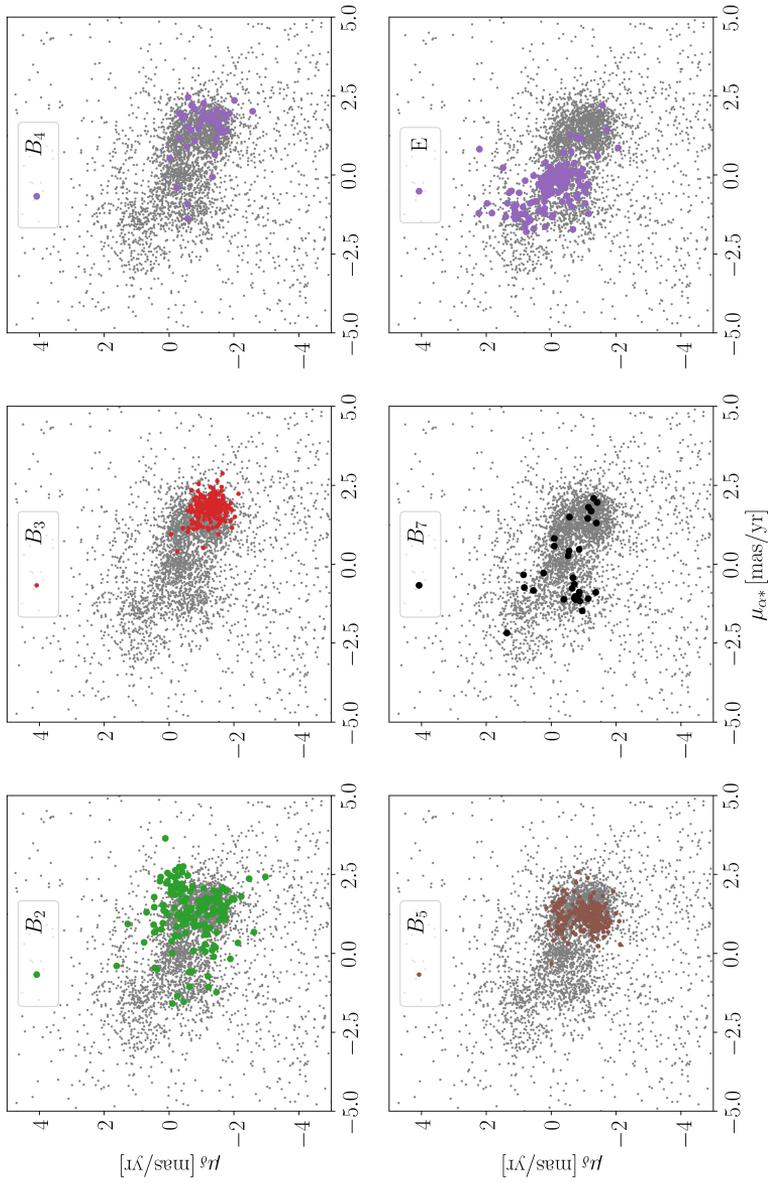
Figure 3.6: Proper motion diagram for all the stars in the Belt region (grey dots). Each panel corresponds to one of the groups identified by DBSCAN (the colours are the same as in Figs. 4.9 and 3.4) except for group B_7, which is indicated by black dots for representation purposes.
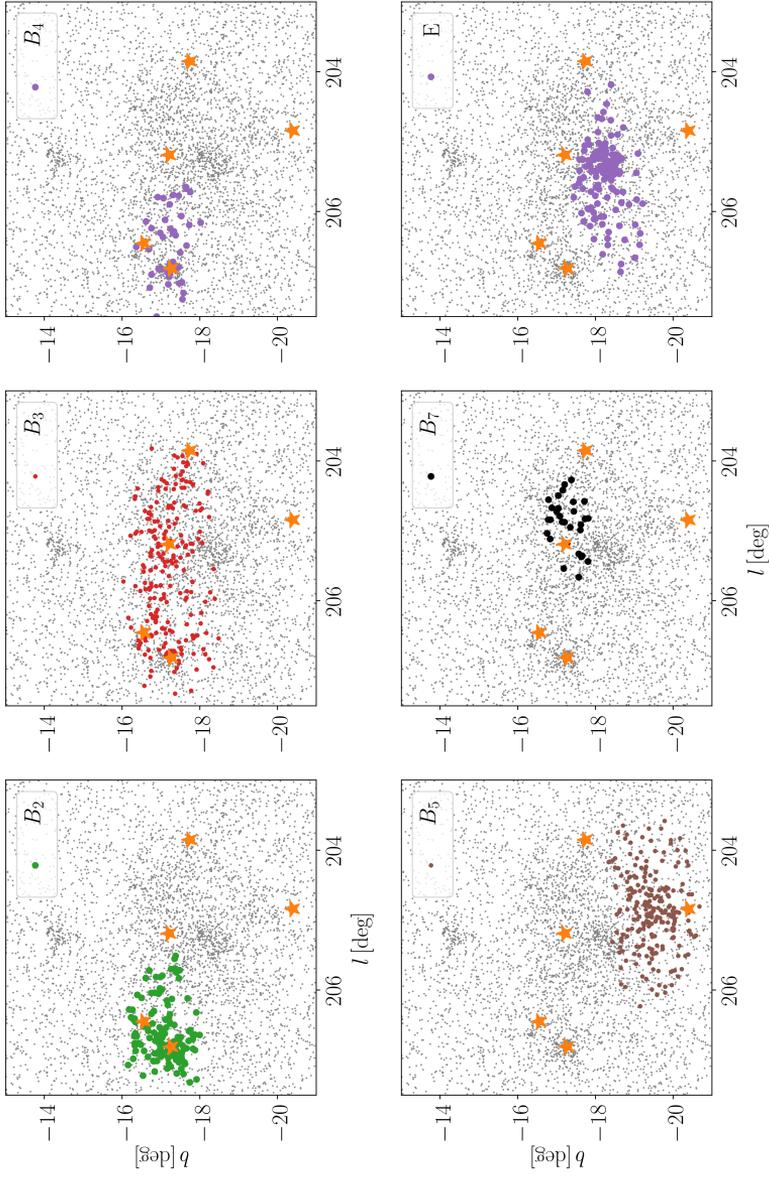
Figure 3.7: Distribution in the sky of the stars in the Belt region. Each panel corresponds to one group identified by DBSCAN. The colour-coding is the same as in Fig. 3.6. The orange stars correspond to, from left to right: σ Ori, ζ Ori, ε Ori, η Ori, δ Ori.

- Group E is the most distant group in the entire Orion region (see Table 3.1 and Fig. 4.9). Since not many radial velocity measurements are available, the kinematic properties are determined with less accuracy than for the other groups, especially in the y_I direction. While $v_{y,I}$ is comparable with those of group A, C, B_1, B_2, B_4 (and $B_{7,b}$ and B_8, see below), the $v_{x,I}$ component is different from the other groups. As for group B_5, the proper motions seem to be divided in two clumps, one of which does not correspond to any other DBSCAN groups. We select group B_8 by applying the following conditions:

$$-2.2\,\text{mas yr}^{-1} < \mu_{\alpha*} < -0.5\,\text{mas yr}^{-1};$$
$$0.4\,\text{mas yr}^{-1} < \mu_\delta < 2.2\,\text{mas yr}^{-1}. \tag{3.13}$$

Similarly as for group $B_{5,b}$, and with the same motivations, we consider again all the sources in the Belt region. The estimated kinematic parameters are reported in Table 3.1. The source distribution in the sky and in 3D Cartesian space is shown in Fig. 4.5, compared to that of group E. The sources are loosely distributed in the entire Belt region, although they seem to clump next to group E.

- DBSCAN identifies only 30 sources belonging to group B_7, none of them with a measured radial velocity, therefore the kinematic modelling code does not succeed in determining reliable parameters. Similarly to what was found for group $B_{5,b}$ and B_8, when considering all the stars in the Belt area, we notice that many more sources clump in the same proper motion region that are excluded when we apply the condition $f(D) > 0.5$ or that are classified as 'noise' stars by DBSCAN. We therefore select group B_{7b} according to the following equations (see Fig.):

$$-2.2\,\text{mas yr}^{-1} < \mu_{\alpha*} < -0.5\,\text{mas yr}^{-1};$$
$$-2\,\text{mas yr}^{-1} < \mu_\delta < 0.4\,\text{mas yr}^{-1}. \tag{3.14}$$

The number of sources is now much larger (see Table 3.1), and the parameters can be accurately determined. Fig. 3.10 shows the source distribution in the sky and in Cartesian galactic co-ordinates. We notice that the sources are distributed in the sky towards the reflection nebulae M78 and NGC 2071, where two groups of young stars are present and towards the centre of the Belt.

- Figure 3.9 shows the dust distribution towards the Belt region, where a bubble is visible (see Ochsendorf et al. 2014, 2015). Some of the groups we identified might be responsible for the origin of the Belt bubble. In particular groups E and B_8 are located in the sky within the dust structure shown in Fig. 3.9, at different distances. Group B_8 is slightly more diffuse than the bubble, but the central over-density is still located within the bubble boundaries. The stellar winds and the supernova explosions coming from these groups might be responsible for the creation of the bubble itself.

Orion A

The DBSCAN groups associated with the Orion A molecular cloud are those labelled B_1, C, and D. Group B_1 and C nearly occupy the same position in the sky and share
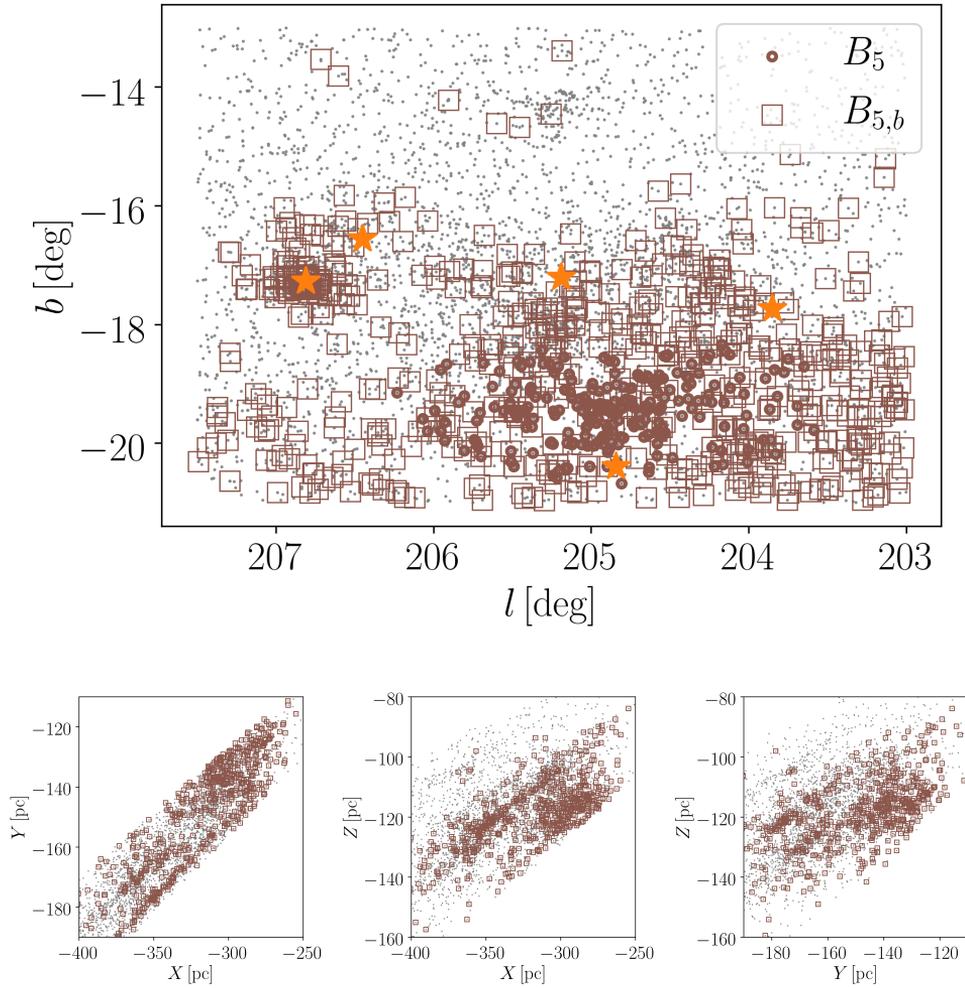
Figure 3.8: The top panel shows the sky distribution of the sources belonging to group $B_{5,b}$ (brown empty squares), group B_5 (brown dots), and all the sources in the Belt region defined in the text (grey dots). The orange stars mark the position of σ Ori, ζ Ori, ϵ Ori, η Ori, and δ Ori. The bottom panels show the distribution in 3D galactic co-ordinates of group $B_{5,b}$ (brown squares), and of all the sources belonging to the Belt region (grey dots).
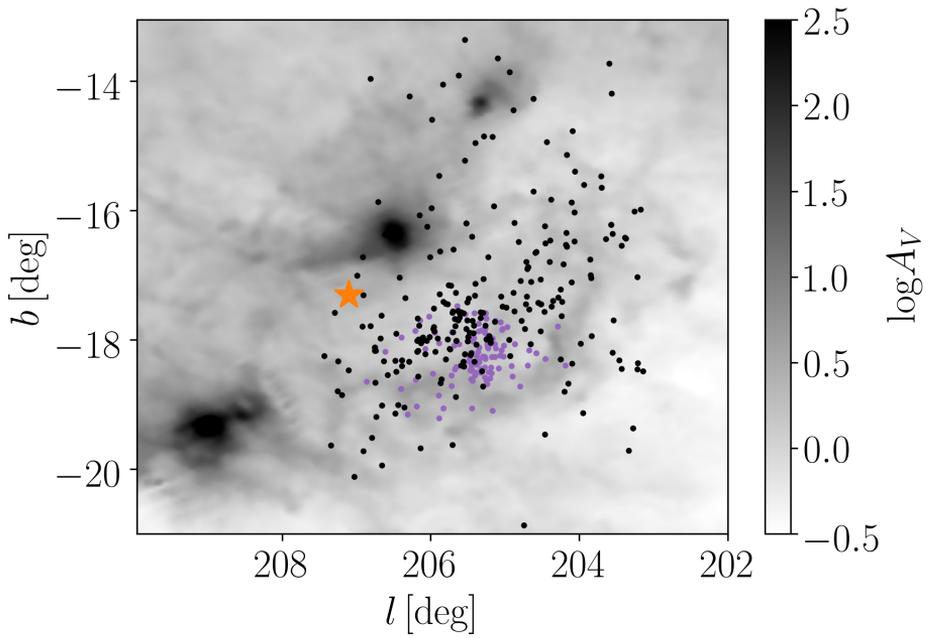
Figure 3.9: *Planck* data and groups E (purple dots) and B_8 (black dots). The orange star represents σ Ori.
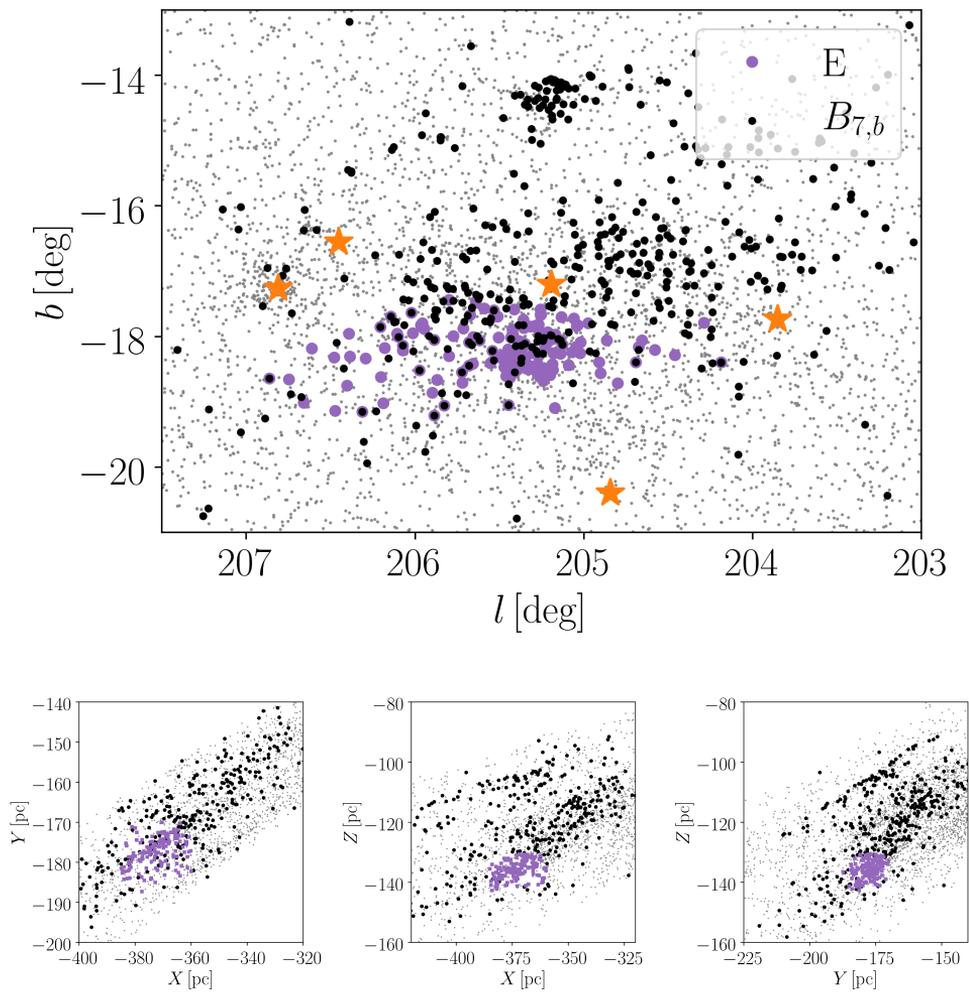
Figure 3.10: Distribution in the sky (top) and in 3D space (bottom) of the stars belonging to group $B_{7,b}$ (black dots) compared to those in group E (purple dots) The grey dots represent all the sources in the Belt region. The orange stars are the same as defined in Fig. 3.7.
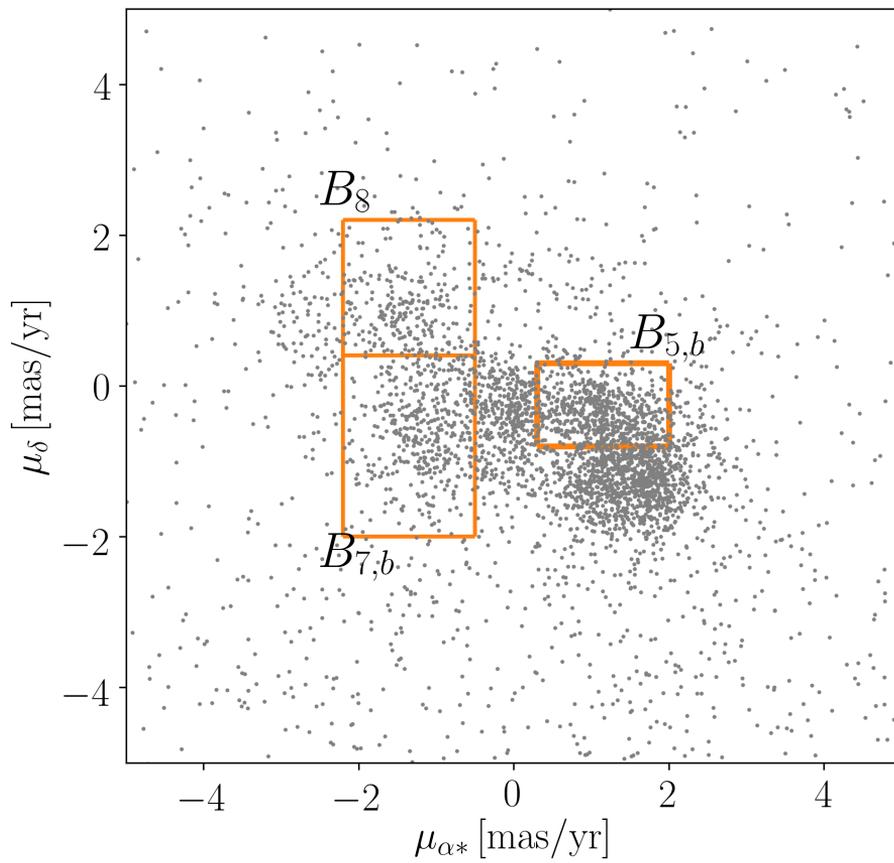
Figure 3.11: Proper motion diagram of all the sources in the Belt region. The orange rectangles are those defined in Eq. 3.12 and 3.13.
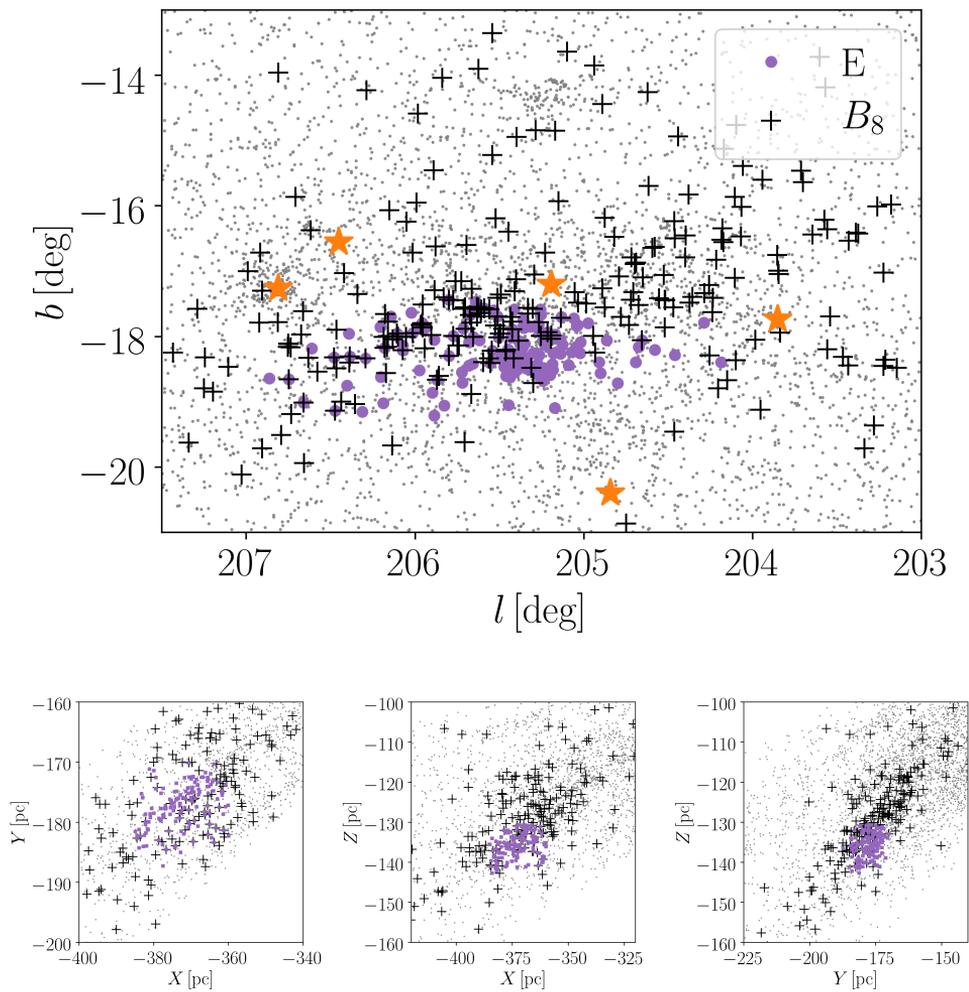
Figure 3.12: Distribution in the sky (top) and in 3D space (bottom) of the stars belonging to group B_8 (black crosses) compared to those in group E (purple dots). The orange stars are the same as defined in Fig. 3.7.

very similar kinematic properties (see Table 3.1), however they are at different distances, with group B_1 being closer to the Sun than group C. This poses interesting questions about their origin: the two groups might be identified separately by DB-SCAN just because of a local under-density of sources. In this case, the Orion A cloud would be even more elongated along the line of sight than previously thought (Großschedl et al. 2018). The radial velocities of the embedded sources in the Orion A molecular cloud are tightly related to the motion of the molecular gas in the cloud (Hacar et al. 2016). So, if the foreground is moving as the stars in the cloud, and stars in the cloud are coupled to the gas, the foreground group might have originated from the same cloud complex. The proper motion diagram of the three groups is shown in Fig. 3.13. We define the Orion A region as:

$$207.5° < l < 216°, \qquad -22° < b < -17°. \tag{3.15}$$

The proper motions of all the sources (grey dots in Fig. 3.13) in the region show a clump in $\mu_{\alpha*}, \mu_\delta \sim (-2., 1)$ (see also left panel of Fig. 3.14). We select the sources with proper motions:

$$-2.5 \, \mathrm{mas\,yr^{-1}} < \mu_{\alpha*} < -1. \, \mathrm{mas\,yr^{-1}};$$
$$0 \, \mathrm{mas\,yr^{-1}} < \mu_\delta < 2. \, \mathrm{mas\,yr^{-1}}, \tag{3.16}$$

(black dots in Fig. 3.14) and we study their distribution in the sky and on the $X - Y$ plane in galactic Cartesian co-ordinates. We label this group as group F. Fig. 3.14 (centre) shows that the sources are loosely distributed in the Orion A region, and seem to cluster at $(l, b) \sim (209, -19)$. Fig. 3.14 (right) show that the members of group F are loosely spread at larger distances than the sources associated with the Orion A molecular cloud. We run the kinematic modelling on group F and we find the parameters reported in Table 3.1. We compare the proper motions of group F with those of the other groups, and we notice that they are roughly the same as those of group B_8 (see Fig. 3.10). Nevertheless the results of the kinematic modelling for the two groups are quite dissimilar. This could be due to the fact that, for both groups, the number of stars with measured radial velocity is small, and therefore the 3D velocity is not well constrained. An inspection of the parallax distribution of group F also shows a number of sources with small parallax ($\varpi < 1.9 \, \mathrm{mas}$), which are most likely field contaminants.
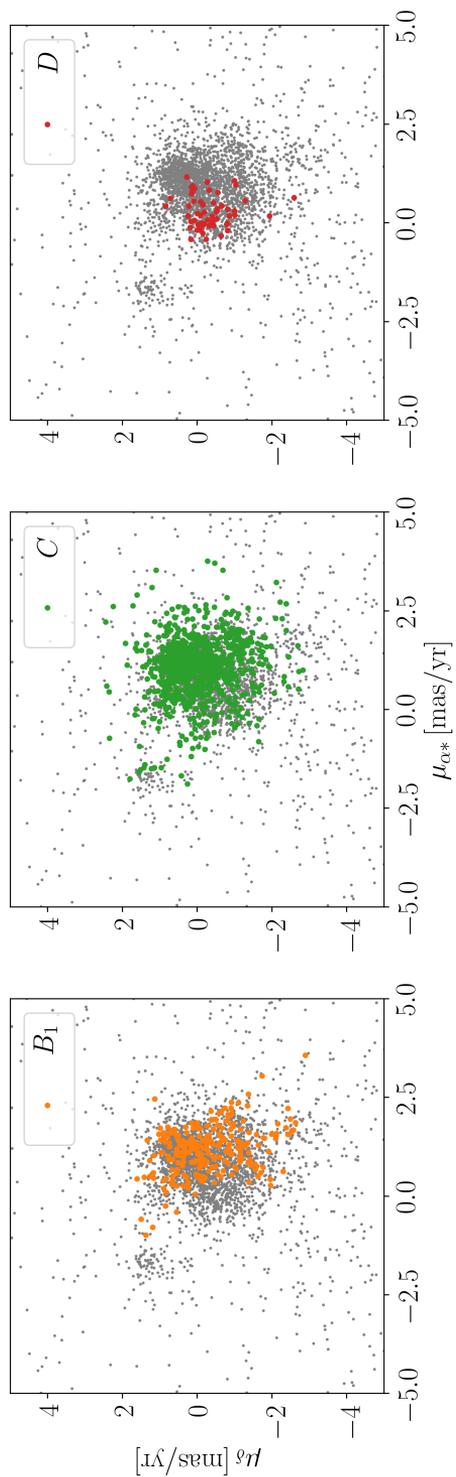
Figure 3.13: Proper motion diagram of the sources towards the Orion A molecular cloud, for group B_1 (orange dots in the left panel), group C (green dots in central panel), and group D (red dots in right panel). The grey dots represent the proper motions of all the sources in the Orion A region (see text).
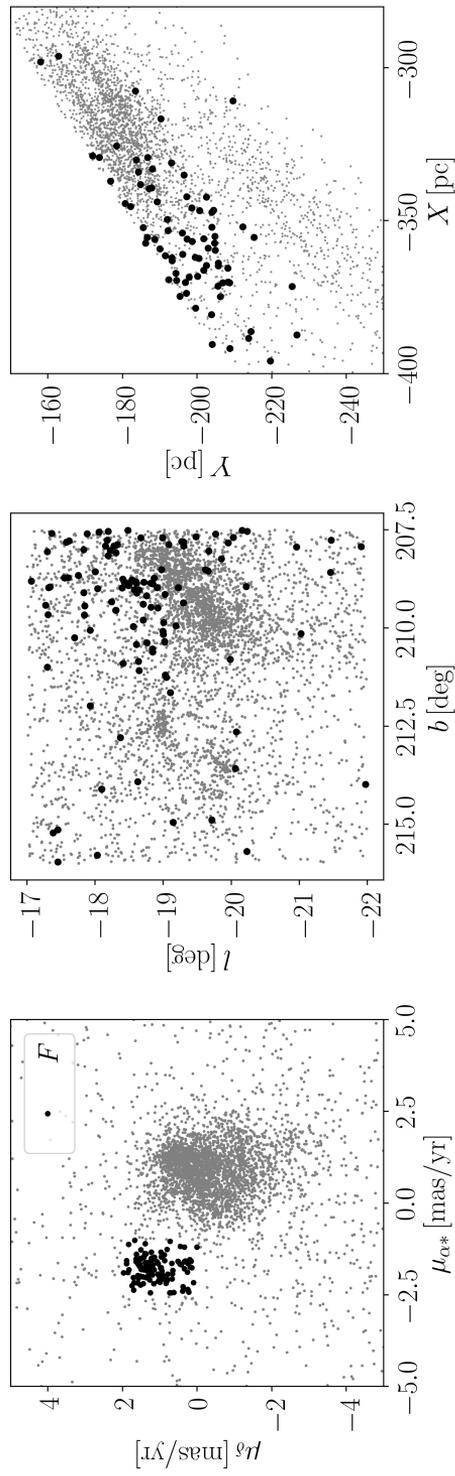
Figure 3.14: Proper motion (left), sky distribution (centre), and distribution in the $X - Y$ plane in Cartesian galactic co-ordinates of all the sources in the Orion A region (grey dots) and of those selected through Eq. 3.16 (black dots).

Table 3.1: Results of the kinematic modelling for the groups identified by DBSCAN. The first column indicates the region in the sky where the groups are located.

	#	N	N_{RV}	$v_{x,I}\,[\mathrm{km\,s^{-1}}]$	$v_{y,I}\,[\mathrm{km\,s^{-1}}]$	$v_{z,I}\,[\mathrm{km\,s^{-1}}]$	$\sigma_v\,[\mathrm{km\,s^{-1}}]$	$\kappa\,[\mathrm{km\,s^{-1}\,pc^{-1}}]$	$\varpi\,[\mathrm{mas}]$
λOri	A	296	81	0.75 ± 0.05	27.4 ± 0.1	0.5 ± 0.05	0.73 ± 0.02	0.122 ± 0.007	$2.48^{0.07}_{0.06}$
OrionA	C	1059	489	0.94 ± 0.06	27.1 ± 0.09	-2.68 ± 0.06	1.63 ± 0.03	0.07 ± 0.006	$2.5^{0.06}_{0.08}$
OrionA	D	69	50	1.1 ± 0.1	21.8 ± 0.2	-3.7 ± 0.1	0.98 ± 0.07	-0.02 ± 0.02	$2.37^{0.02}_{0.03}$
Belt	E	150	10	4.6 ± 0.3	32.4 ± 2.7	-1 ± 0.1	1.28 ± 0.06	-	$2.31^{0.03}_{0.04}$
25Ori	B_0	710	73	0.41 ± 0.04	19.2 ± 0.2	0.045 ± 0.03	0.74 ± 0.02	-	$2.84^{0.08}_{0.08}$
25Ori	B_{0b}	54	4	4.8 ± 0.4	$25. \pm 2.8$	2.1 ± 0.1	0.38 ± 0.04	-	$2.78^{0.06}_{0.06}$
OrionA	B_1	265	73	0.8 ± 0.1	26.8 ± 0.2	-3.0 ± 0.1	1.55 ± 0.05	0.03 ± 0.02	$2.74^{0.04}_{0.07}$
Belt	B_2	174	48	0.17 ± 0.14	27.84 ± 0.3	-2.6 ± 0.1	1.6 ± 0.07	0.08 ± 0.03	$2.48^{0.04}_{0.05}$
Belt	B_3	290	48	-0.74 ± 0.03	22.5 ± 0.2	-2.69 ± 0.05	0.79 ± 0.03	-	$2.78^{0.07}_{0.08}$
Belt	B_4	46	10	-0.2 ± 0.3	26.5 ± 0.6	-3.1 ± 0.2	1.6 ± 0.1	0.07 ± 0.06	$2.62^{0.02}_{0.02}$
Belt	B_5	248	12	0.9 ± 0.1	20.7 ± 0.9	-2.44 ± 0.06	0.7 ± 0.03	-	$2.79^{0.09}_{0.06}$
Belt	$B_{5,b}$	622	48	1.2 ± 0.1	24.8 ± 0.2	-1.42 ± 0.04	0.82 ± 0.02	-	$2.75^{0.22}_{0.36}$
25Ori	B_6	40	0	-	-	-	-	-	-
Belt	B_7	30	0	-	-	-	-	-	-
Belt	$B_{7,b}$	441	63	4.9 ± 0.05	26.7 ± 0.2	-1.55 ± 0.05	0.95 ± 0.03	0.024 ± 0.003	$2.36^{0.2}_{0.23}$
Belt	B_8	245	18	5.8 ± 0.08	28.5 ± 0.4	1.5 ± 0.06	0.9 ± 0.03	0.05 ± 0.05	$2.34^{0.2}_{0.18}$
OrionA	F	116	17	5.8 ± 0.1	21.2 ± 0.5	0.3 ± 0.12	$1. \pm 0.06$	-	$2.27^{0.2}_{0.48}$

3.5 Ages

We determine ages (τ) and extinctions (A_V) of the groups we identified by performing an isochrone fit based on a maximum likelihood approach similar to the methods described in Jørgensen & Lindegren (2005), Valls-Gabaud (2014), and Zari et al. (2017).

Assuming independent Gaussian errors on all the observed quantities we can write the likelihood for a single star to come from an isochrone with certain properties $\boldsymbol{\theta} = (\tau, A_V, Z, ...)$, as:

$$L(\boldsymbol{\theta}, m) = \prod_{i=1}^{n} \left(\frac{1}{(2\pi)^{1/2}\sigma_i} \right) \times \exp\left(-\chi^2/2\right), \tag{3.17}$$

with:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{q_i^{\mathrm{obs}} - q_i(\boldsymbol{\theta}, m)}{\sigma_i} \right)^2, \tag{3.18}$$

where m is the stellar mass, n is the number of observed quantities, and $\boldsymbol{q}^{\mathrm{obs}}$ and $\boldsymbol{q}(\boldsymbol{\theta}, m)$ are the vectors of observed and modelled quantities. To take into account the fact that stars are not distributed uniformly along the isochrone, we weight the jth likelihood with a factor w defined as:

$$w = \sqrt{\frac{n_{redder\,j}}{n_{bluer\,j} + 1}}, \tag{3.19}$$

where n_{redder} is the number of stars with $G_{BP} - G_{\mathrm{RP}}$ colour larger than that of the jth star and n_{bluer} is the number of stars with $G_{BP} - G_{\mathrm{RP}}$ smaller than that of the jth star. This choice gives larger weights to blue, massive stars, to take into account that they are fewer than the low-mass members of the clusters.

The likelihood for N coeval stars is just defined as:

$$L_{combined}(\boldsymbol{\theta}, m) = \prod_{j=1}^{N} L_j(\boldsymbol{\theta}, m)^{w_j} \tag{3.20}$$

Since we are interested in determining the ages and the extinctions of the groups, we fix the metallicity to $Z = Z_\odot = 0.0158$ and we integrate Eq. 13 on the mass, so that the probability density function as a function of age τ and extinction A_V is given by:

$$L_{combined}(\tau, A_V) = \prod_{j=1}^{N} \int L_j(\tau, A_V, m)dm \tag{3.21}$$

To perform the fit we compare the observed G magnitude and $G_{BP} - G_{\mathrm{RP}}$ colour to those predicted by the PARSEC (PAdova and TRieste Stellar Evolution Code Bressan et al. 2012; Chen et al. 2014; Tang et al. 2014) library of stellar evolutionary tracks, using the passbands by Maíz Apellániz & Weiler (2018). We used isochronal tracks from $\log(\mathrm{age/yr}) = 6.0\,(1\,\mathrm{Myr})$ to $\log(\mathrm{age/yr}) = 8.0\,(100\,\mathrm{Myr})$, with a step of $\log(\mathrm{age/yr}) = 0.05$, and from $A_V = 0\,\mathrm{mag}$ to $A_V = 2.5\,\mathrm{mag}$ with a step of $0.1\,\mathrm{mag}$.

Table 3.2: Age estimates for the groups identified in Section 3 and 4. The column $\log(\text{age/yr})$ (τ) indicates the (log-)age estimated by the isochrone fitting procedure. The column τ_{exp} indicates the expansion ages determined by using the formula $\tau_{exp} = 1/(\gamma \kappa)$ for the groups for which it is possible to determine the expansion parameter κ. The number of stars N is different than in Table 3.3 because by applying the kinematic modelling we remove kinematic outliers from the groups.

#	N	$\log(\text{age/yr})$	τ [Myr]	A_V [mag]	τ_{exp} [Myr]
A	274	$6.75^{0.03}_{0.01}$	$5.6^{0.4}_{0.1}$	0.4	8.0
C	943	$6.9^{0.03}_{0.01}$	$8^{0.5}_{0.04}$	0.2	14.0
D	60	$6.85^{0.03}_{0.02}$	$7^{0.6}_{0.2}$	1.3	-
E	139	$7.05^{0.04}_{0.005}$	$11.2^{1}_{0.1}$	0.5	-
B_0	622	$7.05^{0.04}_{0.005}$	$11.2^{1}_{0.1}$	0.2	-
B_{0b}	44	$7.15^{0.1}_{0.004}$	$14^{3}_{0.25}$	0.4	-
B_1	246	$7.0^{0.03}_{0.01}$	$10^{0.7}_{0.23}$	0.4	32.6
B_2	154	$6.6^{0.03}_{0.01}$	$4^{0.3}_{0.1}$	0.3	12.2
B_3	221	$6.9^{0.04}_{0.01}$	$8^{0.7}_{0.04}$	0.2	-
B_4	44	$6.6^{0.03}_{0.01}$	$4^{0.3}_{0.1}$	0	14
B_5	234	$6.9^{0.04}_{0.01}$	$8^{0.7}_{0.04}$	0.2	-
$B_{5,b}$	605	$7.05^{0.03}_{0.005}$	$11.2^{1}_{0.1}$	0.2	-
$B_{7,b}$	418	$7.05^{0.04}_{0.005}$	$11.2^{1}_{0.1}$	0.3	40
B_8	237	$7.15^{0.04}_{0.004}$	$14^{1.5}_{0.25}$	0.3	-
F	108	$7.05^{0.03}_{0.005}$	$11.2^{1}_{0.1}$	0.3	-

Figure 3.15: $\log L$ for the cluster B_0. We note the correlation between age and extinction.

Our fitting procedure does not take into account the presence of unresolved binaries, the photometric variability of young stars, the presence of circumstellar material, or potential age spreads within single groups. These effects can bias our age estimates and this issue is further discussed in Section 6.2

3.5.1   Results

We compute the age τ and the A_V for the groups identified by DBSCAN, and for the groups we selected in Section 4. The results are reported in Table 3.2. Figures 4.12 and 3.16 show the log-likelihood

$$\log L = \log L_{combined}(\tau, A_V)$$

we obtain for group B_0, and the M_G vs. $G-G_{\rm RP}$ (left) and M_G vs. $G_{BP}-G_{\rm RP}$ (right) colour-magnitude diagrams (the colour-magnitude diagrams for the other groups are shown in Appendix B). The orange solid line corresponds to the best-fitting isochrone. As mentioned above, we perform the fit using the $G_{BP}-G_{\rm RP}$ colour, and we show the colour-magnitude diagram in $G-G_{\rm RP}$ as a quality check. We adopt the maximum of $L_{combined}(\tau, A_V)$ as our best estimate of the stellar age, and we compute the confidence intervals by evaluating the 16th and the 84th percentiles after marginalizing over A_V. Figure 4.12 shows a correlation between age and extinction: at large extinction values the isochrones move towards redder colours, and soon they do not intersect the upper main sequence. However they still can fit the low pre-main sequence.
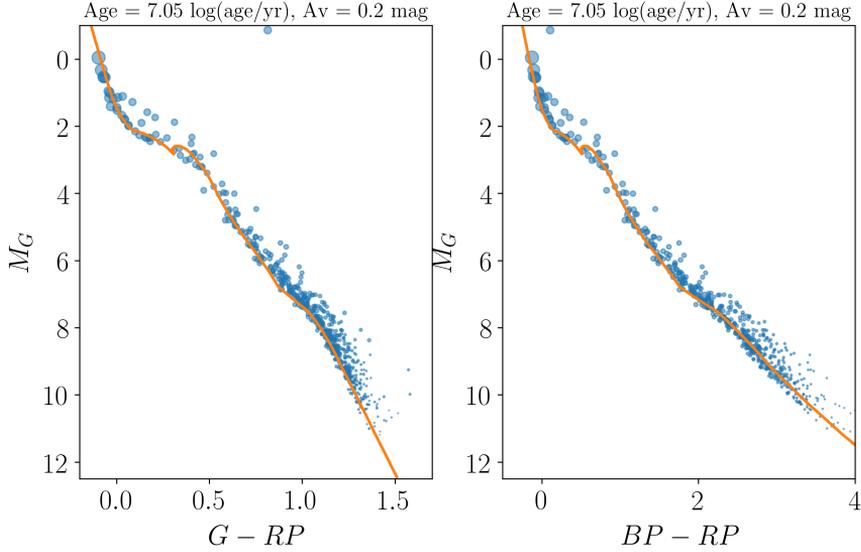
Figure 3.16: M_G vs. $G - G_{\rm RP}$ (left) and M_G vs. $G_{BP} - G_{\rm RP}$ (right) the colour magnitudes for group B_0. The symbol sizes represent the weights assigned to each star. The solid orange line represents the best fit isochrone.

3.6 Dicussion

In this section we summarise and comment the results obtained in the previous Sections and we put them in the broader context of the models of sequential star formation and triggering.

3.6.1 Kinematics

By considering the $v_{y,I}$ velocities, we notice that we can roughly divide them in two groups, the first one with $v_{y,I} \sim 20\,{\rm km\,s^{-1}}$ and the second one with $v_{y,I} \sim 26 - 27\,{\rm km\,s^{-1}}$. We observe a loose correlation between velocity and distances (the farthest objects are also the fastest), while there is no correlation between velocity and age or distance and age.

In the kinematic modelling code we included isotropic expansion, however expansion could be an-isotropic, as observed for example by Cantat-Gaudin et al. (2018b) and Wright & Mamajek (2018), although expansion due to residual gas expulsion is usually thought to be isotropic. The expansion ages determined by using the formula $\tau_{\rm exp} = 1/(\gamma\kappa)$ give a loose indication of the group ages, and confirm the age ordering obtained by the isochrone fitting procedure. The results of the simulations that we performed to test the kinematic modelling code (see Appendix A) showed that the expansion parameter κ always resulted to be under-estimated, thus providing over-estimated expansion ages. This is consistent with the expansion ages obtained for the DBSCAN groups.
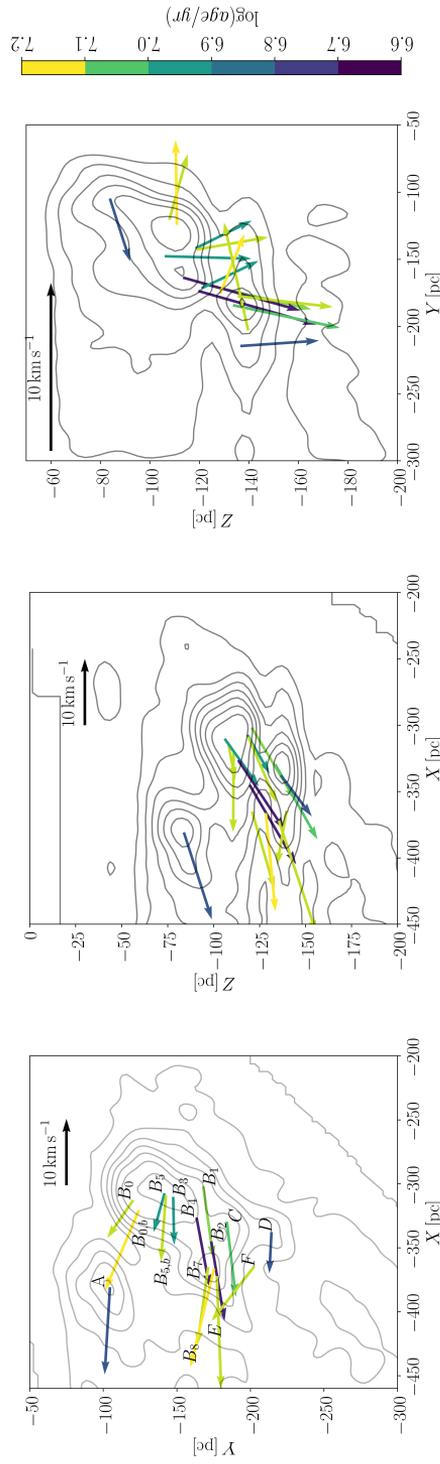
Figure 3.17: The contours represent the density distribution of the selected sources projected in the Galactic plane (left), in the (X, Z) plane (centre), and in the (Y, Z) plane (the Sun is at $(X, Y, Z) = 0, 0, 0$). The arrows represent the velocities (in Galactic co-ordinates) estimated in Section 3 for all the groups, and are corrected for the solar motion. The colours of the arrows represent the $\log(age/yr)$ obtained by fitting the colour-magnitude diagrams of the clusters in Section 4.

As mentioned in Section 3, by using the DBSCAN algorithm we preferentially select clusters that are dense in 3D space, and tend to neglect more diffuse groups. This effect is mitigated by the visual inspection of the proper motion diagrams of the DBSCAN groups, which we use to select groups with common kinematic properties that DBSCAN fails to retrieve. Further, one of the goals of the kinematic modelling code is to exclude outliers from the DBSCAN groups. Outliers are stars that do not share the same kinematic properties as the other cluster members: this implies that also stars that should be considered cluster members, such as binaries, are excluded from the DBSCAN groups.

These considerations suggest that the groups that we analyse are not complete in terms of membership. The aim of this study is however to characterise the global properties of the stellar population in the Orion region. A more detailed analysis of the physical properties for which a complete membership list is important, such as the initial mass function, is left to future studies.

3.6.2 Ages

The results obtained by fitting isochrones to the colour-magnitude diagrams of the groups isolated in Section 4 confirm the existence of the old population towards the 25 Ori group found by Kos et al. (2018), which corresponds to our group $B_{0,b}$. Kos et al. (2018) derive an age of 20 Myr, while we obtain an age of 15 Myr. This could depend on the different extinction values used or by a slightly different membership list. We also found that, towards the Belt, group E, $B_{5,b}$, $B_{7,b}$, and B_8 are older than 10 Myr, and that some older sources are also found in the Orion A region (group F). The population in front of the Orion A cloud (group B_1) is around 10 Myr old. The age is similar to the estimated age for the group related to the Orion A cloud (group C). However, the colour-magnitude diagram of group C (see Appendix B) shows that, not unexpectedly, many sources are brighter than the 10 Myr isochrone, and therefore likely younger.

A substantial luminosity spread has been observed in the colour magnitude diagram of the stellar population towards the ONC (see for example Jeffries et al. 2011; Da Rio et al. 2010). This spread represents the combined effect of a real age spread, possibly due to the presence of multiple populations (Jerabkova et al. 2019; Beccari et al. 2017), and of an apparent spread caused by other physical effects that scatter the measured luminosities, such as stellar variability and scattered light from circumstellar material. Age spreads are not included in our data modelling, therefore our age estimate for group C should be considered as an upper estimate for the age of the stellar population towards the Orion A molecular cloud, which also contains younger sources. The older population is more numerous than the younger ones, and therefore our age estimates are biased toward older ages. The age estimate for group C and for all the other groups is very precise (see Table 2). This is partly an artefact of using a single isochrone set, and ignoring differential extinction as well as the effects mentioned above. The presence of unresolved binaries in our data is also not taken into account, and could introduce biases towards younger age estimates, as unresolved binaries appear brighter than single stars. This could be the case for example for groups B2 and B5 (see Fig. B). For the other groups the single star sequence is usually more numerous than the unresolved binary sequence, thus the fit results are weighted to-

wards the single star sequence.

In terms of age ranking, our age estimates agree with those found by Kounkel et al. (2018): their Fig. 13 indicates indeed the presence of a diffuse older population, which however they find to be around 10 Myr old. The difference in the maximum age they obtain is due to a number of differences in our fitting procedure: for example, they use $A_V = 0$ mag and a previous version of the *Gaia* DR2 filters. Our results contradict instead what was found by Briceño et al. (2019), who derive an age sequence that agrees with the long-standing picture of star formation starting in the 25 Ori region (also called Orion OB1a) and sequentially propagating towards the Belt region (Orion OB1b and 1c) and the Orion A molecular clouds (Orion OB1d).

3.6.3 Sequential star formation and triggering in Orion

The main result of this work is that the star formation history of the Orion region is complex and fragmentary. The Orion region is composed of many subgroups with different kinematic properties. Star formation started around 15 Myr ago (or 20 according to Kos et al. 2018), and still continues in the Orion A and B molecular clouds. The groups that we observe at the present time are sometimes spatially mixed (such as in the 25 Ori region) but their kinematics retain traces of their different origin. Figure 3.17 shows a schematic view of the Orion region, which summarises our results. The arrows represent the velocity vectors (in galactic Cartesian co-ordinates and corrected for the solar motion following Schönrich et al. 2010) of the groups we identified, and are colour-coded by the group ages. The grey contours represent the stellar density integrated in the Z (left), Y (centre), and X direction. The Sun is at $(X, Y, Z) = (0, 0, 0)$ pc.
Cantat-Gaudin et al. (2018b) studied the Vela OB association, finding that a large fraction of the young stars in the region are not concentrated in clusters, but rather distributed in sparse structures, elongated along the Galactic plane. Krause et al. (2018) performed a multi-wavelength analysis of the Scorpius-Centaurus association, and suggested a refined scenario to explain the age sequence of the sub-groups that form the association. Similar to these studies, we find that the star formation history of Orion is not consistent with simple sequential star formation scenarios. Further, the traditional groups in which the Orion OB association is sub-divided are not monolithic episodes of star formation, but exhibit significant kinematic and physical substructure.

We do not observe any clear age gradient nor any clear evidence of triggering in the kinematic properties of the groups (such as those predicted for instance by Hartmann et al. 2001). As Cantat-Gaudin et al. (2018b) suggest, the difference in velocity that are observed might be the result of galactic shear, or the consequence of a velocity pattern already imprinted in the filaments belonging to the parent molecular cloud these young populations formed from. The disposition in space of the clusters might reflect the structure of their parental molecular clouds: however this should be confirmed by specific simulations of the star formation process in the Orion region.

3.7 Conclusions

In this work we study the 3D structure, the kinematics, and the age ordering of the young stellar groups of the Orion star forming region, making use of *Gaia* DR2.

- We select young sources by applying simple cuts in the M_G vs. $G_{BP} - G_{RP}$ colour-magnitude diagram, and we study their density distribution in 3D galactic co-ordinates.

- We normalise our 3D density map between 0 and 1, and we select only the sources above a threshold of 0.5. We then apply the DBSCAN clustering algorithm to identify groups in 3D space and we analyse their properties in terms of ages and kinematics.

- We first inspect the proper motions of all the groups. In some cases we find that single groups in 3D space show sub-structures in their proper motion distribution. In this case we further sub-divide the groups, making simple cuts based on the proper motion distribution. We then apply a kinematic modelling code that we use to retrieve average motions, velocity dispersion, and isotropic expansion for all the groups identified.

- By comparing the 3D velocities of all the groups, we find evidence of kinematic sub-structures.

- We compute ages and extinctions for all the groups by using a 2D maximum likelihood approach.We find that star formation in Orion started around 15 Myr ago in two groups, one towards the Belt region, and one towards the 25 Ori region.

- We do not find any clear age gradient, or any evidence of sequential star formation propagating from the 25 Ori region towards the Belt region and the Orion A and B molecular gas.

In conclusion, the picture of the Orion that we obtain from this study is that of a highly sub-structured ensemble of young stars with different ages, with several kinematic groups, mixed in 3D space and overlapping in the sky. These results do not agree well with sequential star formation models, and would require designated specific simulations to be fully explained.

The limited number of radial velocities available for most of the groups, as well as their large uncertainties, does not allow to characterise fully the internal kinematics of the clusters, or establish the presence of an-isotropic expansion. Future, ground based spectroscopic surveys could provide precise radial velocities for a large sample of sources, which, combined with the next *Gaia* releases, will allow to better probe the internal kinematics of young clusters and OB associations.

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 extensive use of Matplotlib (Hunter 2007), 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.

3.A Testing the kinematic modelling code with simulated clusters

We generate a sample of N = 200 stars which mimics the kinematics properties of young clusters and we test our code by changing a) the position of the sample (in particular its distance to the Sun), b) the velocity dispersion, and c) the expansion coefficient (κ) value. In particular we are interested in the ability of the code to retrieve the correct value for κ, especially when not all the radial velocities of the cluster members are provided.

3.A.1 Simulation set up

The simulated star positions are drawn from Gaussian distributions with $\sigma = 2\,\mathrm{pc}$. The velocity of each simulated star is drawn following the same assumption as in L00, that is from a Gaussian distribution centred in $\boldsymbol{v_0}$ with a small velocity dispersion σ. We include expansion following Eq. 9, chosing $\kappa = 0.1\,\mathrm{km\,s^{-1}\,pc^{-1}}$.
We obtain the observed quantities (positions, parallax, proper motions, and radial velocities)[2] by adding typical *Gaia* errors in the Orion region drawn from Gaussian distribution with widths $0.1\,\mathrm{mas}$, $0.1\,\mathrm{mas\,yr^{-1}}$, and $3\,\mathrm{km\,s^{-1}}$ respectively.

3.A.2 Simple tests

We simulate two clusters at different distances and with different velocities (see Tables 1 and 2, respectively): cluster A is similar in terms of kinematics

$$\boldsymbol{v_{0,I}} = (-5.0, 45.0, 6.0)\,\mathrm{km\,s^{-1}}$$

and distance

$$(x_{0,I}, y_{0,I}, z_{0,I}) = (17.89, 42.14, 13.16)\,\mathrm{pc}$$

to the Hyades cluster; cluster B is instead resembling the 25 Ori cluster:

$$(x_{0,I}, y_{0,I}, z_{0,I}) = (52.96, 343.97, 10.21)\,\mathrm{pc}$$

and

$$\boldsymbol{v_{0,I}} = (0.0, 20.0, 0.0)\,\mathrm{km\,s^{-1}}$$

. We run the simulations in five different scenarios for both the simulated clusters:

1. $\sigma_v = 0.3\,\mathrm{km\,s^{-1}}$ and $\kappa = 0.1\,\mathrm{km\,s^{-1}\,pc^{-1}}$.

---

[2]To do the transformation we make use of the `pygaia` routine `phaseSpaceToAstrometry`.

2. $\sigma_v = 1.0\,\mathrm{km\,s^{-1}}$ and $\kappa = 0.1\,\mathrm{km\,s^{-1}\,pc^{-1}}$;

3. $\sigma_v = 0.3\,\mathrm{km\,s^{-1}}$, $\kappa = 0.1\mathrm{km\,s^{-1}\,pc^{-1}}$, and a fraction $f = [10\%, 50\%, 95\%]$ of stars without measured radial velocities.

The average velocities are always retrieved quite correctly in both cases; σ and κ are retrieved correctly for cluster A, however we notice that for cluster B the value of κ is usually underestimated, while σ is usually slightly over-estimated. When the number of observed radial velocities is too low, the expansion parameter can not be retrieved as it can not be separated from v_0 from astrometric data only. In the cases when this happens, we do not give any estimate for the expansion term κ. When there are no radial velocities available the velocity is very poorly constrained, especially for cluster B: in this case we do not give estimates for the velocities. When 10% or 50% of the measured radial velocities are missing, the errors on the estimated parameters are of the same order of magnitude as in the other cases were all the kinematic data are available. However, not unexpectedly, when only 5% of the radial velocities is available, the error on the v_y parameter is roughly one order of magnitude larger than in the other cases.

3.A.3   Realistic tests

In the real case it is likely that the clusters selected with the DBSCAN algorithm have both stars without measured radial velocities and kinematic outliers. We therefore further tested our code for cluster in two cases (see Table 3). In the first one we include 20 kinematic outliers in our simulated clusters: the kinematic outliers have a broader spatial distribution than the simulated cluster members ($\sigma = 5\,\mathrm{pc}$), and their velocities are drawn from a Gaussian distribution with mean $20\,\mathrm{km\,s^{-1}}$ in x_I, y_I, z_I, and dispersion $\sigma_v = 10\,\mathrm{km\,s^{-1}}$. In the second one we include 20 kinematic outliers and we remove the 10% of measured radial velocities. In both cases, after the exclusion procedure the parameters are retrieved correctly. We notice that also in this case the expansion coefficient κ is under-estimated (roughly by a factor of 2), while σ_v is slightly over-estimated.

3.A.4   Initial conditions

To test whether the initial conditions of the minimisation have an impact on the estimated parameters, we performed 100 runs with initial guesses for the mean cluster velocity components, the velocity dispersion, and the expansion term κ drawn randomly from a Gaussian distribution centred on the mean parameters, with dispersion equal to the 20% of their real values. Reino et al. (2018) performed similar tests on the Hyades cluster (which as said above is kinematically similar to our cluster A), finding essentially no dependence from the estimated parameters from the initial conditions. Thus, we repeat these tests only on our simulated cluster B.
B.1: $\sigma_v = 0.3\,\mathrm{km\,s^{-1}}$ *and* $\kappa = 0.1\,\mathrm{km\,s^{-1}\,pc^{-1}}$. We find that in general the minimisation results do not strongly depend on the initial parameters, however if the velocity dispersion σ_v is over-estimated and (or) the velocity in the x_I component is under- or over-estimated then the velocity in the y_I component is also under- or over-estimated.
B.2: $\sigma_v = 1.\,\mathrm{km\,s^{-1}}$ *and* $\kappa = 0.1\,\mathrm{km\,s^{-1}\,pc^{-1}}$. We find that the minimisation results

do not depend on the initial parameters in any case. This is reassuring, as the values for σ_v in the clusters considered here are larger than $0.3\,\mathrm{km\,s^{-1}}$. In the cases with $\sigma_v = 1.\,\mathrm{km\,s^{-1}}$ and missing radial velocities (for 20, 100, and 190 stars respectively), the estimated parameters are retrieved correctly for any choice of initial conditions, except for the expansion parameter κ, that is underestimated. If outliers are present, the parameters are retrieved correctly after the exclusion procedure.

Table 3.3: Results of the tests of the kinematic modelling for cluster A.

	$v_{x,I}$ [km s⁻¹]	$v_{y,I}$ [km s⁻¹]	$v_{z,I}$ [km s⁻¹]	σ_v [km s⁻¹]	κ [km s⁻¹ pc⁻¹]
Initial values	-5.0	45.0	6.0	0.3	0.1
A.1	-5.9	45.6	5.57	0.3	0.1
	-5.88 ± 0.03	45.57 ± 0.05	5.56 ± 0.027	0.32 ± 0.01	0.1 ± 0.01
A.2	-5.9	45.6	5.57	1.0	0.1
	-5.96 ± 0.08	45.5 ± 0.1	5.54 ± 0.08	1.01 ± 0.03	0.06 ± 0.02
A.3					
20/200 missing radial velocities	-5.873 ± 0.03	45.6 ± 0.06	5.586 ± 0.03	0.31 ± 0.01	0.1 ± 0.01
100/200 missing radial velocities	-5.91 ± 0.035	45.55 ± 0.07	5.564 ± 0.03	0.3 ± 0.01	0.1 ± 0.1
190/200 missing radial velocities	-6.0 ± 0.4	45.0 ± 1.0	5.5 ± 0.3	0.3 ± 0.01	0.1 ± 0.03

Table 3.4: Results of the tests of the kinematic modelling for cluster B.

	$v_{x,I}$ [km s^{-1}]	$v_{y,I}$ [km s^{-1}]	$v_{z,I}$ [km s^{-1}]	σ_v [km s^{-1}]	κ [km s^{-1} pc^{-1}]
Initial values					
B.1	0.0	20.	0.0	0.3	0.1
	0.66	19.73	0.53	0.3	0.1
	0.65 ± 0.03	19.83 ± 0.063	0.51 ± 0.026	0.36 ± 0.013	0.07 ± 0.01
B.2	0.66	19.73	0.53	1.0	0.1
	0.65 ± 0.07	19.8 ± 0.1	0.56 ± 0.07	1.0 ± 0.03	0.05 ± 0.01
B.3					
20/200 missing radial velocities	0.69 ± 0.03	19.80 ± 0.06	0.57 ± 0.02	0.33 ± 0.01	0.05 ± 0.006
100/200 missing radial velocities	0.67 ± 0.03	19.87 ± 0.08	0.52 ± 0.03	0.38 ± 0.01	0.05 ± 0.007
190/200 missing radial velocities	0.67 ± 0.043	19.89 ± 0.232	0.56 ± 0.025	0.32 ± 0.01	0.09 ± 0.008

Table 3.5: Results of the tests of the kinematic modelling for cluster B, with missing radial velocities and outliers.

	$v_{x,I}$ [km s^{-1}]	$v_{y,I}$ [km s^{-1}]	$v_{z,I}$ [km s^{-1}]	σ_v [km s^{-1}]	κ [km s^{-1} pc^{-1}]
No missing radial velocities and 20 outliers					
Initial parameters	0.0	20.	0.0	0.3	0.1
Real values	0.66	19.73	0.53	0.3	0.1
First iteration	0.62 ± 6.	19.88 ± 6.	0.71 ± 6	88.09 ± 2.4	-0.17 ± 0.4
After exclusion procedure	0.63 ± 0.03	19.75 ± 0.06	0.52 ± 0.03	0.33 ± 0.01	0.05 ± 0.006
20 missing radial velocities and 20 outliers					
Initial parameters	0.0	20.	0.0	0.3	0.1
Real values	0.66	19.73	0.53	0.3	0.1
First iteration	0.64 ± 4.85	19.88 ± 5.2	0.6 ± 4.85	72. ± 2.	0.4 ± 0.3
After exclusion procedure	0.68 ± 0.03	19.83 ± 0.06	0.55 ± 0.03	0.37 ± 0.01	0.07 ± 0.007

3.B Colour magnitude diagrams

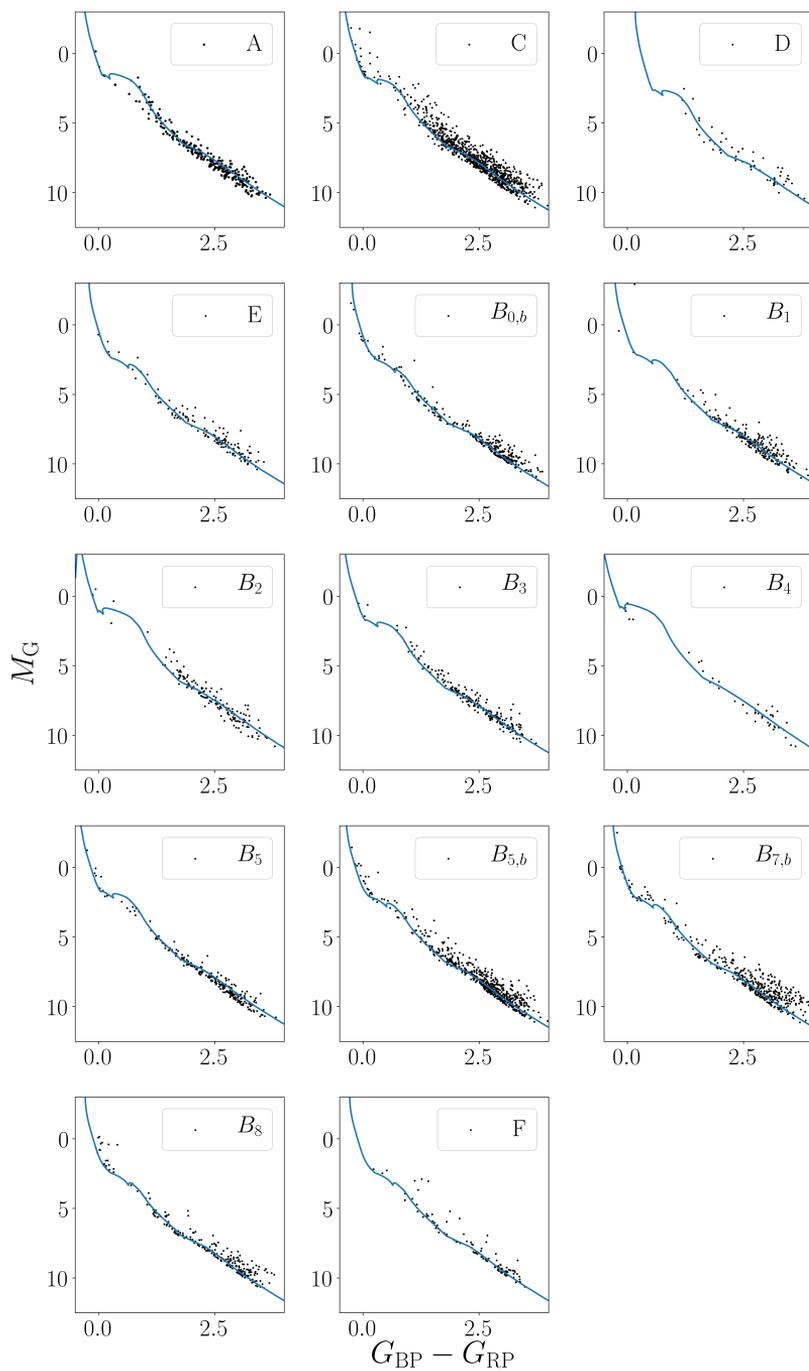Fig. 3.18 shows the colour magnitude diagram for the groups that we identified in Section 4.

Figure 3.18: M_G vs. $G_{BP} - G_{RP}$ colour magnitude diagram for the groups selected in Section 4. The blue solid lines correspond to the best fitting isochrones, derived in Section 5.

4

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

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 $500\,\mathrm{pc}$ 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 $l, b \sim 218.5°, -2°$, 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.

4.1 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 $\sim 20°$ 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) suggests 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. 2006b,a, 2009; Bouy & Alves 2015) and our knowledge of the structure of the solar neighbourhood amplified.

In particular, Elias et al. (2006b) first studied the three dimensional spatial distribution of early type stars within $1\,\text{kpc}$ 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. (2006b) 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 dimensional 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. 4.2 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. 4.3 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. 4.4 we discuss our findings. Finally in Sec. 4.5 we summarize our results and draw our conclusions.

4.2 Data

In this section we present the selction criteria that we used. We refer to Gaia Collaboration et al. (2016a, 2018a) 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 $d = 500\,\mathrm{pc}$ from the Sun ($\varpi \geqslant 2\,\mathrm{mas}$), 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[1]. We shortly describe the catalogue columns in Appendix F.

4.2.1 Extinction correction

G band extinction, A_{G}, and colour excess, $E(G_{\mathrm{BP}} - G_{\mathrm{RP}})$, 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 $E(G_{\mathrm{BP}} - G_{\mathrm{RP}})$ based on the 3D map. In this way, we aim at producing a de-reddened 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,

---

[1]The UMS and PMS catalogues are only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via `http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/`

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 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 $\varpi > 2\,\mathrm{mas}$, $\varpi/\sigma_\varpi > 5$ and a measured A_G value. Then, we compute the source galactic Cartesian coordinates, x, y, z. We define a volume $N = 1000 \times 1000 \times 1000\,\mathrm{pc}$ centred on the Sun and we divide it in cubes n of $10 \times 10 \times 10\,\mathrm{pc}$ 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.

4.2.2   Upper Main Sequence

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

$$M_G \leqslant 4.4\,\mathrm{mag};$$
$$(G_{\mathrm{BP}} - G_{\mathrm{RP}}) \leqslant 1.7\,\mathrm{mag};$$
$$\varpi >= 2\,\mathrm{mas}; \tag{4.1}$$
$$\varpi/\sigma_\varpi > 5 \tag{4.2}$$

By using the extinction A_G and colour excess $E(G_{BP} - G_{RP})$ values computed in Section 4.2.1, we correct the colour-magnitude diagram for extinction and reddening, and apply the following selection criteria:

$$M_{G,0} \leqslant 3.5\,\mathrm{mag};$$
$$(G_{\mathrm{BP}} - G_{\mathrm{RP}})_0 \leqslant 0.4\,\mathrm{mag};$$

$$\tag{4.3}$$

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, $(d = 1000/\varpi\,\mathrm{pc})$, and this holds only when parallax errors are small (Bailer-Jones 2015). Fig. 4.1(left) shows the initial colour-magnitude diagram used for the selection. Fig. 4.1(right) shows the conditions on colour and magnitude as black dashed lines.
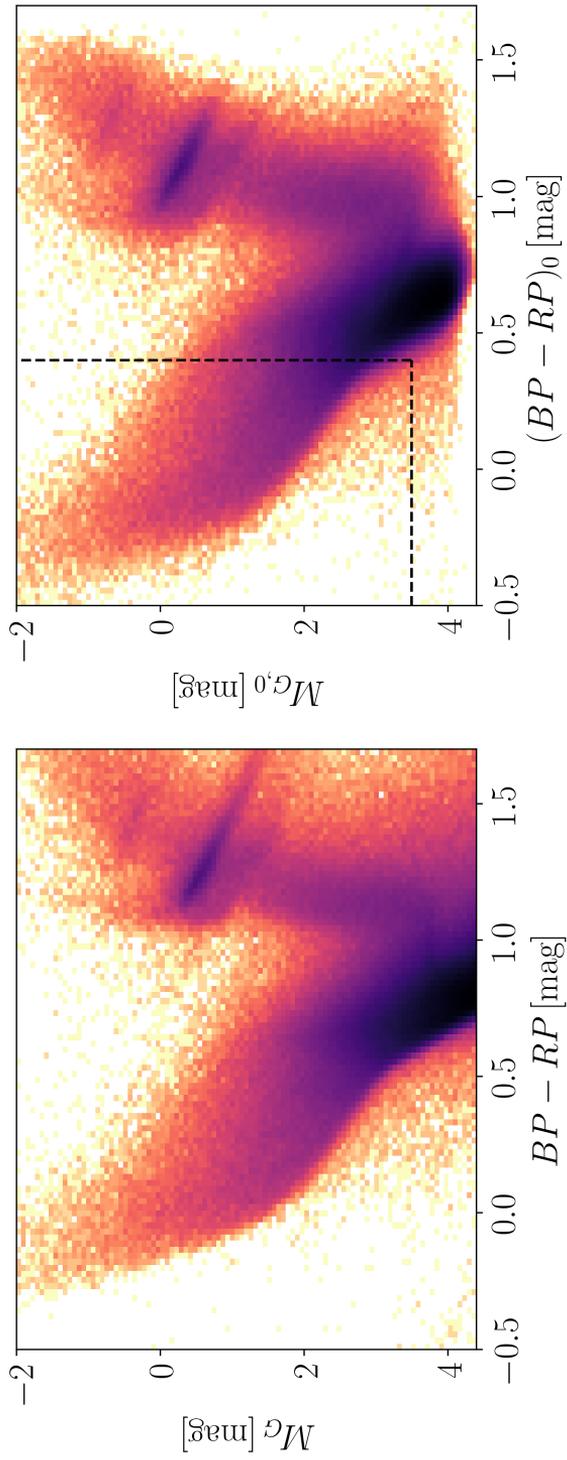
Figure 4.1: Upper main sequence colour magnitude diagrams. Left: colour magnitude diagram before correcting for extinction and colour excess. Right: colour magnitude diagram after accounting for extinction and reddening. The dashed lines limit the region we considered as upper-main sequence in this study.
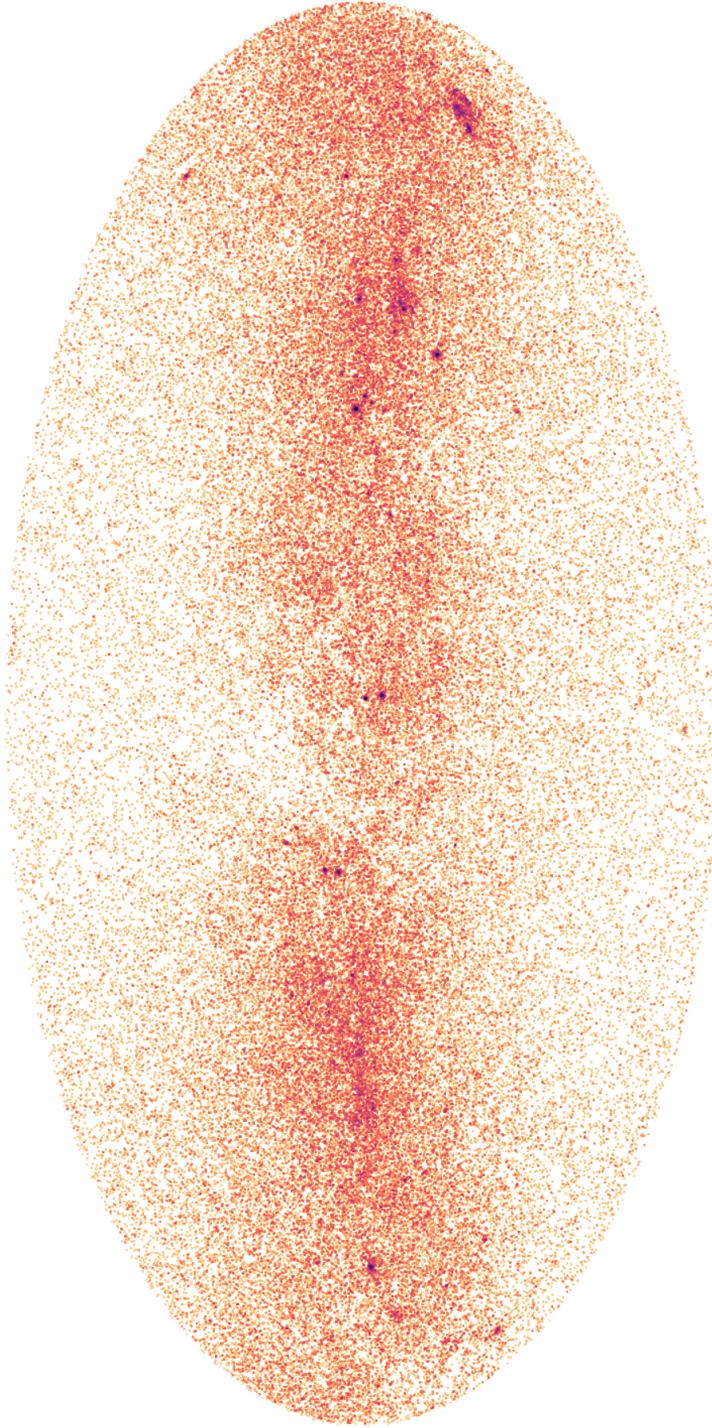
Figure 4.2: Upper main sequence sources selected by applying the conditions in Sec. 4.2.2. The sources are concentrated towards the galactic plane, and their density decreases towards the poles. Clumps corresponding to known open clusters and associations are visible.
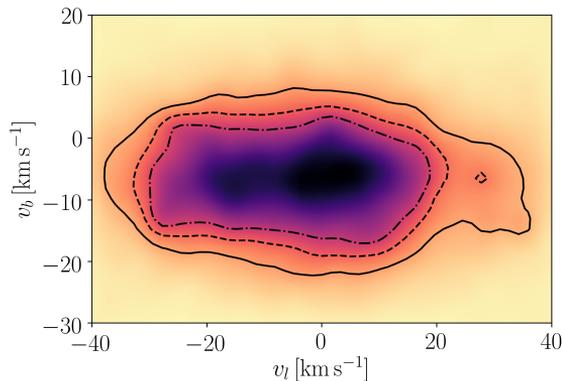
Figure 4.3: Smoothed tangential velocity distribution of the UMS sample, defined in Eq. 3 in the text. The contours represent the $S = 1, 2, 3$ levels. The density enhancements correspond to known clusters and associations. Note also that the distribution is not centred in $v_l, v_b = (0, 0)$ due to the solar motion.

Tangential velocities

Fig. 4.2 shows the distribution of the UMS sources selected in Sec. 4.2.2. 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 $A = 4.74047 \, \mathrm{km s^{-1} yr^{-1}}$) of the stars we have selected so far.

Fig. 4.3 shows an unsharp mask of the tangential velocity distribution of the UMS sample. We use a two-dimensional gaussian filter, with bandwidth $= 30 \, \mathrm{km s^{-1}}$ 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 $= 1 \, \mathrm{km s^{-1}}$. Finally we compute the quantity:

$$S = \frac{I_1 - I_{30}}{I_{30}}, \tag{4.4}$$

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. 4.3. Fig. 4.4 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. 4.2: 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 early-type population and not on the kinematic outliers, which represent a small fraction of the population.
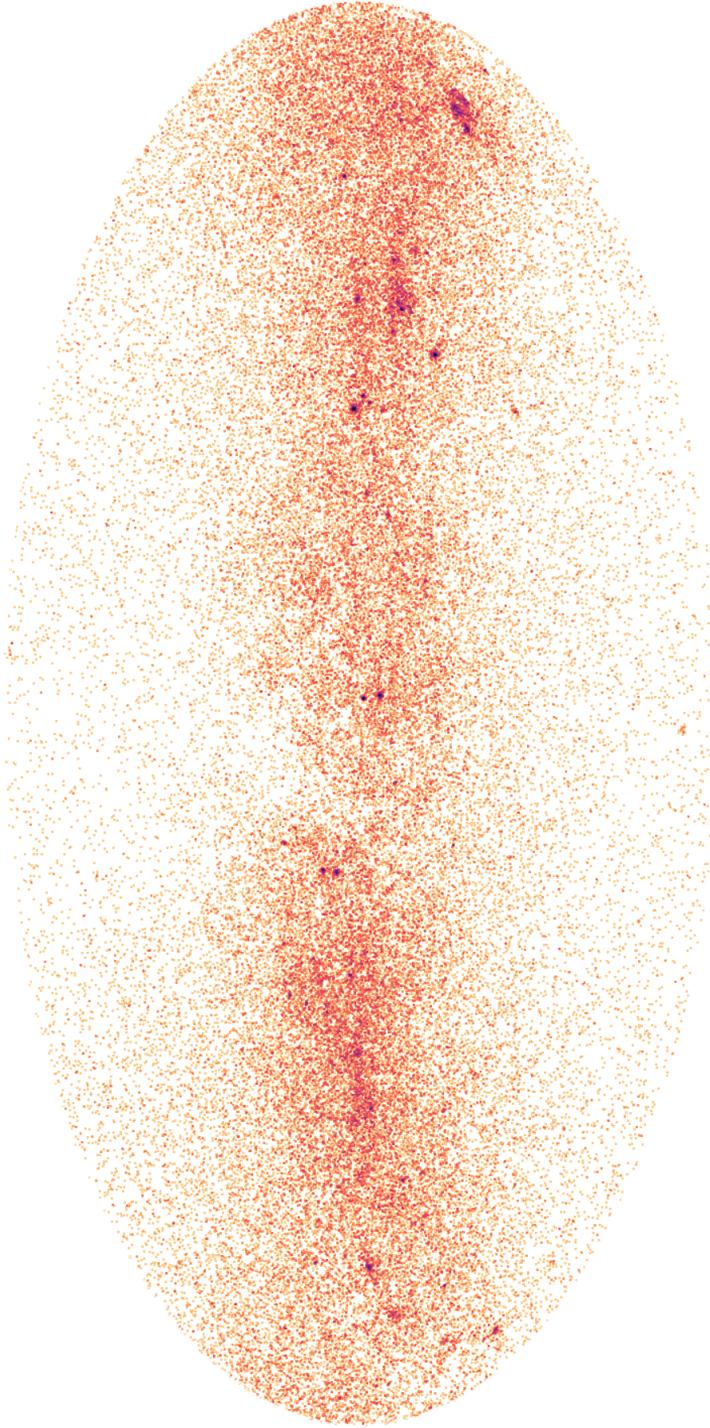
Figure 4.4: Distribution of the sources in the Sky after the selection based on tangential velocities. The number of sources at high galactic latitudes has decreased with respect to Fig. 4.2, which indicates that many contaminants have been discarded.

4.2.3 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.

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 ($d < 100 \, \text{pc}$). 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 $40 \, \text{km s}^{-1}$:
$v_t = \sqrt{v_l^2 + v_b^2} < 40 \, \text{km s}^{-1}$.
The condition on the tangential velocity follows Gaia Collaboration et al. (2018b). 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.

Extinction correction and selection of the PMS

We first corrected for extinction and reddening using the procedure described in Section 4.2.1. Then, we used the PARSEC Isochrones (Bressan et al. 2012) version 1.2S (Chen et al. 2014, 2015; Tang et al. 2014) with $A_V = 0 \, \text{mag}$ 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 $M_{G,0} > 4 \, \text{mag}$: 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. 4.5 shows the color magnitude diagram of the selection. We note that for $M_{G,0} \sim 7 \, \text{mag}$ 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 isochrone [2]. Fig. 4.6 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 $l < 220°$;

- *Vela*, at $240° < l < 270°$;

---

[2]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.
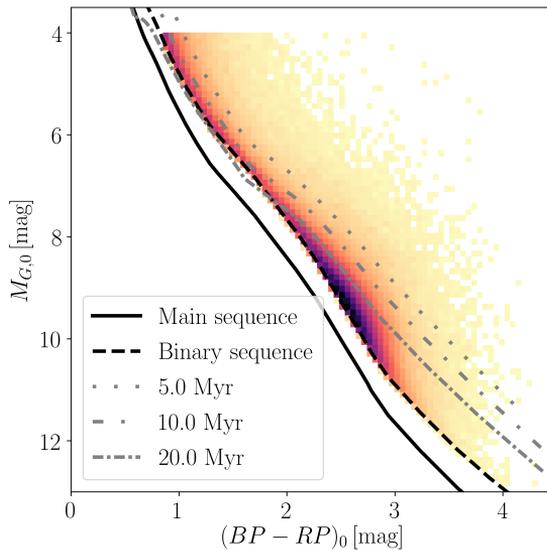
Figure 4.5: $G_{\mathrm{BP}} - G_{\mathrm{RP}}$ vs. M_G colour-magnitude diagram of the sources selected in Section 2.2.2. The density of sources increases towards the binary sequence.

- *Scorpius-Centaurus* and *Ophiucus*, at $l > 280°$ and positive b;

- *Chamaeleon*, at $l, b \sim (300°, -16°)$;

- The *Aquila rift*, at $l, b \sim (30°, +3°)$;

- *Lacerta*, at $\sim (100°, -20°)$;

- *Cepheus* and *Cassiopeia*, at $l > 100°$, above and slightly below the galactic plane;

- *Taurus* and *Perseus*, at $l > 140°$, 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 $A_{\mathrm{G}} > 0.92\,\mathrm{mag}$. We chose this threshold after studying the extinction distribution of our sample: the median of the distribution is $0.51\,\mathrm{mag}$, while the $16th$ percentile is $0.30\,\mathrm{mag}$ and the $84th$ percentile is $0.92\,\mathrm{mag}$. 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. 4.7 shows the distribution in the sky of the sources remaining after the extinction cut.
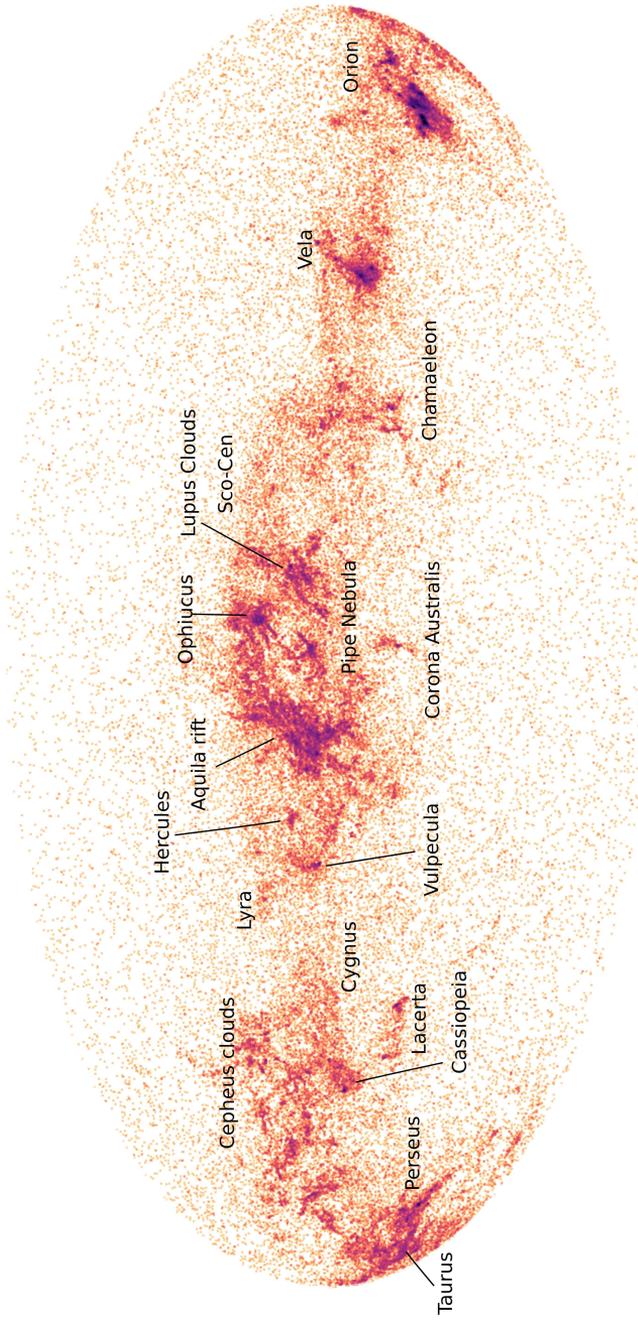
Figure 4.6: Pre-main sequence sources younger than 20 Myr.
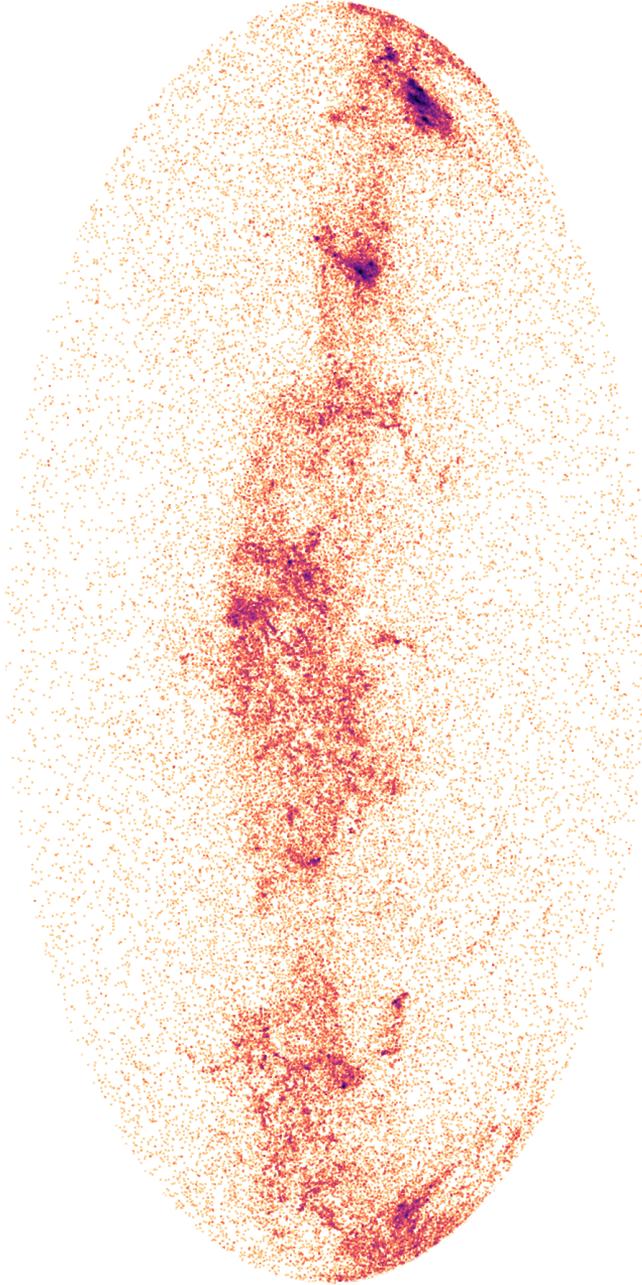
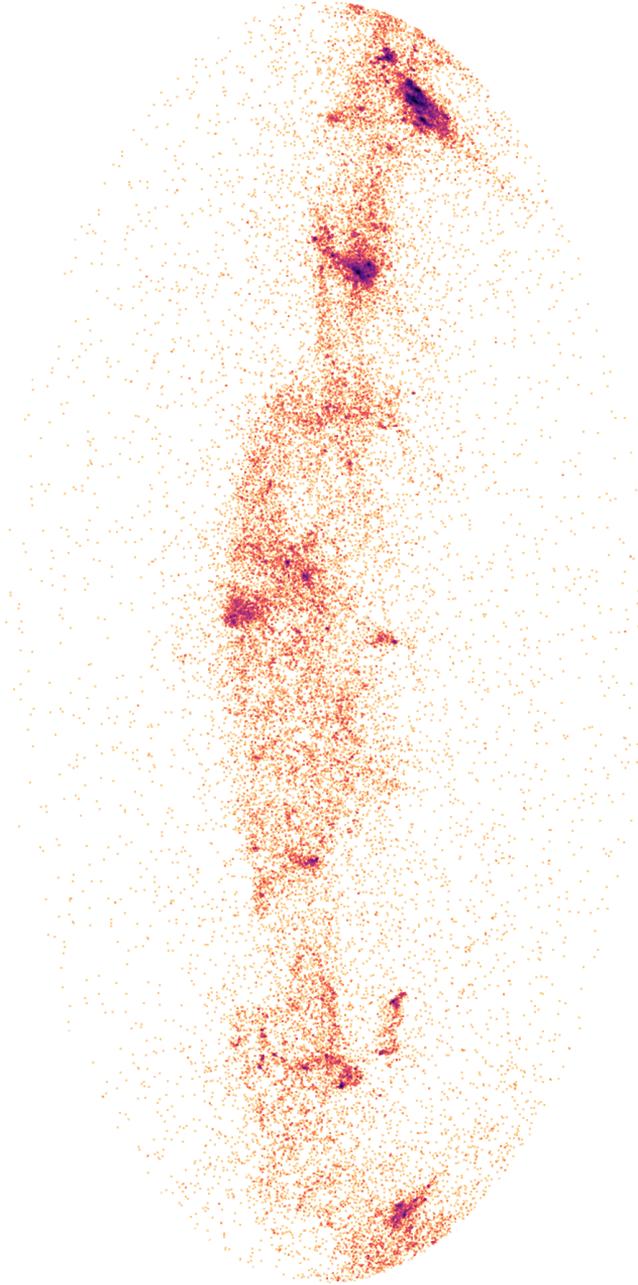**Figure 4.7:** Pre-main sequence sources younger than 20 Myr, with $A_G < 0.92\,\mathrm{mag}$

Figure 4.8: Pre-main sequence sources younger than 20 Myr, with $A_G < 0.92$ mag, and within the $S = 1$ level of Fig. 4.9.

Tangential velocities

As in Sec. 4.2.2, 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. 4.9 shows the tangential velocity distribution defined in Eq. 4.3 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. 4.8. 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.

4.3 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 4.4.

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 $V = 1000 \times 1000 \times 700 \, \text{pc}$ centred on the Sun. We divide the cube in volume elements v of $3 \times 3 \times 3 \, \text{pc}$. 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 ([3]). 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:

$$f(\text{x}) = \frac{L}{1 + e^{-k(\text{x} - \text{x}_0)}} - 1, \tag{4.5}$$

where $\text{x} = D(x, y, z)$, 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, $f(\text{x})$ 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]The python function used for the smoothing is scipy.ndimage.filters.gaussian_filter()
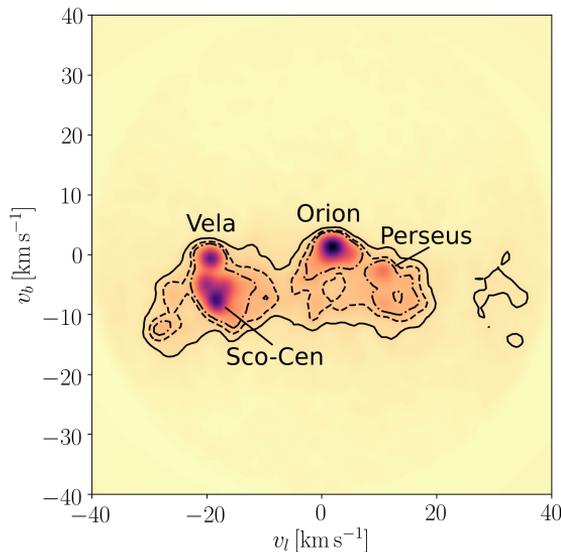
Figure 4.9: Tangential velocity plot of the PMS sample selected in Section 2. Many clumps are visible and correspond to known associations and clusters. The four most prominent structures are: Orion, Sco-Cen, Vela and Perseus. Note the gap around $20\,\mathrm{km\,s^{-1}}$, visible also in Fig. 4.13.

4.3.2 Results

Fig. 4.10 (left) shows the density distribution of pre-main sequence sources younger than $20\,\mathrm{Myr}$ on the galactic plane (X is directed towards the galactic Centre, Y towards galactic rotation, the Sun is at $(0, 0, 0)$). Fig. 4.10 (right) shows the density distribution perpendicular to the plane. Fig. 4.11 shows the density distribution of the UMS sample. The axes are the same as in Fig. 4.10.

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 $\sim 10\,\mathrm{Myr}$ ago, while NGC 2547 and a newly discovered cluster formed $\sim 30\,\mathrm{Myr}$

ago. Cantat-Gaudin et al. (2018a) 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 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. 4.11: 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. 4.11 and the PMS map of Fig. 4.10). 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 $(x, y) \sim (-250, -250)$ pc ($l, b \sim 218.5°, -2°$) of Fig. 4.10. 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 $0 < y < 100$ 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. (2002a,b) studied an area of $\sim 20°$ 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. 4.11, 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 $250 < y < 500$ 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 $d < 500\,\mathrm{pc}$ 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 $900\,\mathrm{pc}$, and therefore beyond the boundaries of our region. The groups in Fig. 4.10 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 $\mathrm{LkH}\alpha\,198$ and associated with the dark cloud L 1265, plus other small cluster in the same area.

3. *Aquila*: $x > 100$ and $50 < y < 200$ 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. 4.10, while Cygnus is visible in both Fig. 4.10 and 4.11, 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 4.4. We note here that Cyg OB4 and Cyg OB7 de Zeeuw et al. (1999) are beyond the region studied in this work ($d > 500\,\mathrm{pc}$). 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 $50-100\,\mathrm{pc}$ groups younger than $20\,\mathrm{Myr}$ 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 4.4. 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 4.4.
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.

4.3.3 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. $\tau \leqslant 5\,\mathrm{Myr}$;

2. $5 \leqslant \tau \leqslant 10\,\mathrm{Myr}$;

3. $10 \leqslant \tau \leqslant 20\,\mathrm{Myr}$;

Fig. 4.12 shows the density distribution of stars $\leqslant 5\,\mathrm{Myr}$ (red), $\leqslant 10\,\mathrm{Myr}$ (green), $\leqslant 20\,\mathrm{Myr}$ (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 $\sim 10\,\mathrm{Myr}$ for the PMS population in Vela. In Perseus, the young cluster IC 348 is visible. The red cluster in $(X, Y) \sim -30, 0\,\mathrm{pc}$ belongs to the Taurus star forming region. The groups at large positive Y values are instead more evolved.

4.3.4 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 ($|b| > 45°$) 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
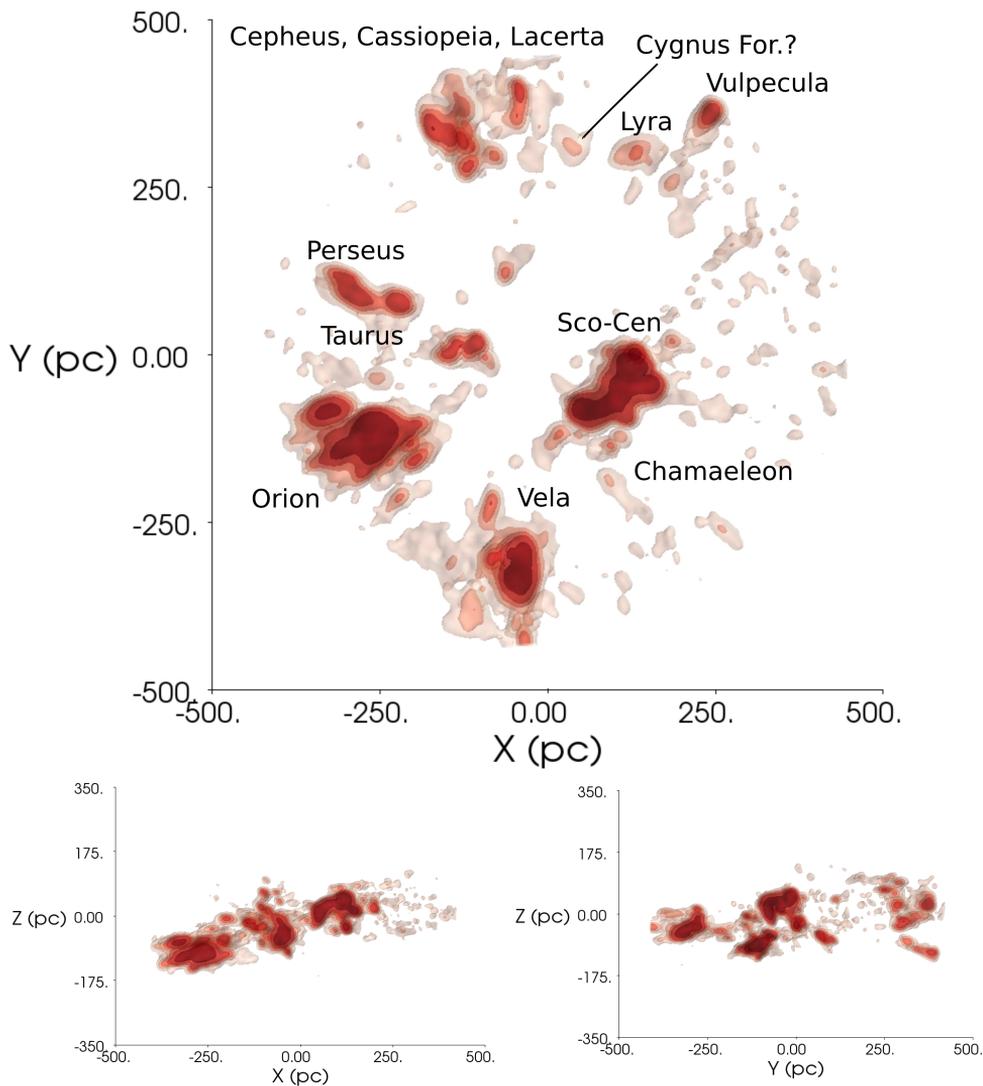
Figure 4.10: Top: 3D density distribution of PMS sources younger than 20 Myr on the galactic plane. The Sun is in $(0, 0)$, the x-axis is directed towards the galactic centre, and the y-axis towards the direction of the galactic rotation. The z-axis is perpendicular to the plane. The contours represent the 0.2, 0.4, 0.6, 0.8, and 1 density levels. Bottom left: 3D density distribution of the PMS sample (age < 20 Myr) perpendicular to the galactic plane. Contour levels are the same as on the left. Bottom right: 3D density distribution of the PMS sample (age < 20 Myr) along the rotation axis.
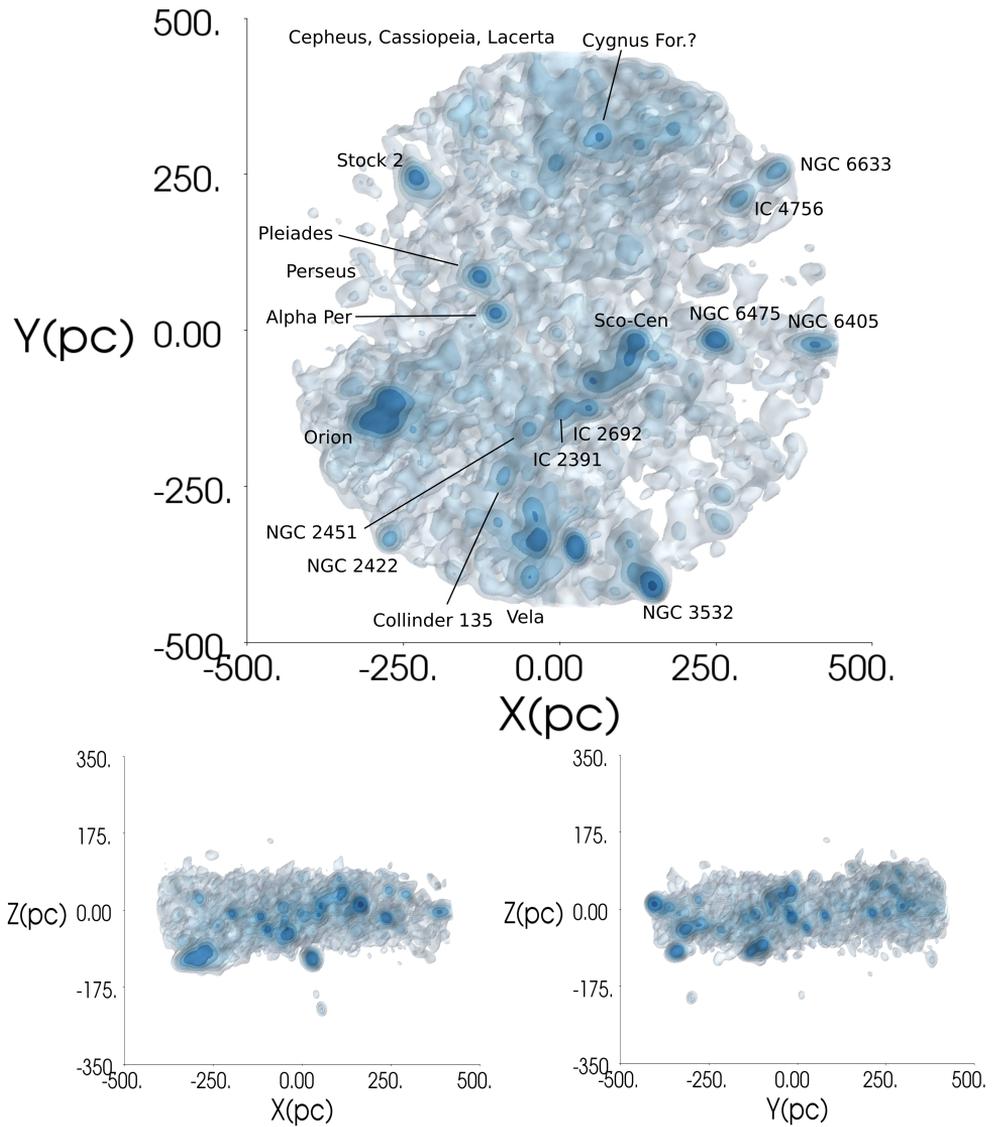
Figure 4.11: Same as Fig. 4.10, but for the upper-main sequence sample selected in Section 2.1. The contours represent the 0.2, 0.3, 0.4, 0.6, 0.8, and 1 density levels.
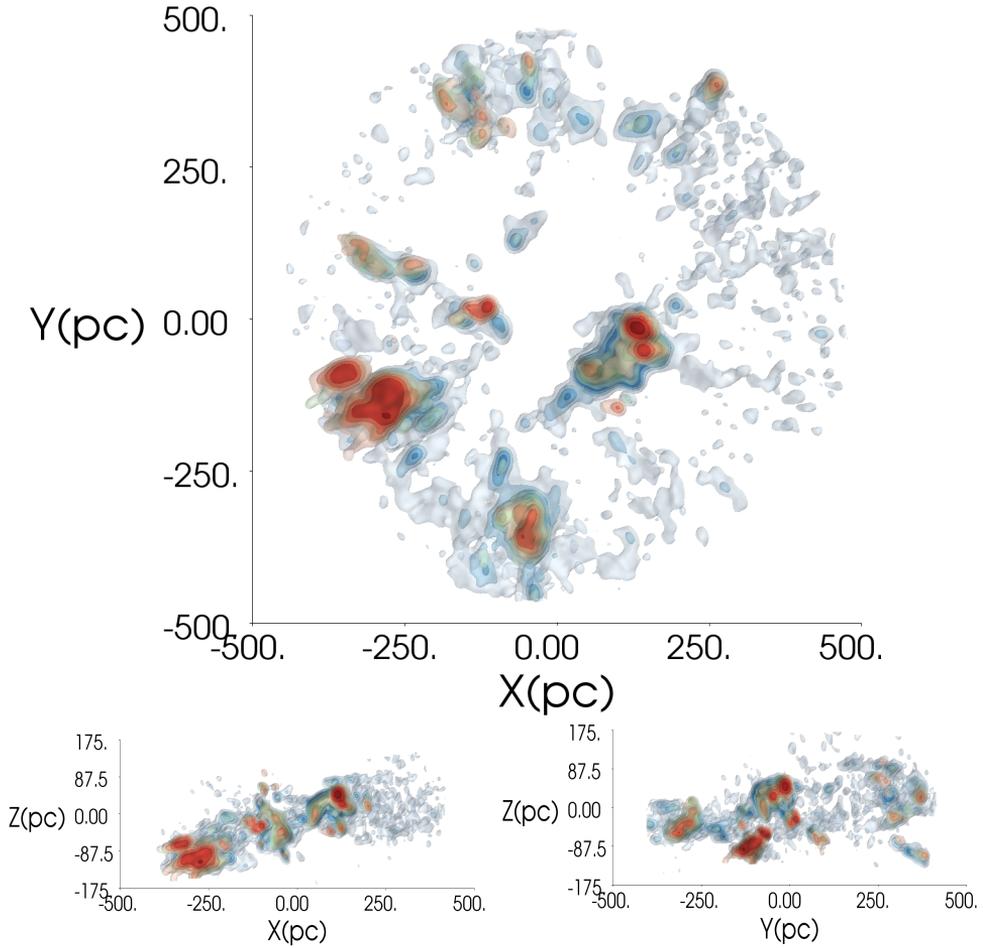
Figure 4.12: 3D maps of sources younger than 20 Myr and older than 10 Myr (blue), younger than 10 Myr and older than 5 Myr (green), and younger than 5 Myr (red). The contours are the same as in Figs. 4.10. In Fig. 4.16 we show separate maps of the $X - Y$ plane for each of the age intervals.

ecliptic poles (see Appendix 4.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. 4.10, right): the density does not look as a uniform slab (compare with the UMS distribution of Fig. 4.11, 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. 2018a). 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 $G = 21 \, \text{mag}$ and parallax error $\sigma_\varpi \sim 1\text{mas}$ (see Fig. 7 in Gaia Collaboration et al. 2018a), would be considered part of our sample until $\varpi = 5 \, \text{mas}$ ($d = 200 \, \text{pc}$) and excluded for smaller parallaxes ($d > 200 \, \text{pc}$). 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. 4.2 and 4.4 show essentially a uniform distribution of sources on the galactic plane, without any evident sign of extinction, Fig. 4.6 clearly show the outline of nearby molecular clouds. To exclude extinted sources we resolved to eliminate all the PMS sources with $A_G > 0.92 \, \text{mag}$. 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. 4.10 and 4.19 (where in the latter the condition $A_G < 0.92 \, \text{mag}$ 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

4.4   Discussion

In the previous sections, we analysed the spatial distribution and the age ordering of young stellar population within $d = 500 \, \text{pc}$ 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. 4.10 and 4.12. 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 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. 4.13 shows the tangential velocities along galactic latitude of the UMS (top) and the PMS sample (bottom) older than $20\,\mathrm{Myr}$, 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 $(U_\odot, V_\odot, W_\odot = (11.1, 12.24, 7.25)\,\mathrm{km\,s^{-1}}$, Schönrich et al. 2010). The location of the groups in the v_l 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 4.3. 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. Further, 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. 4.13 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. 4.13 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 4.3, PMS stars show a gap in their distribution in the inner $\sim 50\,\mathrm{pc}$. This is not unexpected as vicinity of the Sun ($d < 50\,\mathrm{pc}$) 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 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.
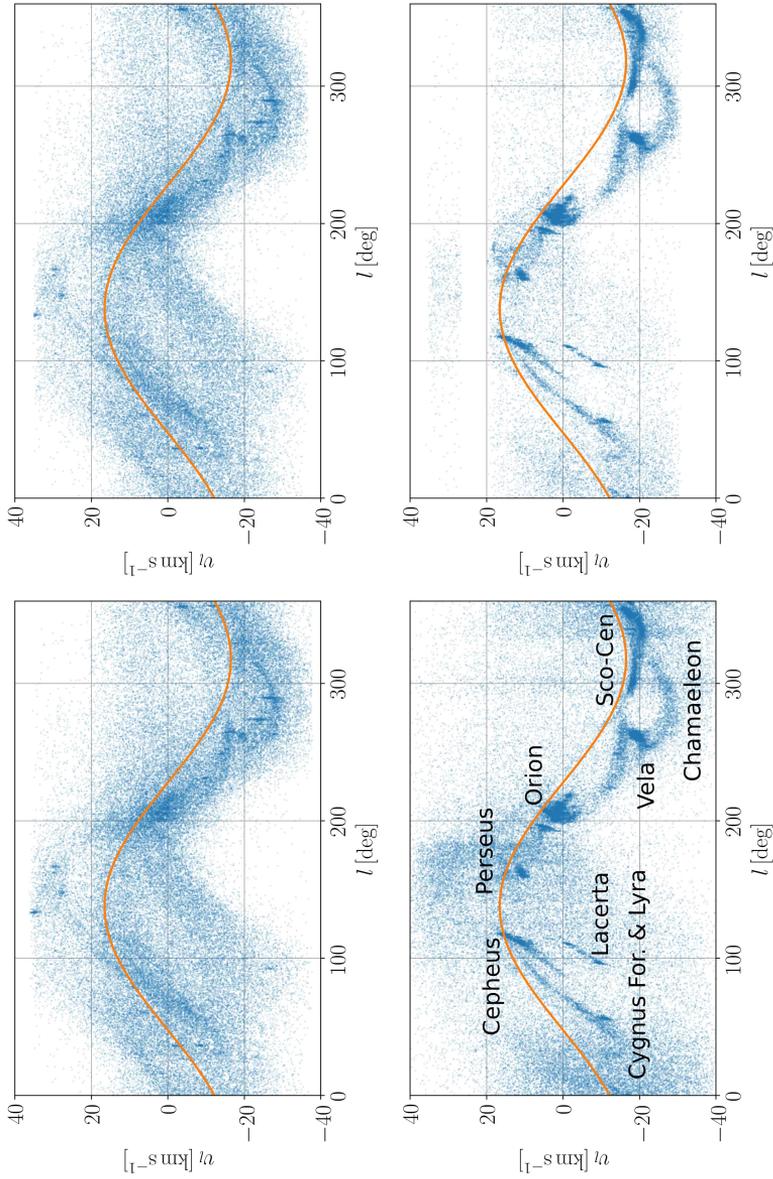
Figure 4.13: Tangential velocity along galactic latitude *vs.* latitude for the UMS (top) and the PMS (bottom) sample, before (left) and after (right) the tangential velocity selection. The solid orange line shows the projection of the Sun motion. The 'gaps' in the scatter plots on the left are due to the tangential velocity selection (see Section 2 in the text).

The age map of Fig. 4.9 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. 4.14, which shows the same sources as in Fig. 4.12, 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 binaries stand out as a separate sequence, which, being brighter by $\sim 0.75\,\mathrm{mag}$ 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. 4.14 we observe a density enhancement at coordinates $l, b \sim 343°, +5°$: 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. 4.13). The candidate open cluster at $l, b \sim 220°, -2°$, $X, Y \sim (-250, -250)$ pc, has an age $> 10\,\mathrm{Myr}$ and might be related to the Orion dust ring discovered by Schlafly et al. (2015). Cantat-Gaudin et al. (2018a) 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. 4.14: 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. 4.13 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. 4.5, 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. 4.15. 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. 4.10. 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. 4.11). The region surrounding the Sun presents a lack of sources, which is however less pronounced than in the PMS map of Fig. 4.10. This represents additional evidence that there is a real gap for the youngest stars, extending out to $\sim 100\,\mathrm{pc}$ towards Scorpius-Centaurus and reaching $\sim 200\,\mathrm{pc}$ 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 $20-30\,\mathrm{Myr}$ ago due to the events that created the Local Bubble (Alves et al. 2018; Lallement et al. 2014; Puspitarini et al. 2014).

4.5 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 $d = 500\,\mathrm{pc}$ 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
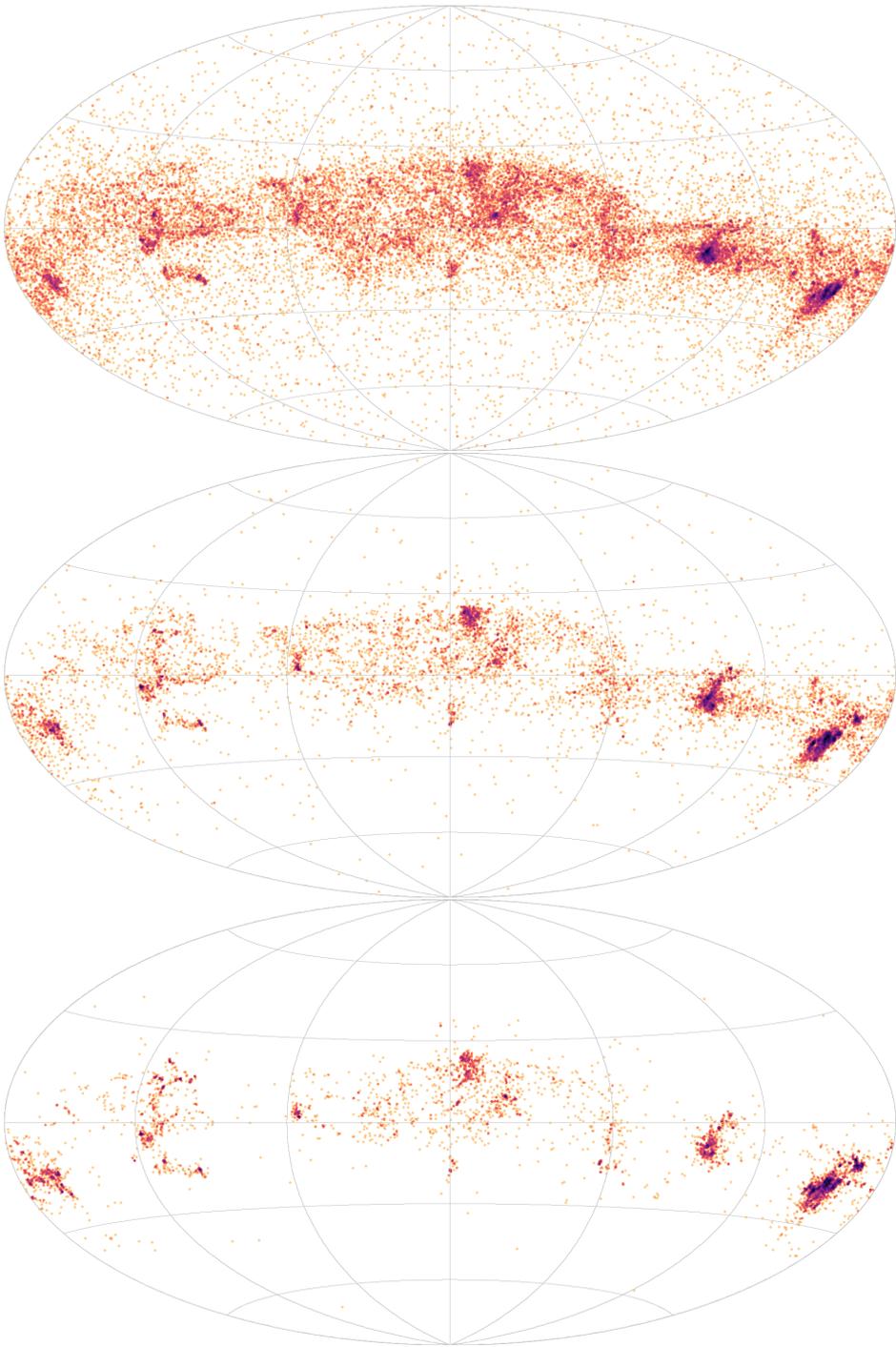
Figure 4.14: Sky projection of sources with different ages. Top: sources with $10 < t < 20\,\mathrm{Myr}$; centre: sources with $5 < t < 10\,\mathrm{Myr}$; bottom: sources with $t < 5\mathrm{Myr}$.
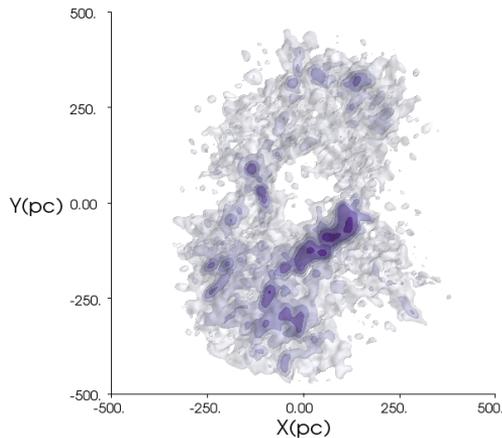
Figure 4.15: 3D map of sources older than 20 Myr. The contours represent the 0.2, 0.4, 0.6, 0.8, and 1 density levels.

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. 4.10) of the sources younger than 20 Myr shows a gap in the innermost $50 - 100 \, \text{pc}$. This is due to 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. 4.11), 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. 4.10 and 4.11: 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 $l, b \sim 220°, -2°$. Due to its position, distance, and age, this cluster might be related to the Orion star forming complex.

- 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.

4.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 $\varpi > 2\,\mathrm{mas}$, 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
```

4.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-pyhsical over-densities in the data due to the fact that *Gaia*'s scanning law favours the ecliptic poles ($|b| > 45°$). This effect is well visible when studying the distribution in the sky of *all* the sources within $d = 500\,\mathrm{pc}$ before and after applying the condition $\sigma_\varpi/\varpi > 5$. Fig. 4.16 (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 law[4]. The region where we observe the lowest values of completeness is towards the galactic plane for small positive b values.

4.C New cluster at $l, b \sim (218.5°, -2°)$

As mentioned in the main text of the paper, we report the discovery of a candidate young cluster centred roughly at $l, b = (218.5°, -2°)$. Fig. 4.17 shows the proper motion diagram (left), the parallax distribution (centre), and the colour-magnitude

---

[4]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).
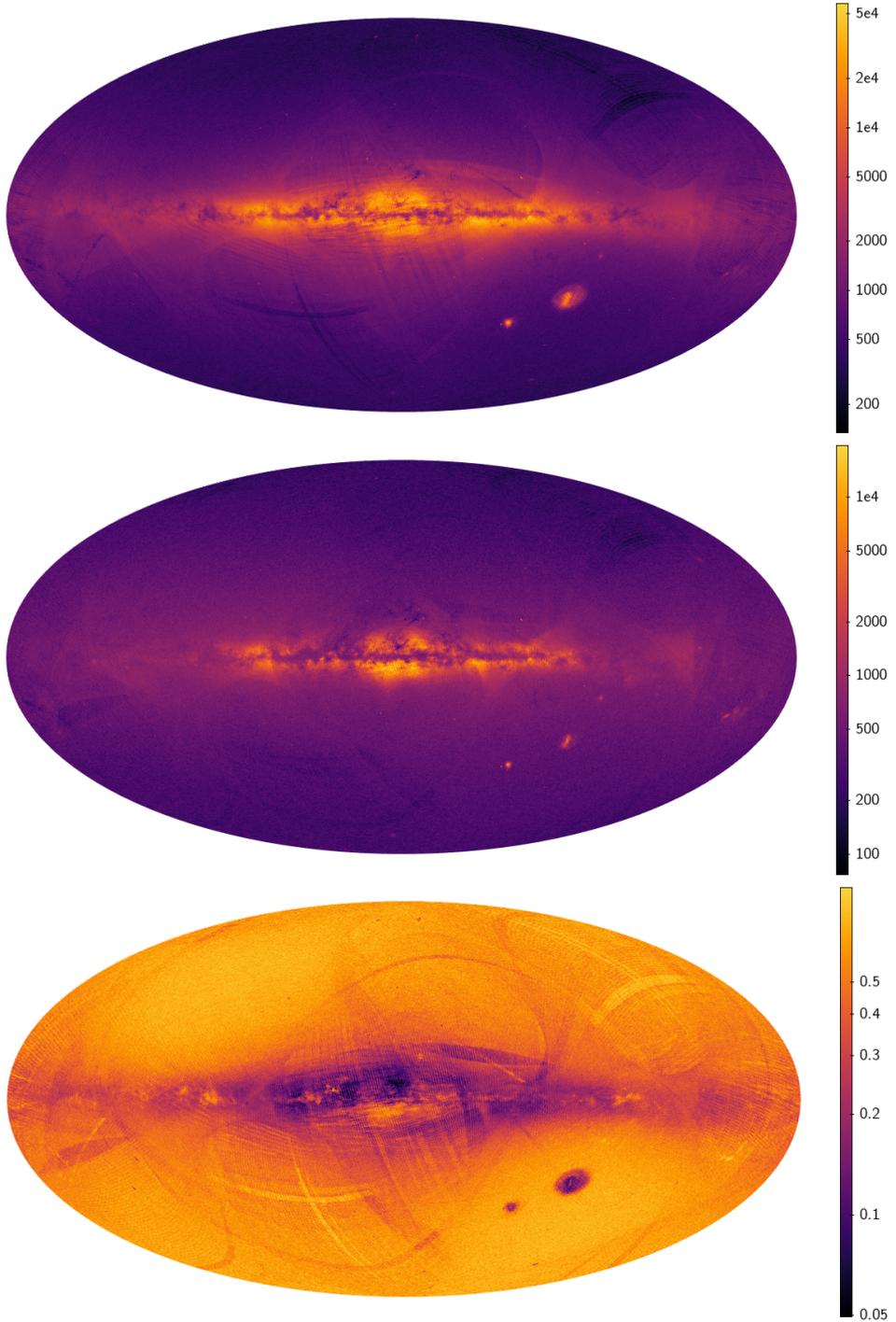
Figure 4.16: Top: Distribution in the sky of the sources within $d = 500\,\mathrm{pc}$. Centre: Distribution in the sky of the sources within $d = 500\,\mathrm{pc}$ and $\sigma_\varpi/\varpi > 5$. Bottom: Ratio between the distributions shown in the central and left panels.
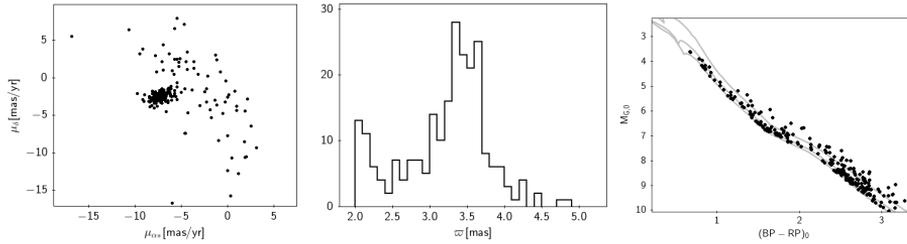
Figure 4.17: Left: Proper motion diagram of the sources selected in the region defined in the text. Proper motions cluster at $\mu_{\alpha*}, \mu_\delta \sim (-7., -2.5)\,\mathrm{mas\,yr^{-1}}$, with few, scattered outliers. Centre: Parallax histogram of the candidate cluster members. The histogram peaks 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.

diagram (right) of the sources within $215° \leqslant l \leqslant 222°$ and $-5° \leqslant b \leqslant 0°$. 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.

4.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. 4.16, right), younger than 10 Myr and older than 5 Myr (green, Fig. 4.16, centre), and younger than 5 Myr (red, Fig. 4.16, left).

4.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. 4.19 (left) shows the density map corresponding to the top panel of Fig. 6, while Fig. 4.19 (right) shows the density map corresponding to the central panel of Fig. 6. Fig. 4.19 (left) does not show any additional clustering with respect to Fig. 4.19 (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. 4.19 (left). Additional contaminants are removed by selecting stars according to their tangential velocity (compare Fig. 4.19 (right) with Fig. 4.10).

4.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 docu-
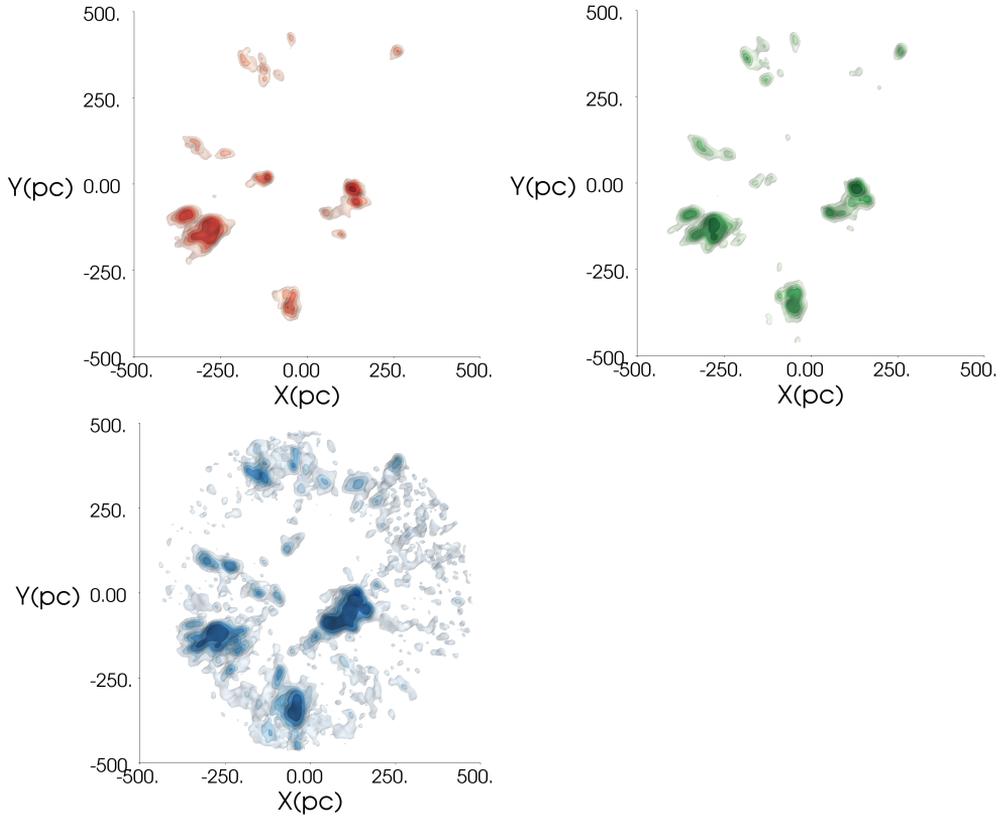
Figure 4.18: 3D density map of sources with age $10 < \tau < 20\,\mathrm{Myr}$ (top left), $5 < \tau < 10\,\mathrm{Myr}$ (top right), $\tau < 5\,\mathrm{Myr}$ (bottom).
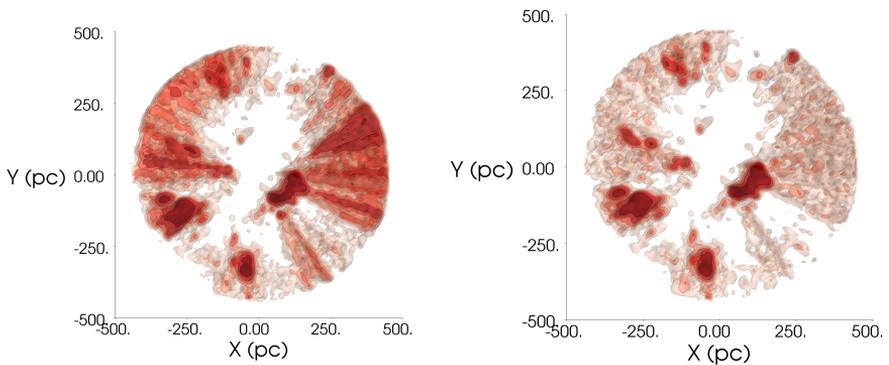

Figure 4.19: Left: 3D density map of the sources in the top panel of Fig. 6. Right: 3D density map of the sources in the central panel of Fig. 6.

mentation. 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];

- `pml_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).

5

Searching for runaway stars in Gaia DR2

We search for early type runaway stars within 1 kpc from the Sun by using the data of the second data release of the *Gaia* satellite (*Gaia* DR2) and the stellar parameters provided in the `StarHorse` catalogue. We select upper main sequence (UMS) sources by applying simple photometric cuts. Our sample consists of O-, B- and early A-type sources. We study the tangential velocity, and, when possible, the total velocity distribution of our sample, and we classify as candidate runaway stars those sources that have tangential velocities significantly different from the rest of the population (2σ) or total velocities higher than $30\,\mathrm{km\,s^{-1}}$. We study the orbits of the candidate runaway stars with literature radial velocities, and we find that around half of our candidates originated from beyond 1 kpc. We focus on the runaway star candidates in the Orion and Scorpius-Centaurus (Sco-Cen) regions. In Orion, we confirm previously known runaway stars and we enlarge the sample by adding 6 new runaway candidates. In Sco-Cen we identify two runaway star candidates that likely share the same origin. Finally, we discuss our findings in the context of other studies, and we estimate the completeness of our sample. More radial velocities are needed to obtain a more complete sample.

5.1 Introduction

O and B-type stars are often found in isolated locations, and do not appear to be members of clusters or associations. A large fraction of them moves at very high velocities: these are referred to as "runaway" stars (Blaauw 1952; Ambartsumian 1955; Hoogerwerf et al. 2001). The typical velocity threshold adopted to define a runaway star is $v > 30\,\mathrm{km\,s^{-1}}$ (Blaauw 1956; Gies & Bolton 1986; De Donder et al. 1997; Hoogerwerf et al. 2001; Dray et al. 2005; Eldridge et al. 2011), although sometimes $v > 40\,\mathrm{km\,s^{-1}}$ is used (Blaauw 1961; de Wit et al. 2005; Boubert & Evans 2018). Contrary to the majority of young, un-evolved massive stars, runaway stars present an almost complete absence of multiplicity (Chini et al. 2012; Sana et al. 2012; Almeida et al. 2017). Further, they show large rotational velocities and enhanced helium surface abundances (Hoogerwerf et al. 2001).

Two main mechanisms have been suggested to explain the origin of runaway stars: the binary supernova scenario (BSS) and the dynamical ejection scenario (DES).

According to the BSS (Blaauw 1961; Zwicky 1957; Boersma 1961), a runaway star receives its velocity when the primary component of the massive binary system explodes as a supernova. What remains of the binary after the explosion is a compact object, which, depending on the details of the preceding binary evolution, the eccentricity of the orbit, and the kick velocity due to the asymmetry of the supernova explosion, might or not remain bound to the runaway star (Renzo et al. 2019b).

In the DES (Poveda et al. 1967; Leonard 1991), runaways are formed through gravitational interactions between stars in dense, compact clusters. DES runaways have the following characteristics: 1) they are formed in high density environments, i.e. in young open clusters; 2) they do not show signs of binary evolution; 3) they are expected to be mostly single stars.

Which of the two formation mechanisms is mostly responsible for runaway stars is still unclear. The relative importance of the two scenarios has been investigated by studying the statistical properties of the ensemble of runaway stars or by focusing on individual runaways in detail.

By analysing a sample of 56 nearby runaway stars and 9 radio pulsars, Hoogerwerf et al. (2001) estimated that the disruption of binaries is responsible for roughly two thirds of observed runaways, however this claim could not be confirmed in the re-analysis of the same sample by Jilinski et al. (2010). Boubert et al. (2017) searched for runaways from the progenitors of nearby Galactic core-collapse supernova remnants (SNRs) by using the first data release of the *Gaia* satellite (*Gaia* DR1) and found likely companions for four SNRs. Tetzlaff et al. (2011) select young (age < 50 Myr) stars of any spectral type in the Hipparcos catalogue, and identify those with large peculiar velocities, finding in total 2547 candidate runaway stars. Boubert & Evans (2018) measured the fraction of runaway B-type emission-line stars (Be stars), and conclude that all Be stars could be explained by an origin in mass-transfer binaries. Maíz Apellániz & Weiler (2018) detected runaway stars by using *Gaia* DR1 proper motions, and suggest that the majority of them is produced by supernova explosions. Renzo et al. (2019a) investigated the kinematics of VFTS682, located in 30 Doradus (in the Large Magellanic Cloud), by combining the second *Gaia* data release (*Gaia* DR2) and *Hubble Space Telescope* data, and conclude that if was ejected from the central cluster.

Theoretical and numerical studies have focused on deriving analytic relations to iden-

tify dynamically formed runaway stars (Ryu et al. 2017), and predicting the peculiar velocities that massive binary systems obtain when the primary star collapses and disrupts the system, and investigating which physical processes leave a clear imprint and may therefore be constrained observationally (Renzo et al. 2019b). Renzo et al. (2019b) find in particular that the majority of the systems ejects a star at velocity below $30 \, \mathrm{km \, s^{-1}}$, i.e., a walkaway star (de Mink et al. 2012), and that runaways resulting from the disruption of binaries rarely exceed $60 \, \mathrm{km \, s^{-1}}$.

In this study we search for Upper Main Sequence (UMS) runaway stars within 1 kpc by using the `StarHorse` catalogue (Anders et al. 2019), which is based on the combination of *Gaia* DR2 with other photometric catalogues (PanSTARRS-1, 2MASS, and All-WISE). Our goal is to construct a catalogue of fast moving sources in the (extended) solar neighbourhood and to analyse their statistical properties. The data used are presented in Section 5.2, where we describe the quality flags applied and the UMS selection criteria. In Section 5.3 we describe the methods used to identify the runaway candidates and to study their orbits. In Section 5.4 we present our results, which are discussed in Section 5.5. Conclusions are drawn in Section 5.6.

5.2 Data

To create the `StarHorse` catalogue, Anders et al. (2019) combine parallaxes and optical photometry from *Gaia* DR2 with the photometric catalogues PanSTARRS-1, 2MASS, and AllWISE, and employ a Bayesian approach to derive stellar parameters, distances and extinction for *Gaia* DR2 sources brighter than $G = 18 \, \mathrm{mag}$. `StarHorse` does not contain all the *Gaia* DR2 sources, since the stars for which the code did not converge are not reported in the catalogue. This usually occurs for objects which cannot be fitted onto the model grid within the extinction limits used in the paper (e.g. highly reddened stars, F. Anders, private communication). The draw-back of this approach is that the sample is not complete, and most importantly, that estimating our incompleteness level is not trivial (see Section 5.5). However, we decided to use `StarHorse` to take properly into account extinction and reddening, as our selection of early type stars is strongly based on their position in the colour-magnitude diagram.

We consider stars nominally closer than 1 kpc ($\varpi > 1 \, \mathrm{mas}$), and we select them by applying the conditions `SH_OUTFLAG` = "00000" and `SH_GAIAFLAG` = "000", as recommended by Anders et al. (2019). The three digits of `SH_GAIAFLAG` correspond to:

- *Re-normalized unit weight error* (*RUWE*) *flag* (see Lindegren et al. 2018, Appendix C for the definition of the unit weight error and the web-page:https://www.cosmos.esa.int/web/gaia/dr2-known-issues for the definition of the RUWE). The first digit corresponds to 0 if the RUWE < 1.4.

- *Colour excess factor flag*. It corresponds to 0 if Eq. C.2 in (Lindegren et al. 2018) is satisfied.

- *Variability flag*. It corresponds to the *Gaia* DR2 `phot_variable_flag`.

The five digits of `SH_OUTFLAG` correspond to:

- *Main* `StarHorse` *reliability flag*: it is equal to 1 when the distance posterior probability is very broad (see Anders et al. 2019, for the definition of "broad").

- *Large distance flag*: for some stars `StarHorse` delivers very large distances, many of which are affected by significant biases. The second digit equal to 0 corresponds to median posterior distance smaller than 20 kpc.

- *Unreliable extinction flag*: it is set to 0 to exclude stars with significantly negative extinctions, or A_V values close to the prior boundary at $A_V = 4$.

- *Large A_V uncertainty flag*: it is set to 0 to exclude stars with very large uncertainty on A_V, probably due to uncertain or incomplete input data.

- *Very small uncertainty flag*: it indicates that the posterior uncertainty on distance, or extinction, or temperature, or $\log g$, or mass is small and most likely underestimated. By setting the fifth digit to 0, stars with unrealistic uncertainties are excluded.

In the following, we use the `StarHorse` parameter `dist50`, which corresponds to the median of the distance posterior distribution, as the distance to a star, and we call it d for brevity. The `StarHorse` distances are derived by taking into account different parallax offsets, which depend on the source G-magnitude, and by re-scaling the parallax errors following a slightly modified version of the re-calibration proposed in Lindegren et al. (2018).
Similarly to Zari et al. (2018), we select Upper Main Sequence (UMS) stars by using the following conditions:

$$M_{G,0} \leqslant 0 \,\text{mag};$$
$$(G_{\mathrm{BP}} - G_{\mathrm{RP}})_0 \leqslant 0.5 \,\text{mag}, \tag{5.1}$$

where M_G is the absolute magnitude in the *Gaia G* band, and $(G_{\mathrm{BP}} - G_{\mathrm{RP}})$ is the colour in the *Gaia* G_{BP} and G_{RP} bands. The de-reddened absolute magnitude and colour $M_{G,0} = M_G - A_G$ and $(G_{\mathrm{BP}} - G_{\mathrm{RP}})_0 = G_{\mathrm{BP}} - G_{\mathrm{RP}} - E(BP - RP)$ are provided in the `StarHorse` catalogue. [1] The sample consists of 17408 sources. By cross-matching with Simbad we find that ~80% of them has a measured spectral type. Most of the sources are B-type stars (7684, i.e. 44%), 5993 (34%) stars are early A-type stars, and only 40 sources are O-type stars.
To select UMS candidate runaway stars we rely primarily on their tangential velocity. To trace back their orbits in 3D space radial velocities are however needed. We therefore cross-match our sample with external radial velocity catalogues, in particular: the XHIP catalogue (3088 sources, Anderson & Francis 2012, and references therein) and RAVE DR5 (254 sources, Kunder et al. 2017).

5.3   Method

We first select as candidate runways stars those whose tangential velocity is significantly different from the rest of the population (Section 5.3.1). At this stage we do not consider radial velocities because the majority of the stars selected in Section 5.2 does not have a measured radial velocity. This however results in a loss of completeness, as we neglect sources with high radial velocity but tangential velocity comparable to the

---

[1] The column names are: `mg0` and `bprp0` respectively.
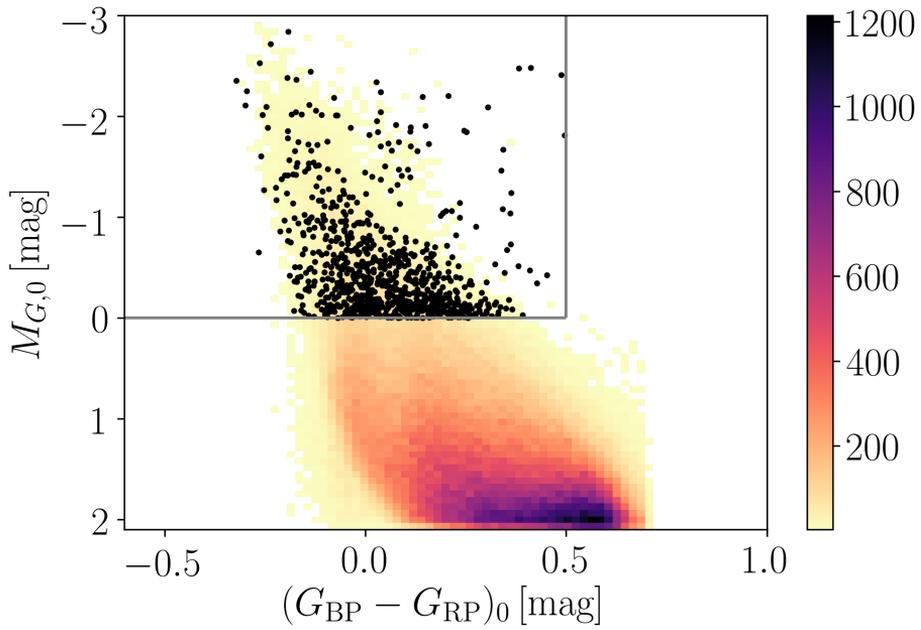
Figure 5.1: Colour-magnitude diagram corrected for extinction and reddening of the sources selected by applying the conditions `SH_OUTFLAG = "00000"` and `SH_GAIAFLAG = "000"`. They grey lines correspond to $M_{\mathrm{G},0} = 0\,\mathrm{mag}$ and $G_{\mathrm{BP}} - G_{\mathrm{RP}} = 0\,\mathrm{mag}$. The black dots represent the candidate runaway stars selected in Section 3.1.
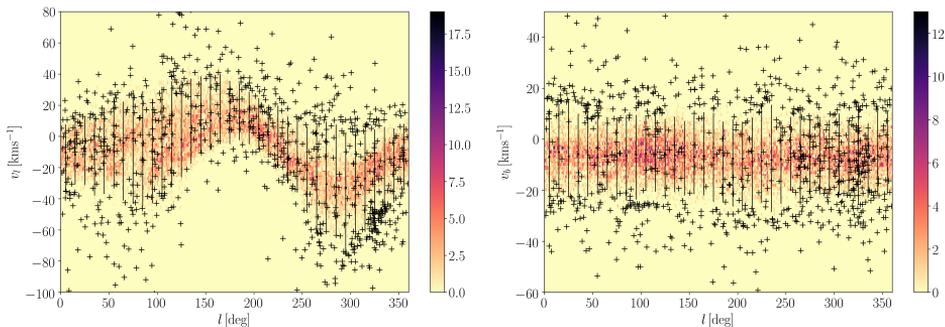
Figure 5.2: The 2D histogram represent the tangential velocity along longitude (v_l, left) and latitude (v_b, right) as a function of longitude for the UMS sample. The black crosses represent the candidate runaway stars selected in Sec. 5.3.1. The vertical bars represent the 16th and 84th percentiles of the v_l and v_b distributions at each longitude bin.

reflex solar motion. Therefore, in Section 5.3.2 we consider the sources with measured radial velocity, we estimate the peculiar total velocity v_{tot}, and we include as candidate runaway stars the sources with total peculiar velocity $v_{tot} > 30\,\mathrm{km\,s^{-1}}$ which were not already selected in Section 5.3.1. In this way we obtain 1197 candidate runaway stars, of which 385 are classified as O- and B- type stars (although there are only 4 O-type stars), and 223 do not have any spectral type from literature. The rest of the sources is of spectral type A or later. We will focus the rest of our analysis (in particular the orbit integration) on the O and B-type stars with measured radial velocities. We include the sources with spectral type A, the sources without a spectral type from literature, and the sources without radial velocities (of any spectral type) in the catalogue which will accompany the paper, but we will not analyse those sources further in this study. Finally, to understand the origin of our runaway candidates, and to determine whether their ejection location was within 1 kpc from the Sun, we perform a three-dimensional (3D) trace back in a galactic potential by integrating our candidate orbits back in time (5.3.3).

5.3.1 Selection of sources with high tangential velocity

Fig. 5.2 shows the tangential velocity in the longitude and latitude direction (v_l and v_b) vs. longitude for the stars selected in Section 2; v_l and v_b are defined as:

$$v_l = A\,\mu_{l*}/\varpi,$$
$$v_b = A\,\mu_b/\varpi, \tag{5.2}$$

where $A = 4.74047\mathrm{km\,s^{-1}\,yr}$. To select our candidates, we define the quantity:

$$\Delta = \sqrt{\left(\frac{v_l - \bar{v}_l}{\sigma_l}\right)^2 + \left(\frac{v_b - \bar{v}_b}{\sigma_b}\right)^2} \tag{5.3}$$

and

$$\sigma_{l,b} = \frac{v_{l,b,16} - v_{l,b,84}}{2}, \tag{5.4}$$

where $\bar{v}_{l,b}$, $v_{l,b,16}$, and $v_{l,b,84}$ are the 50th, 16th and 84th percentiles of the tangential velocities per longitude bin. The quantities $v_{l,pec} = v_l - \bar{v}_l$ and $v_{b,pec} = v_b - \bar{v}_b$ correspond to the peculiar velocity along l and b. We select stars with $\Delta > 3.$, which corresponds roughly to the 95th percentile of the Δ distribution. With this criterion we select stars with tangential velocities strongly deviating from the mean tangential motion as a function of longitude angle. The mean tangential velocity mainly reflects the projection of the solar motion in the different directions on the sky and also to differential galactic rotation. The sample consist of 857 sources, which corresponds to the 5% of the sources selected in Section 5.2. Fig. 5.3 (left) shows the reduced tangential velocity distribution of the stars selected through the Δ condition (orange histogram) and of all the stars in our sample (grey histogram). Both histograms are normalised. The minimum reduced tangential velocity is $\approx 17\,\mathrm{km\,s}^{-1}$, while the peak of the distribution is at $v_t \approx 41\,\mathrm{km\,s}^{-1}$.

5.3.2 Selection of sources with high total velocity

25% of the sources selected in Section 5.2 have literature radial velocities, and 300 of those are selected as runaway stars (35%) only based on their tangential velocity. Fig. 5.4 shows the radial velocity v_r as a function of l for our UMS sample (gray dots), and for our candidate runaway stars (blue crosses). The orange dots correspond to the median radial velocity per latitude bin. Similarly to Fig. 5.2, the sinusoidal trend visible in Fig. 5.4 reflects the solar motion. Many sources that have a high radial velocity compared to the solar motion are not selected as candidate runaway stars by considering the tangential velocity only. Following Hoogerwerf et al. (2001) we therefore select stars whose total peculiar velocity is larger than $30\,\mathrm{km\,s}^{-1}$, where the radial peculiar velocity is defined as $v_{r,pec} = v_r - \bar{v}_r$, and $\bar{v}_r$ is the mean radial velocity per latitude bin and the total peculiar velocity is: $v_{tot,pec} = \sqrt{v_{r,pec}^2 + v_{l,pec}^2 + v_{b,pec}^2}$. In this way we add 341 sources to those selected by using only the proper motions. This brings the total number of runaway candidates to 1197 sources. Fig. 5.3 shows the total peculiar velocity distribution of our sample (light grey histogram), and the total peculiar velocity distribution of the runaway star candidates (blue histogram).

Fig. 5.5 shows the sky distribution of UMS sources selected by applying Eq. 5.1 (grey dots, top) and the distribution of candidate runaway stars (black dots, bottom). UMS stars are located in the galactic plane, and their dispersion along the Z-axis (where Z is the third component of the vector (X, Y, Z), indicating the position of a star in Cartesian galactic coordinates) is around 200 pc.

5.3.3 3D trace back

To derive the birthplace of the candidate runaway stars, we perform an orbit integration using the python package `gala` (Price-Whelan 2017). We use the `gala` *Milky Way Potential*, which consist of a spherical nucleus and bulge, a Miyamoto-Nagai disk, and a spherical NFW dark matter halo. We define the Sun's velocity (in Galactic coordinates) as $(U, V, W)_\odot = (11.1, 12.24, 7.25)\,\mathrm{km\,s}^{-1}$ following Schönrich (2012). The circular rotation velocity at the Sun position is $V_{circ} = 238\,\mathrm{km\,s}^{-1}$, and the Sun's height with respect to the Galactic plane is 25 pc (Bland-Hawthorn & Gerhard 2016).
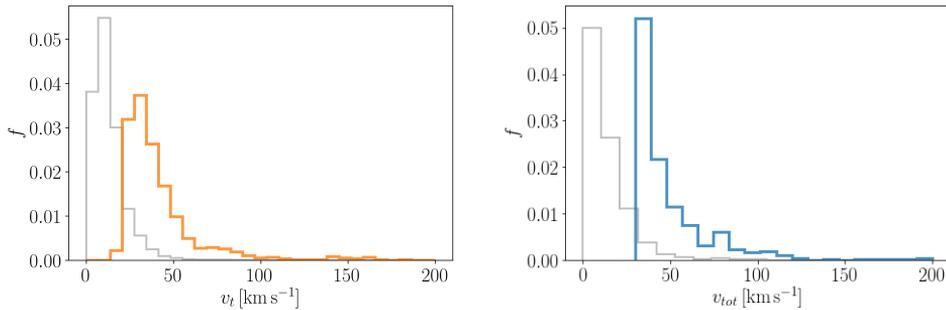
Figure 5.3: Left: Normalised histograms of the reduced total tangential velocity for all the stars selected (grey histogram) and for the star following the condition $\Delta > 3$ (orange histogram). The minimum of the candidate runaway tangential velocity distribution is around $17\,\mathrm{km\,s^{-1}}$, the maximum is around $164\,\mathrm{km\,s^{-1}}$ and the mean $\sim 41\,\mathrm{km\,s^{-1}}$. Right:Normalised histograms of the reduced total velocity for all the stars with radial velocity (lgrey histogram), and for the sources with reduced total velocity higher than $30\;\mathrm{km\,s^{-1}}$ (blue histogram).
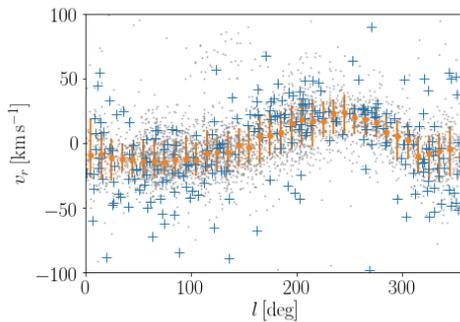
Figure 5.4: Radial velocity versus galactic longitude for the stars selected by applying the conditions `SH_OUTFLAG = "00000"` and `SH_GAIAFLAG = "000"`. The blue crosses represent the candidate runaway stars selected in Section 3.1. The orange dots correspond to the median radial velocity values per latitude bin, and the error bars the 16th and 84th percentile respectively.

Figure 5.5: Top: distribution in the sky of the sources selected with Eq. 1. Bottom: distribution of the candidate runaway stars selected in Section 3.1 and 3.2. The distribution of UMS sources peaks towards the Galactic Plane, as well as the distribution of runaway candidates.

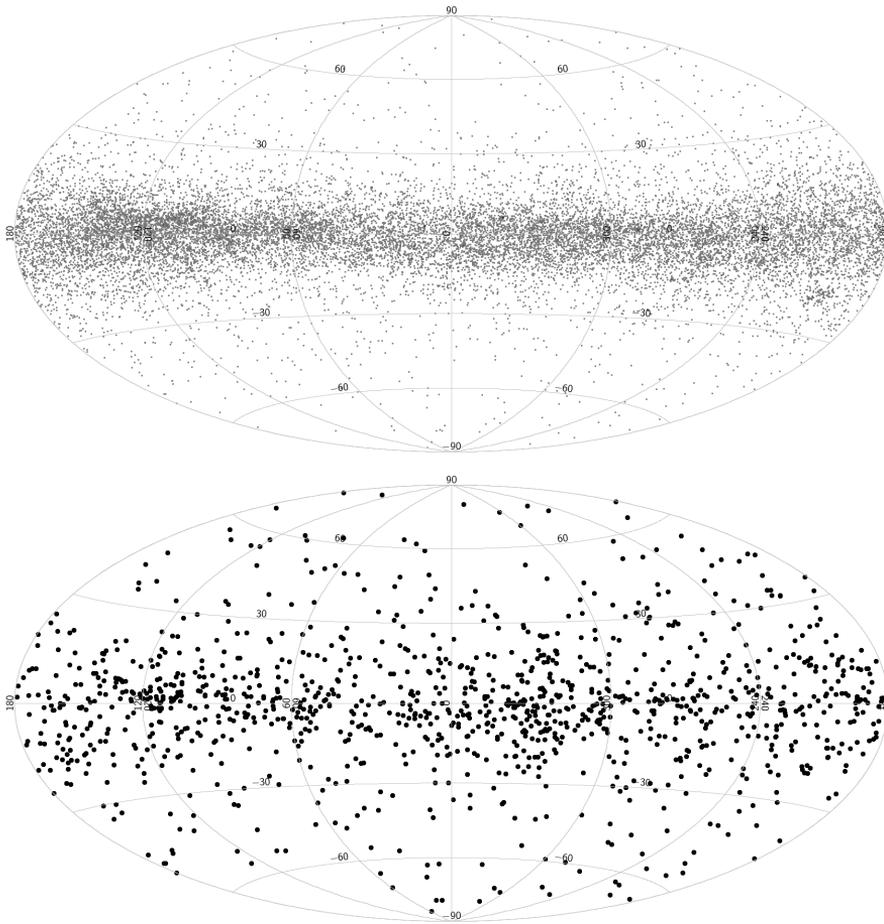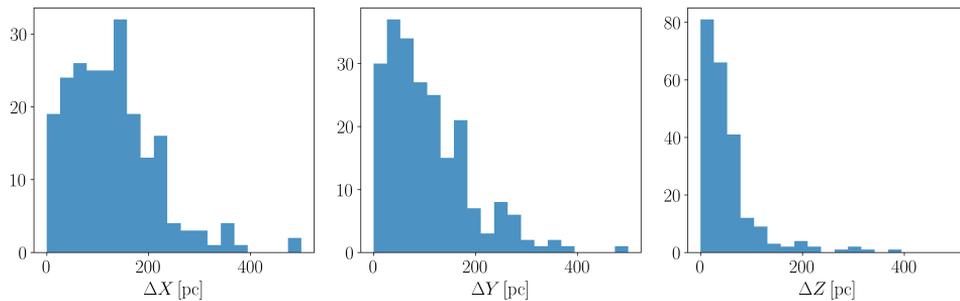**Figure 5.6:** Histograms showing the quantities $(X, Y, Z)_{50} - (X, Y, Z)_{16}$ (left, centre, right respectively) at $t = -10$ Myr. Th distributions indicate the uncertainty in the trace-back of the candidate runaway stars. The median of the distributions are: $\approx 123, 89$, and 38 pc respectively.

The trace back consists of two steps. We first perform a trace back with a long integration time ($t = 100$Myr) and with a large time step $\delta t = 0.5$ Myr to estimate which candidate runaway stars originate from the solar neighbourhood ($d < 1$ kpc). A star with a velocity of $30\,\mathrm{km\,s^{-1}}$ travels ~ 1 kpc in 33 Myr, however we chose a longer trace-back time to account for the possibility of the ejection taking place at distances much farther than 1 kpc. For stars with $|Z_0| > 0.2$ kpc (the suffix 0 indicates the current position of the star), we compute the "ejection distance" as the distance at the time when the star enters the disc, which we define as the slab with $|Z| < 0.2$ kpc. The value $|Z| = 0.2$ kpc corresponds to the dispersion along the Z axis of the stars selected in Section 5.2. For stars with $|Z_0| < 0.2$ kpc, i.e. stars that are already in the disc, we define the ejection distance as the distance when the star first crosses the plane $Z = 0$ kpc.

We repeat the trace back for a shorter time and a smaller time step ($t = 10$ Myr and $\delta t = 0.1$ Myr) to precisely derive the birth place of the stars ejected within $d < 1$ kpc. The total trace back time is arbitrary, and it corresponds to an average age of the star forming regions in the solar neighbourhood. Many of our sources can live for much longer ages that 10 Myr [2], therefore they might come from a very different location than the one derived with a 10 Myr trace back. Another issue is that many of our sources are not traced back to star forming regions within 10 Myr. These problems are further discussed in Section 5.5.

To take into account errors on the measured quantities, we draw $N = 1000$ Monte Carlo (MC) realisations of each stars distance, proper motions, and radial velocity. Fig. 5.6 shows the distribution of the $(X, Y, Z)_{50} - (X, Y, Z)_{16}$ at $t = -10$ Myr. The distribution $(X, Y, Z)_{84} - (X, Y, Z)_{50}$ is similar. The typical error on the final position of each star after a trace back time of 10 Myr are large, with medians of $123, 89$, and 38 pc in X, Y, and Z respectively.

---

[2]For example, a B0.5V type star main sequence lifetime is around 11 Myr, a B5V type star main sequence lifetime is around 220 Myr, and a B9V type star main sequence lifetime is 700-800 Myr

5.4 Results

Fig. 5.7 shows the trajectories of 225 OB-type stars with measured radial velocity, colour-coded by the distance from the Sun at each trace back time. The total integration time is 10 Myr, however in the figure only the first 2 Myr are shown to avoid crowding. The current source positions are marked with a black cross. Of the 225 O and B-type candidate runaway stars with measured radial velocities, 115 (51 % of the sample) have distances smaller than 1 kpc when they enter or cross the plane for more than the 50% of the Monte Carlo simulations. The sources with ejection locations further than 1 kpc might have been ejected from clusters or star forming regions outside the solar neighbourhood.

We now focus on the candidates in two regions, Orion and Scorpius-Centaurus. The runaway star candidates in the Orion region are listed in Table 5.1. Their trajectory is shown in Fig. 5.8. The first two entries in the table (53 Ari and HD 43112) were recognised as runaway by Blaauw (1961) and Hoogerwerf et al. (2001). The origin of 53 Ari is not certain. It could have been ejected by one of the clusters in the Orion region (see Zari et al. 2019) or by a supernova explosion. HD 43112 (HIP 29678) is a B1V star likely ejected from the λ Ori cluster, together with HIP 22061 (which is not in our sample, see Table 5.3) as found by **?**. The other eight candidates in Table 5.1 have distances consistent with Orion, however further analysis is required to determine whether they have been ejected as a result of a supernova explosion or dynamical interactions within a cluster in the region. The trajectories of *Gaia* DR2 301...392 and *Gaia* DR2 320...656 seem to cross at $l, b \approx (195°, -8°)$ (see Fig. 5.8), however the distance from the Sun at which the orbits cross is $< 100\,\mathrm{pc}$, and, most importantly, the closest approach between the orbits occurs at different trace back times (between 7 and 8 Myr and between 6 and 7 Myr). This is therefore likely to be a chance alignment between the orbits projected in the sky, and the stars might have been ejected in the surrounding of the Orion nebula cluster or in the Belt region. The known Orion runaways AE Aur and μ Col are not in our candidate runaway list. AE Aur is not in the `StarHorse` catalogue (see Section 2), while μ Col is in `StarHorse` but it is removed by the conditions on the flags (in particular the *large distance flag* $= 1$, see Table 5.3). Since both of them are in the *Gaia* catalogue however, we confirmed that they are indeed coming from the Orion Nebula Cluster.

In Sco-Cen we noticed that *Gaia* DR2 349...472 and *Gaia* DR2 589...640 seem to come from the same point in the sky (their properties are reported in Table 5.2). By performing Monte Carlo simulations we found that both stars originate from Upper Centaurus Lupus (a sub-group of Sco-Cen, gray box in Fig. 5.9). The distances of the stars are consistent with the distance to the association (de Zeeuw et al. 1999), and the travel times are comparable (colour map in Fig. 5.9, right).

The known runaway ζ Oph (**?**Hoogerwerf et al. 2001), which was found to be ejected from a supernova explosion in the Upper Scorpius (US) sub-group of Sco-Cen, is not included in the `StarHorse` catalogue. Similarly to μ Col and AE Aur, we confirmed that it can be traced back to US by using *Gaia* DR2 astrometry.

The analysis of the orbits of the other runaway star candidates is currently on-going.

Table 5.1: *Gaia* `source_id`, Simbad identifiers, spectral types, observed radial velocities, and total peculiar velocities of the runaway sources whose trajectories cross the Orion region.

Gaia `source_id`	Name	Spectral type	$v_r\,[\mathrm{km\,s^{-1}}]$	$v_{tot}\,[\mathrm{km\,s^{-1}}]$
58850711337487616	53 Ari	B1.5V	21.7 ± 1.5	49.2
3344576352924411136	HD 43112	B1V	36.9 ± 0.7	71.7
3332953896541347200	HD 36824	B3V	17 ± 3.5	23.3
3015379032512681856	HD 37492	B8III	52 ± 1.5	36.9
3017367396223983616	HD 37061	O9V	67 ± 1	50.2
3223150304544227072	HD 38528	B9/9.5IV	65 ± 16.9	47.7
3012264940704649984	HD 38185	B8Ib/II	94 ± 6.9	78.1
3016120962343387392	HD 37889	B3II/III	52.6 ± 6	35.5
3209067866991398656	HD 36120	B8V	69.9 ± 4	51.8
3016713083716619520	HD 36487	B6IV	183.5 ± 39.4	166.7

Table 5.2: *Gaia* `source_id`, Simbad identifiers, spectral types, observed radial velocities, and total peculiar velocities of the runaway sources in the Sco-Cen region.

Gaia `source_id`	Name	Spectral type	$v_r\,[\mathrm{km\,s^{-1}}]$	$v_{tot}\,[\mathrm{km\,s^{-1}}]$
3498480561739049472	HD 111226	B8V	49 ± 7.4	65.6
5895765142704352640	V* V716 Cen	B5V	66 ± 10	76
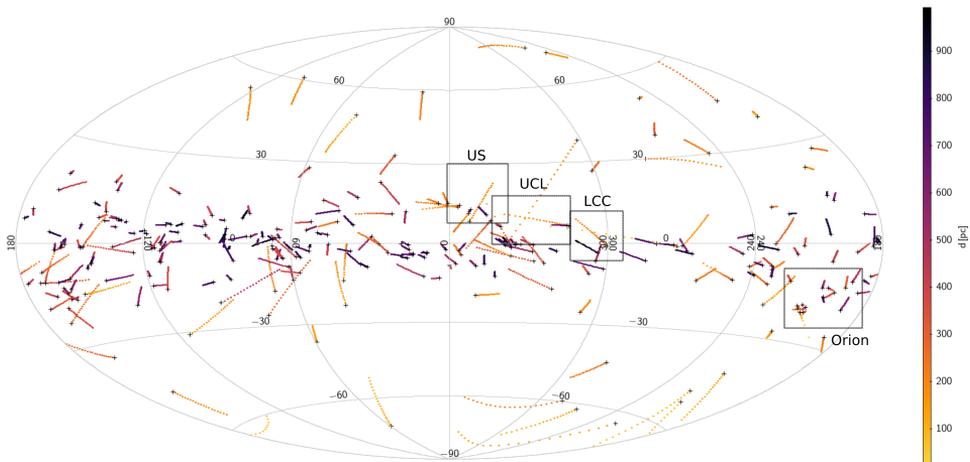
Figure 5.7: Trajectories of the OB-type runaway star candidates with measured radial velocities projected in the sky. The orbits are calculated for 2 Myr. The crosses represent the current position of the sources. The colour map represents the distance of the sources at each time step. The grey boxes represent the approximate locations of the regions analysed in more detail in Section 5.4. Upper Scorpius (US), Upper Centaurus Lupus (UCL), and Lower Centaurus Crux (LCC) are the sub-groups of the Scorpius-Centaurus association (Sco-Cen).
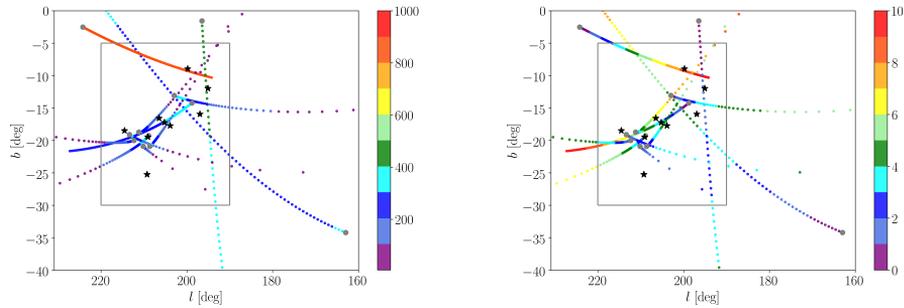
Figure 5.8: Trajectories of the OB-type runaway star candidates in the Orion region, colour-coded according to their distance (in parsec, left) and trace-back time (in Myr, right). The current source positions are shown by the gray dots. The black stars show the Orion constellation, and the grey box draws the boundaries of the complex (see Zari et al. 2017, 2019).

Figure 5.9: Same as Fig. 5.8 but for the two candidate runaway stars in Upper Centaurus Lupus. The background image shows dust extinction from the Planck satellite. The box corresponds to the boundaries of the Upper Centaurus Lupus region.

5.5 Discussion

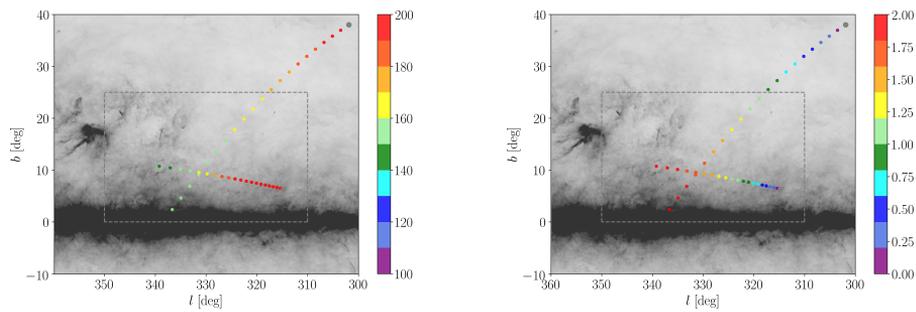In this Section we compare our runaway star candidates with other catalogues, in particular Hoogerwerf et al. (2001), Maíz Apellániz & Weiler (2018), and Tetzlaff et al. (2011), and with the results of the simulations by Renzo et al. (2019b) and Schoettler et al. (2019). Finally we discuss the completeness of our catalogue.

5.5.1 Comparisons with other runaway stars catalogues

Out of the 56 stars in Hoogerwerf et al. (2001), 55 are in the *Gaia* catalogue; 31 sources are left when cross-matching with the `StarHorse` catalogue, cleaned with the conditions `sh_outflag = "00000"` and `sh_gaiaflag = "000"`; 29 sources are left when applying the conditions in Eq. 5.1. Finally 20 stars are classified as runaways using the conditions $\Delta > 3$ or $v_{tot} > 30\,\mathrm{km\,s^{-1}}$. A summary of the comparison between our catalogue and Hoogerwerf et al. (2001) runaway star list is given in Table 5.3. As mentioned in Section 5.4, some of the known runaways are not included in the `StarHorse` catalogue (likely because of convergence problems), or are removed by one of the flags (see Section 5.2) or by the photometric criteria of Eq. 5.1.

We do not have any of the Maíz Apellániz & Weiler (2018) sources in our selection, mainly because of distance: indeed only 10 of their sources are within 1 kpc, 7 remain after we apply the StarHorse quality flags, and only 1 is left after applying Eq. 5.1. This source is not selected by the condition $\Delta > 3$. Some of the sources selected as runaway star candidates by Maíz Apellániz & Weiler (2018) show evidence of bow-shocks, which we do not find for any of the sources in our sample. This could depend on the fact that most of our stars are of spectral type B, and they might not have winds that are strong enough to produce bow shocks or that the medium in which they are located is not dense enough for the bow shock to be created: this is the case for stars at high galactic latitudes.

Tetzlaff et al. (2011) provide a catalogue of 2547 runaway stars of all spectral types (of which 835 are classified as O and B-type stars); 83 of our candidate runaway stars are in common with this sub-set. Tetzlaff et al. (2011) select runaway stars by using their complete kinematics when available, and tangential velocities only for sources without measured radial velocities.The large discrepancy in numbers is due to the same reasons why some of the runaway stars in Hoogerwerf et al. (2001) are not in our sample: some of them are not reported in the `StarHorse` catalogue, others do not pass the quality flags, others are left out as they do not comply with Eq. 5.1.

5.5.2 Comparison with simulations

In their simulations, Renzo et al. (2019b) study the evolution of massive binary systems to predict the peculiar velocities that stars obtain when their companion collapses and disrupts the system. In particular they investigate which physical processes leave a clear imprint and may therefore be constrained observationally. On of their main results is that the fraction of O-type runaway stars is at best of a few percent: this is however in tension with the observational result that 10-20% of the O-type stars are runaways. Our initial sample contains only 40 confirmed O-type stars, and

only four of them are selected as candidate runaways, therefore we can hardly draw any conclusion, without any further analysis of our candidates.

By performing N-body simulations of young stellar clusters, which do not include stellar evolution or primordial binaries, Schoettler et al. (2019) suggest that dynamical interactions during the early evolution of the clusters can produce runaway and walkaway stars. However, the runaways stars ejected in these simulations do not exceed masses of $0.5 M_\odot$. This result, together with the fact that the classical runaway production mechanisms may include also low-mass stars, imply that runaway and walkaway stars should be searched among all spectral types, and not only early type stars. For such studies however, it is necessary to focus on single star forming regions and their neighbourhood, and most importantly to be able to determine accurate ages for large samples.

5.5.3 Completeness

The first cut that might limit the completeness of our sample is the fact that we select stars nominally closer than 1 kpc ($\varpi > 1\,\mathrm{mas}$). By doing so, we do not consider the error on the parallax measurement, therefore we might be excluding stars with a measured parallax smaller than 1 mas, that are however compatible with being closer than 1 kpc.

As mentioned above, the `StarHorse` catalogue does not contain all the sources in *Gaia* DR2, but those that have converged. This is one of the main reasons for the very low number of sources in common with Tetzlaff et al. (2011). We also lose sources when selecting our sample by applying the *Gaia* and `StarHorse` quality flags. For instance, we might be loosing very bright sources by applying the condition $RUWE < 1.4$ (see Section 2).

In Section 5.3.1 we noticed that many sources with total peculiar velocity higher than $30\,\mathrm{km\,s^{-1}}$ were not selected as candidate runaway stars based only on their peculiar tangential velocity. Literature radial velocities are available only for 4352 stars, of which 340 ($\approx 8\%$) were not classified as runaway candidates based on their tangential velocity. If we assume this fraction to stay the same when considering the entire sample, we would be missing around a thousand runaway candidates from our selection (the total number of sources $\times$ 0.08 - 340). By adding these stars to our list, the fraction of runaways in our sample would be around 12-13%.

5.6 Conclusions

We use the `StarHorse` catalogue to search for early-type runaway stars in *Gaia* DR2. We identify O-, B-, and early A-type stars within 1 kpc from the Sun by performing a photometric selection in the $M_{G,0}$ vs. $(G_{\mathrm{BP}} - G_{\mathrm{RP}})_0$ colour-magnitude diagram, after correcting for extinction and reddening. The selection of candidate runaway stars is performed in two steps. We study the tangential velocity distribution of the early type stars, and we select as candidate runaway sources those whose tangential velocity is significantly different than the average tangential velocity at the same longitude angle. After cross-matching with radial velocity catalogues, we include in the candidate runaway source list also the stars whose total peculiar velocity is larger than

$30 \, \mathrm{km \, s^{-1}}$. This is because stars with high radial velocity but tangential velocities comparable with those of field stars are not selected by using only tangential velocities. We integrate back in time the orbits of our candidate stars and we find that all of them are coming from the disc, although around half of our sample was probably originated at distances larger than 1 kpc. We study in more detail the runaway candidates in the Orion and Scorpius-Centaurus star forming regions, leaving to further studies a detailed analysis of all our candidate runaway stars. We compare our findings with previous studies, in particular Hoogerwerf et al. (2001), Maíz Apellániz & Weiler (2018), and Tetzlaff et al. (2011), and we discuss the (in)completeness of our sample. Future surveys (such as SDSS-V, WEAVE, 4MOST, and *Gaia* DR3) and targeted proposals for the brightest stars will increase the number of radial velocities available for the community, and will greatly improve the completeness of our sample.

5.A   Hoogerwerf et al. (2001) runaways in *Gaia* and `StarHorse`

Table 5.3: Hoogerwerf et al. (2001) Table 1 sources in *Gaia* DR2; ζ Puppis is not included because it is not in the *Gaia* DR2 catalogue. The ✓ symbol indicates that a condtion is satisfied. The ✗ symbol signifies that a condition is not satisfied, and therefore the star is excluded in the following columns (- symbol).

Gaia DR2 source_id	HIP ID	$d < 1\,\mathrm{kpc}$	StarHorse	StarHorse flags	$(G_{BP} - G_{RP})_0 \leq 0.5$ & $M_{G,0} \leq 0$	runaway selected
3904419434422220800	HIP 3478	✓	✓	✓	✓	✓
2890122152481960576	HIP 28756	✓	✓	✓	✓	✓
5625488726258364544	HIP 43158	✗	-	-	-	-
3959631234670040704	HIP 61602	✓	✗	-	-	-
4155626054317058048	HIP 91599	✗	-	-	-	-
2083681294654108672	HIP 101350	✓	✓	✓	✓	✓
3750108131390530368	HIP 3881	✓	✓	✓	✓	✗
3344576352924411136	HIP 29678	✓	✓	✓	✓	✓
1132056347324894592	HIP 45563	✗	-	-	-	-
5855802671032914944	HIP 62322	✓	✗	-	-	-
6631708338599353856	HIP 92609	✓	✗	-	-	-
2196984971452315392	HIP 102274	✗	-	-	-	-
4718219005712946432	HIP 9549	✓	✓	✓	✗	-
2891652569588715392	HIP 30143	✓	✓	✗	-	-
5195896888895916288	HIP 46928	✓	✓	✓	✓	✓
6070376179105170048	HIP 66524	✓	✓	✓	✓	✓
6639975875765813888	HIP 94899	✓	✓	✓	✓	✓
6477764130748015488	HIP 103206	✓	✓	✓	✓	✗
4550304862741022272	HIP 10849	✓	✓	✓	✓	✓
3028075299444718080	HIP 35951	✓	✓	✓	✓	✓
5312887640522116992	HIP 46950	✓	✓	✓	✓	-
5895765142704352640	HIP 69491	✓	✓	✓	✓	✓
2049168552364523520	HIP 94934	✓	✓	✓	✓	✓
1963714268134133120	HIP 105811	✗	-	-	-	-
58850711337487616	HIP 14514	✓	✓	✓	✓	✓
5611882304229988864	HIP 36246	✓	✓	✓	✓	✗

HIP	Source ID
HIP 48715	5242303495131015808
HIP 70574	6099307559838681216
HIP 95818	2017735817073815680
HIP 106620	1849597571190605952
HIP 18614	2193759000005174272
HIP 38455	5537733610870346880
HIP 48943	5660602555779590 3616
HIP 76013	5794330865689870592
HIP 96115	2024645246873799040
HIP 109556	2199904449710846720
HIP 20330	5511780021165932 80
HIP 38518	5518235833732209536
HIP 49934	5258611382897321088
HIP 81377	4337352305315545088
HIP 97774	2086465429887466368
HIP 22061	3230156087694893952
HIP 52161	5352159374060797696
HIP 82171	5927377029712024576
HIP 97845	2086460069767734656
HIP 24575	1820715707157130 24
HIP 40341	5547952884262112256
HIP 57669	5333141671190375040
HIP 82868	5936759471698278272
HIP 99435	2249990571599168512
HIP 27204	2901155648586891648
HIP 42038	5321174487856385664
HIP 59607	6151098715037383552
HIP 86768	4168881422863453568
HIP 99580	2062426085622944256

Bibliography

Abolfathi B., et al., 2018, *ApJS*, 235, 42

Almeida L. A., et al., 2017, *A&A*, 598, A84

Alves J., Bouy H., 2012, *A&A*, 547, A97

Alves M. I. R., Boulanger F., Ferrière K., Montier L., 2018, *A&A*, 611, L5

Ambartsumian V. A., 1947, in Stellar Evolution and Astrophysics, Armenian Acad. of Sci. (German translation, 1951, Abhandl. Sowjetischen Astron., 1, 33)

Ambartsumian V. A., 1955, The Observatory, 75, 72

Anders F., et al., 2019, arXiv e-prints, p. arXiv:1904.11302

Anderson E., Francis C., 2012, Astronomy Letters, 38, 331

Andrae R., et al., 2018, *A&A*, 616, A8

Arenou F., et al., 2018, *A&A*, 616, A17

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. p. 459

Bally J., Walawender J., Johnstone D., Kirk H., Goodman A., 2008, The Perseus Cloud. p. 308

Banerjee A., Dhillon I. S., Ghosh J., Sra S., 2005, Journal of Machine Learning Research, 6, 1345

Beccari G., et al., 2017, *A&A*, 604, A22

Beccari G., Boffin H. M. J., Jerabkova T., Wright N. J., Kalari V. M., Carraro G., De Marchi G., de Wit W.-J., 2018, preprint, (`arXiv:1807.07073`)

Bekki K., 2009, *MNRAS*, 398, L36

Belikov A. N., Kharchenko N. V., Piskunov A. E., Schilbach E., Scholz R.-D., Yatsenko A. I., 2002a, *A&A*, 384, 145

Belikov A. N., Kharchenko N. V., Piskunov A. E., Schilbach E., Scholz R.-D., 2002b, *A&A*, 387, 117

Bell C. P. M., Naylor T., Mayne N. J., Jeffries R. D., Littlefair S. P., 2013, *MNRAS*, 434, 806

Bensby T., Feltzing S., Oey M. S., 2014, *A&A*, 562, A71

Blaauw A., 1952, Bulletin of the Astronomical Institutes of the Netherlands, 11, 405

Blaauw A., 1956, *ApJ*, 123, 408

Blaauw A., 1961, Bulletin of the Astronomical Institutes of the Netherlands, 15, 265

Blaauw A., 1964, *ARA&A*, 2, 213

Blaauw A., 1991, in Lada C. J., Kylafis N. D., eds, Vol. 342, NATO Advanced Science Institutes (ASI) Series C. p. 125

Blaauw A., Morgan W. W., 1954, *ApJ*, 119, 625

Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529

Boersma J., 1961, Bulletin of the Astronomical Institutes of the Netherlands, 15, 291

Boubert D., Evans N. W., 2018, *MNRAS*, 477, 5261

Boubert D., Fraser M., Evans N. W., Green D. A., Izzard R. G., 2017, *A&A*, 606, A14

Bouy H., Alves J., 2015, *A&A*, 584, A26

Bouy H., Alves J., Bertin E., Sarro L. M., Barrado D., 2014, *A&A*, 564, A29

Bravi L., et al., 2018, *A&A*, 615, A37

Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127

Briceño C., Preibisch T., Sherry W. H., Mamajek E. A., Mathieu R. D., Walter F. M., Zin-

necker H., 2007a, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 345 (`arXiv:astro-ph/0602446`)

Briceño C., Hartmann L., Hernández J., Calvet N., Vivas A. K., Furesz G., Szentgyorgyi A., 2007b, *ApJ*, 661, 1119

Briceño C., et al., 2019, *AJ*, 157, 85

Briceno C., 2008, The Dispersed Young Population in Orion. p. 838

Brown A. G. A., de Geus E. J., de Zeeuw P. T., 1994, *A&A*, 289, 101

Brown A. G. A., Blaauw A., Hoogerwerf R., de Bruijne J. H. J., de Zeeuw P. T., 1999, in Lada C. J., Kylafis N. D., eds, NATO Advanced Science Institutes (ASI) Series C Vol. 540, NATO Advanced Science Institutes (ASI) Series C. p. 411 (`arXiv:astro-ph/9902234`)

Cantat-Gaudin T., Mapelli M., Balaguer-Núñez L., Jordi C., Sacco G., Vallenari A., 2018a, preprint, (`arXiv:1808.00573`)

Cantat-Gaudin T., et al., 2018b, *A&A*, 618, A93

Cantat-Gaudin T., et al., 2019, *A&A*, 626, A17

Carraro G., Villanova S., Demarque P., Moni Bidin C., McSwain M. V., 2008, *MNRAS*, 386, 1625

Carrasco J. M., et al., 2016, *A&A*, 595, A7

Chambers K. C., et al., 2016, preprint, (`arXiv:1612.05560`)

Chen W. P., Lee H. T., 2008, The Lacerta OB1 Association. p. 124

Chen Y., Girardi L., Bressan A., Marigo P., Barbieri M., Kong X., 2014, *MNRAS*, 444, 2525

Chen Y., Bressan A., Girardi L., Marigo P., Kong X., Lanza A., 2015, *MNRAS*, 452, 1068

Chini R., Hoffmeister V. H., Nasseri A., Stahl O., Zinnecker H., 2012, *MNRAS*, 424, 1925

Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *ApJ*, 823, 102

Clark P. C., Bonnell I. A., Zinnecker H., Bate M. R., 2005, *MNRAS*, 359, 809

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., Soderblom D. R., Panagia N., Hillenbrand L. A., Palla F., Stassun K. G., 2010, *ApJ*, 722, 1092

Da Rio N., Robberto M., Hillenbrand L. A., Henning T., Stassun K. G., 2012, *ApJ*, 748, 14

Da Rio N., Tan J. C., Jaehnig K., 2014, *ApJ*, 795, 55

Da Rio N., et al., 2016, *ApJ*, 818, 59

Dame T. M., 1993, in Holt S. S., Verter F., eds, American Institute of Physics Conference Series Vol. 278, Back to the Galaxy. pp 267–278, doi:10.1063/1.43985

Damiani F., Prisinzano L., Jeffries R. D., Sacco G. G., Randich S., Micela G., 2017, *A&A*, 602, L1

Damiani F., Prisinzano L., Pillitteri I., Micela G., Sciortino S., 2018, preprint, (`arXiv:1807.11884`)

De Donder E., Vanbeveren D., van Bever J., 1997, *A&A*, 318, 812

Dotter A., 2016, *ApJS*, 222, 8

Dray L. M., Dale J. E., Beer M. E., Napiwotzki R., King A. R., 2005, *MNRAS*, 364, 59

Drimmel R., 2000, *A&A*, 358, L13

Eldridge J. J., Langer N., Tout C. A., 2011, *MNRAS*, 414, 3501

Elias F., Cabrera-Caño J., Alfaro E. J., 2006a, *AJ*, 131, 2700

Elias F., Alfaro E. J., Cabrera-Caño J., 2006b, *AJ*, 132, 1052

Elias F., Alfaro E. J., Cabrera-Caño J., 2009, *MNRAS*, 397, 2

Elmegreen B. G., 1982, *ApJ*, 253, 655

Elmegreen B. G., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. p. 97

Elmegreen B. G., Efremov Y. N., 1998, arXiv e-prints, pp astro–ph/9801071

Evans D. W., et al., 2018, preprint, (`arXiv:1804.09368`)

Fang M., et al., 2017, *AJ*, 153, 188

Finkbeiner D. P., 2003, *ApJS*, 146, 407

Gaia Collaboration et al., 2016a, *A&A*, 595, A1

Gaia Collaboration et al., 2016b, *A&A*, 595, A2

Gaia Collaboration et al., 2018a, *A&A*, 616, A1

Gaia Collaboration et al., 2018b, *A&A*, 616, A10

Getman K. V., Feigelson E. D., Kuhn M. A., 2014, *ApJ*, 787, 109

Gies D. R., Bolton C. T., 1986, *ApJS*, 61, 419

Gould B. A., 1874, Proc. AAAS, p. 115

Großschedl J. E., et al., 2018, *A&A*, 619, A106

Gualandris A., Portegies Zwart S., Eggleton P. P., 2004, *MNRAS*, 350, 615

Guillout P., Sterzik M. F., Schmitt J. H. M. M., Motch C., Neuhaeuser R., 1998, *A&A*, 337, 113

Gutermuth R. A., et al., 2008, *ApJ*, 674, 336

Gvaramadze V. V., Gualandris A., Portegies Zwart S., 2009, *MNRAS*, 396, 570

Hacar A., Alves J., Forbrich J., Meingast S., Kubiak K., Großschedl J., 2016, *A&A*, 589, A80

Hacar A., Tafalla M., Alves J., 2017, *A&A*, 606, A123

Hartmann L., Ballesteros-Paredes J., Bergin E. A., 2001, *ApJ*, 562, 852

Hernández J., et al., 2007, *ApJ*, 662, 1067

Herschel Sir J. F. W., 1847, Results of astronomical observations made during the years 1834, 5, 6, 7, 8, at the Cape of Good Hope; being the completion of a telescopic survey of the whole surface of the visible heavens, commenced in 1825

Hirota T., et al., 2007, PASJ, 59, 897

Hoogerwerf R., Aguilar L. A., 1999, *MNRAS*, 306, 394

Hoogerwerf R., de Bruijne J. H. J., de Zeeuw P. T., 2001, *A&A*, 365, 49

Hunter J. D., 2007, Computing In Science & Engineering, 9, 90

Jeffries R. D., 2007, *MNRAS*, 376, 1109

Jeffries R. D., Maxted P. F. L., Oliveira J. M., Naylor T., 2006, *MNRAS*, 371, L6

Jeffries R. D., Naylor T., Walter F. M., Pozzo M. P., Devey C. R., 2009, *MNRAS*, 393, 538

Jeffries R. D., Littlefair S. P., Naylor T., Mayne N. J., 2011, *MNRAS*, 418, 1948

Jeffries R. D., et al., 2014, *A&A*, 563, A94

Jerabkova T., Beccari G., Boffin H. M. J., Petr-Gotzens M. G., Manara C. F., Prada Moroni P. G., Tognelli E., Degl'Innocenti S., 2019, *A&A*, 627, A57

Jilinski E., Ortega V. G., Drake N. A., de la Reza R., 2010, *ApJ*, 721, 469

Jordi C., et al., 2010, *A&A*, 523, A48

Jørgensen B. R., Lindegren L., 2005, *A&A*, 436, 127

Katz D., et al., 2018, preprint, (arXiv:1804.09372)

Kennicutt R. C., Evans N. J., 2012, *ARA&A*, 50, 531

Kenyon S. J., Dobrzycka D., Hartmann L., 1994, *AJ*, 108, 1872

Kenyon M. J., Jeffries R. D., Naylor T., Oliveira J. M., Maxted P. F. L., 2005, *MNRAS*, 356, 89

Kim M. K., et al., 2008, PASJ, 60, 991

Kos J., et al., 2018, arXiv e-prints,

Kounkel M., Hartmann L., Calvet N., Megeath T., 2017a, *AJ*, 154, 29

Kounkel M., et al., 2017b, *ApJ*, 834, 142

Kounkel M., et al., 2018, *AJ*, 156, 84

Kraus S., et al., 2009, *A&A*, 497, 195

Krause M. G. H., et al., 2018, *A&A*, 619, A120

Kroupa P., Jeřábková T., Dinnbier F., Beccari G., Yan Z., 2018, *A&A*, 612, A74

Kubiak K., et al., 2016, preprint, (arXiv:1609.04948)

Kubiak K., et al., 2017, *A&A*, 598, A124

Kun M., Kiss Z. T., Balog Z., 2008, Star Forming Regions in Cepheus. p. 136

Kunder A., et al., 2017, *AJ*, 153, 75

Kutner M. L., Tucker K. D., Chin G., Thaddeus P., 1977, *ApJ*, 215, 521

Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57

Lallement R., Vergely J.-L., Valette B., Puspitarini L., Eyer L., Casagrande L., 2014, *A&A*, 561, A91

Lallement R., et al., 2018, *A&A*, 616, A132

Leonard P. J. T., 1991, *AJ*, 101, 562

Lindblad P. O., 1967, Bull. Astron. Inst. Netherlands, 19, 34

Lindblad P. O., Palous J., Loden K., Lindegren L., 1997, in Bonnet R. M., et al., eds, ESA Special Publication Vol. 402, Hipparcos - Venice '97. pp 507–512

Lindegren L., Madsen S., Dravins D., 2000, *A&A*, 356, 1119

Lindegren L., et al., 2016, *A&A*, 595, A4

Lindegren L., et al., 2018, *A&A*, 616, A2

Lombardi M., Lada C. J., Alves J., 2017, preprint, (arXiv:1707.08594)

Magnier E. A., et al., 2016, preprint, (arXiv:1612.05242)

Maíz Apellániz J., Weiler M., 2018, *A&A*, 619, A180

Mathieu R. D., 2008, The λ Orionis Star Forming Region. p. 757

McKee C. F., Ostriker E. C., 2007, *ARA&A*, 45, 565

Menten K. M., Reid M. J., Forbrich J., Brunthaler A., 2007, *A&A*, 474, 515

Michalik D., Lindegren L., Hobbs D., 2015, *A&A*, 574, A115

Muench A., Getman K., Hillenbrand L., Preibisch T., 2008, Star Formation in the Orion Nebula I: Stellar Content. p. 483

Myers P. C., 2009, *ApJ*, 700, 1609

Ochsendorf B. B., Cox N. L. J., Krijt S., Salgado F., Berné O., Bernard J. P., Kaper L., Tielens A. G. G. M., 2014, *A&A*, 563, A65

Ochsendorf B. B., Brown A. G. A., Bally J., Tielens A. G. G. M., 2015, *ApJ*, 808, 111

Olano C. A., 1982, *A&A*, 112, 195

Olano C. A., 2001, *AJ*, 121, 295

Olver F. W. J., Lozier D. W., F. B. R., W. C. C., eds, 2010, NIST Handbook of Mathematical Functions. Cambridge University Press

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., et al., 2011, Journal of Machine Learning Research, 12, 2825

Pérez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21

Perryman M. A. C., 1997, in Bonnet R. M., et al., eds, ESA Special Publication Vol. 402, Hipparcos - Venice '97. pp 1–4

Perryman M. A. C., et al., 2001, *A&A*, 369, 339

Planck Collaboration et al., 2014, *A&A*, 571, A11

Poggio E., et al., 2018, *MNRAS*,

Poppel W., 1997, Fund. Cosmic Phys., 18, 1

Portegies Zwart S. F., McMillan S. L. W., Makino J., 2007, *MNRAS*, 374, 95

Poveda A., Ruiz J., Allen C., 1967, Boletin de los Observatorios Tonantzintla y Tacubaya, 4, 86

Preibisch T., Zinnecker H., 2007, in Elmegreen B. G., Palous J., eds, IAU Symposium Vol. 237, Triggered Star Formation in a Turbulent ISM. pp 270–277 (arXiv:astro-ph/0610826), doi:10.1017/S1743921307001597

Price-Jones N., Bovy J., 2019, arXiv e-prints,

Price-Whelan A. M., 2017, The Journal of Open Source Software, 2, 388

Puspitarini L., Lallement R., Vergely J.-L., Snowden S. L., 2014, *A&A*, 566, A13

Reggiani M., Robberto M., Da Rio N., Meyer M. R., Soderblom D. R., Ricci L., 2011, *A&A*, 534, A83

Reino S., de Bruijne J., Zari E., d'Antona F., Ventura P., 2018, *MNRAS*, 477, 3197

Renzo M., et al., 2019a, *MNRAS*, 482, L102

Renzo M., et al., 2019b, *A&A*, 624, A66

Rizzuto A. C., Ireland M. J., Robertson J. G., 2011, *MNRAS*, 416, 3108

Röser S., Schilbach E., Goldman B., Henning T., Moor A., Derekas A., 2018, *A&A*, 614, A81

Roslund C., 1960, *PASP*, 72, 205

Ryu T., Leigh N. W. C., Perna R., 2017, *MNRAS*, 470, 2

Sacco G. G., et al., 2015, *A&A*, 574, L7

Sana H., et al., 2012, Science, 337, 444

Sancisi R., Goss W. M., Anderson C., Johansson L. E. B., Winnberg A., 1974, *A&A*, 35, 445

Sandstrom K. M., Peek J. E. G., Bower G. C., Bolatto A. D., Plambeck R. L., 2007, *ApJ*, 667, 1161

Scelsi L., et al., 2007, *A&A*, 468, 405

Schlafly E. F., et al., 2015, *ApJ*, 799, 116

Schneider N., et al., 2012, *A&A*, 540, L11

Schoettler C., Parker R. J., Arnold B., Grimmett L. P., de Bruijne J., Wright N. J., 2019, arXiv e-prints, p. arXiv:1905.10317

Schönrich R., 2012, *MNRAS*, 427, 274

Schönrich R., Binney J., Dehnen W., 2010, *MNRAS*, 403, 1829

Sherry W. H., Walter F. M., Wolk S. J., 2004, *AJ*, 128, 2316

Sherry W. H., Walter F. M., Wolk S. J., Adams N. R., 2008, *AJ*, 135, 1616

Skrutskie M. F., et al., 2006, *AJ*, 131, 1163

Soler J. D., Bracco A., Pon A., 2018, *A&A*, 609, L3

Stahler S. W., Palla F., 2005, The Formation of Stars

Stassun K. G., Mathieu R. D., Vaz L. P. R., Stroud N., Vrba F. J., 2004, *ApJS*, 151, 357

Tang J., Bressan A., Rosenfield P., Slemer A., Marigo P., Girardi L., Bianchi L., 2014, *MNRAS*, 445, 4287

Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, Astronomical Society of the Pacific Conference Series Vol. 347, Astronomical Data Analysis Software and Systems XIV. p. 29

Tetzlaff N., Neuhäuser R., Hohle M. M., 2011, *MNRAS*, 410, 190

Torra J., Gomez A. E., Figueras F., Comeron F., Grenier S., Mennessier M. O., Mestres M., Fernand ez D., 1997, in Bonnet R. M., et al., eds, ESA Special Publication Vol. 402, Hipparcos - Venice '97. pp 513–518

Valls-Gabaud D., 2014, in EAS Publications Series. pp 225–265 (`arXiv:1607.03000`), doi:10.1051/eas/1465006

Vanderplas J., Connolly A., Ivezić Ž., Gray A., 2012, in Conference on Intelligent Data Understanding (CIDU). pp 47 –54, doi:10.1109/CIDU.2012.6382200

Villa Vélez J. A., Brown A. G. A., Kenworthy M. A., 2018, Research Notes of the American Astronomical Society, 2, 58

Walter F. M., Sherry W. H., Wolk S. J., Adams N. R., 2008, The σ Orionis Cluster. p. 732

Wright N. J., Mamajek E. E., 2018, *MNRAS*, 476, 381

Wright N. J., Bouy H., Drew J. E., Sarro L. M., Bertin E., Cuilland re J.-C., Barrado D., 2016, *MNRAS*, 460, 2593

Zapatero Osorio M. R., Béjar V. J. S., Pavlenko Y., Rebolo R., Allende Prieto C., Martín E. L.,

García López R. J., 2002, *A&A*, 384, 937

Zari E., Brown A. G. A., de Bruijne J., Manara C. F., de Zeeuw P. T., 2017, *A&A*, 608, A148

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

Zari E., Brown A. G. A., de Zeeuw P. T., 2019, arXiv e-prints, p. arXiv:1906.07002

Zwicky F., 1957, Morphological astronomy

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 Mink S. E., Brott I., Cantiello M., Izzard R. G., Langer N., Sana H., 2012, in Drissen L., Robert C., St-Louis N., Moffat A. F. J., eds, Astronomical Society of the Pacific Conference Series Vol. 465, Proceedings of a Scientific Meeting in Honor of Anthony F. J. Moffat. p. 65

de Wit W. J., Testi L., Palla F., Zinnecker H., 2005, *A&A*, 437, 247

de Zeeuw P. T., Hoogerwerf R., de Bruijne J. H. J., Brown A. G. A., Blaauw A., 1999, *AJ*, 117, 354

van Leeuwen F., et al., 2017, *A&A*, 599, A32

English Summary

The figure of Orion the Hunter is a familiar sight in the winter sky of the Northern hemisphere. The entire area, which is shown in the large panel of Fig. 5.10, is an extraordinarily active site of star formation and has received intense astronomical scrutiny. Orion hosts many young stellar clusters, superimposed along the line of sight and at different evolutionary stages. Explaining the detailed sequence of events (the so-called *star formation history*) that caused the formation of such a numerous population is one of the main topics of this thesis.

Almost all the bright blue stars visible in Fig. 5.10 belong to the so-called "Orion OB association". OB associations were first identified as loose groups of young, massive stars. These large structures, whose physical sizes are of order of hundreds of parsecs[3], are the last stage of the massive star formation process and the context in which new stars are born. For example, the current star formation in the Orion Nebula (see Fig. 5.10, bottom) is linked to the earlier generations of massive stars in the adjacent groups of the Orion association.

Similarly, the formation of single OB associations is related to large scale star formation events, occurring on scales of hundreds of parsecs. In the solar neighbourhood (which is the region within 500 pc from the Sun), such events produced numerous associations, that historically were thought to form a ring-like structure which is usually referred to as the *Gould Belt*. The associations and clusters that compose the Gould Belt are very young compared for example to the Sun, which is around 5 billion years old, and were about to form when dinosaurs became extinct (around 66 Myr ago). The first members of the genus *Homo* (*Homo habilis* and *Homo erectus*) appeared between 4.5 and 2 Myr ago: this is roughly the age of the youngest clusters in the Gould Belt, such as the Orion Nebula Cluster shown in Fig. 5.10, bottom.

Assuming that these archaic humans would look at the stars on a clear night, they would have seen a slightly different sky than what we observe now. Stars indeed move on the sky. They orbit around the centre of our galaxy, the Milky Way, in an orderly fashion, but there are also local velocity patterns that differ from region to region. The members of OB associations were likely born from the same complexes of gas and dust, and thus they not only share the same rotational velocity but also the same local velocity. This property is often used to separate the members of OB associations from other stars.

A fraction of O and B-type stars moves with velocities higher than expected: these are referred to as *runaway* stars. Runaway stars do not acquire their velocity at birth, but during their life. To explain the origin of such high velocities, two mechanisms have been proposed. The first mechanism predicts that a runaway star might originally have been a member of a binary star system consisting of two massive stars. If one of the stars explodes as a supernova, the other is suddenly released from the gravitational attraction and can be launched away at high speed on a new trajectory. The second mechanism predicts that a runaway might have formed in a very dense young cluster, like the Orion Nebula Cluster. In this environment, two binary star systems

---

[3]1 parsec corresponds to approximately 3×10^{16} m, or 3.28 light years

can pass close to each other, and interact gravitationally. Such interaction can disrupt both systems, and one or more stars can be ejected at high speed.

Gaia

Unravelling the structure and star formation history of the young associations requires accurate knowledge of stellar distances, motions, and ages. The data of the ESA *Gaia* spacecraft are crucial in this respect. The main goal of the *Gaia* mission is to make the largest, most precise three-dimensional map of our Galaxy by measuring the positions and motions of more than a billion stars in our Galaxy and beyond. A star's position on the sky changes with time due to its motion relative to the Sun and the motion of the Earth around the Sun. The projection of a star's space motion relative to the Sun onto the celestial sphere is called *proper motion*. This is an angular velocity (angle per time). The corresponding velocity is the *tangential velocity*. When the distance to an object is known, the tangential velocity can be calculated from its angular velocity. If the projection of a star's velocity on the line of sight (the *radial velocity*) is also measured, the *total velocity* can be obtained by adding up the tangential and the radial velocities. The apparent motion of a star due to the rotation of the Earth around the Sun is called *parallactic motion*, and it is related to the distance of the star. The smaller the parallactic motion is, the larger is the distance to the star with respect to the Sun.

Gaia is not the first mission of this kind. In the 90's, the *Hipparcos* satellite produced a catalogue of positions and motions for around one-hundred thousands stars. *Hipparcos* allowed for an extensive census of the stellar content of nearby OB associations. However, the data were not precise enough to determine the three-dimensional structure of even the nearest association, Scorpius-Centaurus, or to make significant progress in Orion. *Gaia* significantly improves on Hipparcos for a number of reasons. For example, *Gaia* measures star positions and motions 200 times more accurately than *Hipparcos*. As a comparison, Gaia's precision is so high that it can measure the angle that corresponds to the diameter of a coin on the Moon, while *Hipparcos* could measure only the height of the astronaut holding it in their hands.

This thesis

This thesis makes use of the first *Gaia* data releases (*Gaia* DR1 and DR2) to obtain a detailed census of the young stellar populations in the solar neighbourhood, exploring the distribution and the properties of both high-mass, early-type stars and low-mass, pre-main sequence stars. Figure 5.11 shows where high-mass stars and pre-main sequence stars are located in a *colour-magnitude diagram* of the Orion region. A colour-magnitude diagram is a tool used to classify stars according to their luminosity, temperature, mass, and evolutionary stage. When a star starts the nuclear fusion of hydrogen in its central regions, it enters the main sequence (dashed line in Fig. 5.11). The position of a star on the main sequence depends on its mass: the upper main sequence (grey box in the upper left corner of the plot) is where massive, early type stars are located. Just before stars enter the main sequence, they are located on the pre-main sequence (grey ellipse): in this stage, stars are contracting and their temperature is rising, but hydrogen fusion has yet to start. New born stars enter the pre-main sequence after they have acquired almost all their mass and blown away their envelope of gas and dust. The pre-main sequence stage lasts from a few million years to a few tens of million years depending on the star's mass: this is very short
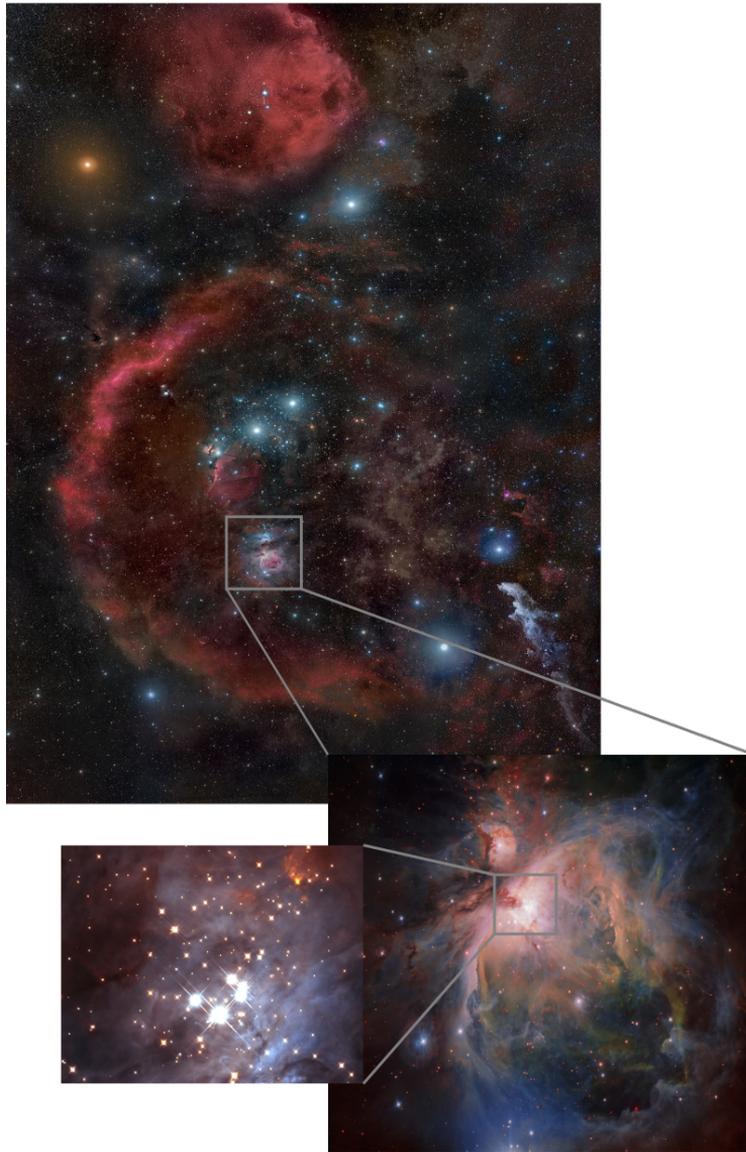
Figure 5.10: Top: stars and gas in the Orion region. The bright red circular filament arcing down from the middle is Barnard's loop (Rogelio Bernal Andreo, *DeepSkyColors.com*). Bottom right: the Orion Nebula, where star formation is currently taking place (ESO/G. Beccari). Bottom left: central part of the Orion Nebula Cluster (K.L. Luhman (Harvard-Smithsonian Center for Astrophysics, Cambridge, Mass.); and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thompson (Steward Observatory, University of Arizona, Tucson, Ariz.) and NASA/ESA).
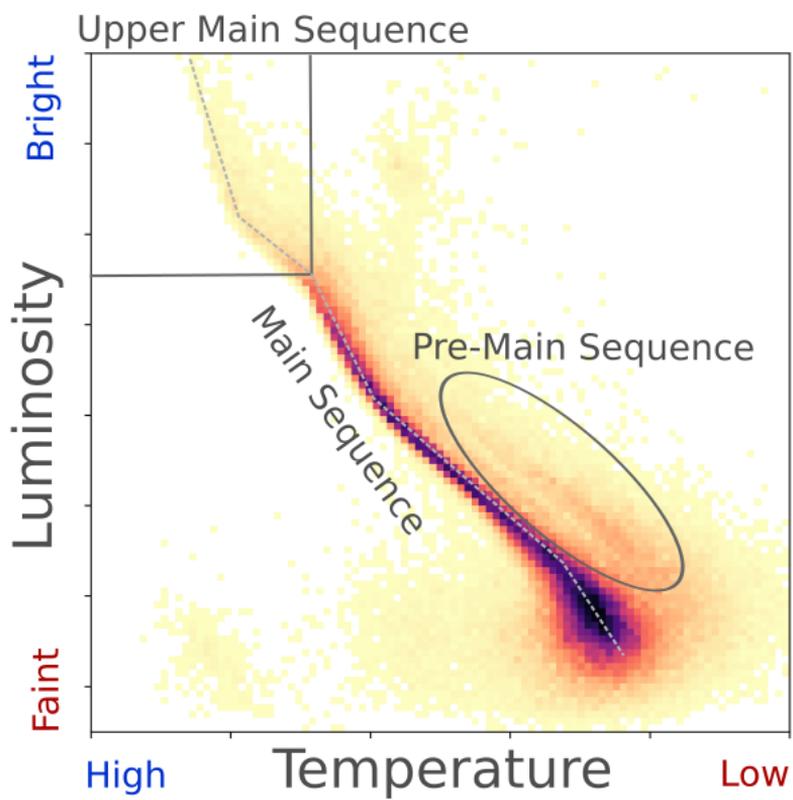
Figure 5.11: Colour-magnitude diagram of the stars in the Orion region. The dashed grey line indicates the main sequence; the grey box is a rough boundary for the upper main sequence; the grey ellipse highlights the pre-main sequence.

compared to the average time stars spend on the main sequence.

The goal of this thesis is to understand how OB associations form and disperse, what are the characteristics of the stellar populations within single associations, what are the properties of the ensemble of OB associations in terms of their disposition in space, and how this compares with what is observed in other galaxies. These topics are addressed by using the Orion OB association as a benchmark to study the mechanisms leading to the formation of an OB association, by studying the disposition of OB associations and star forming regions within 500 pc from the Sun, and finally by analysing the kinematic properties of O and B stars within 1 kpc from the Sun. In particular, the focus is on answering the following questions:

- What are the stellar populations in the Orion OB association?

- What is the star formation history of the Orion OB association?

- What is the structure of the solar neighbourhood as traced by young stars?

- How many runaway stars are there in the solar neighbourhood?

The study presented in **Chapter 2** represents the first step to unravel the complexity of the star formation history of Orion, in terms of the various star formation episodes, their duration, and their effects on the surrounding interstellar medium. The *Gaia* DR1 data provided evidence for the presence of a young stellar population loosely distributed around known clusters. The estimated ages of the members of this population suggested the presence of an age sequence in the association.

These conclusions are partially revisited in **Chapter 3**. The better accuracy of *Gaia* DR2 compared with *Gaia* DR1 and the larger number of sources for which it was possible to determine distances and tangential velocities allow for a thorough study of the three dimensional configuration of the stellar groups composing the Orion OB associations and of their motions and ages. The main finding of this chapter is that the star formation events in Orion follow a complex history, which caused kinematic and physical sub-structure.

The focus of **Chapter 4** is the entire solar neighbourhood. Three-dimensional maps of the spatial distribution of pre-main sequence (PMS) and upper main sequence stars show three prominent structures, Scorpius-Centaurus, Orion, and Vela (see Fig. 5.12). The distribution of the PMS stars as a function of their age shows that younger stars cluster in dense, compact clumps, and are surrounded by older sources, whose distribution is instead more diffuse. Strikingly, the maps do now provide any evidence for the existence of the ring-like structure which is usually referred to as the Gould Belt.

Chapter 5 presents a search for runaway stars within 1 kpc from the Sun. Candidate runaway stars are selected among upper main sequence stars, and classified as such by using their tangential velocity, and, when possible, their total velocity. In particular, candidate runaway stars are defined as sources that have tangential velocities significantly different from the rest of the population or total velocities higher than $30\,\mathrm{km\,s^{-1}}$. The analysis is focused on the runaway star candidates in the Orion and Scorpius-Centaurus (Sco-Cen) regions. In Orion, six new runaway star candidates are added to the sample of previously known runaway stars. In Sco-Cen, two runaway star candidates that likely share the same origin are identified.
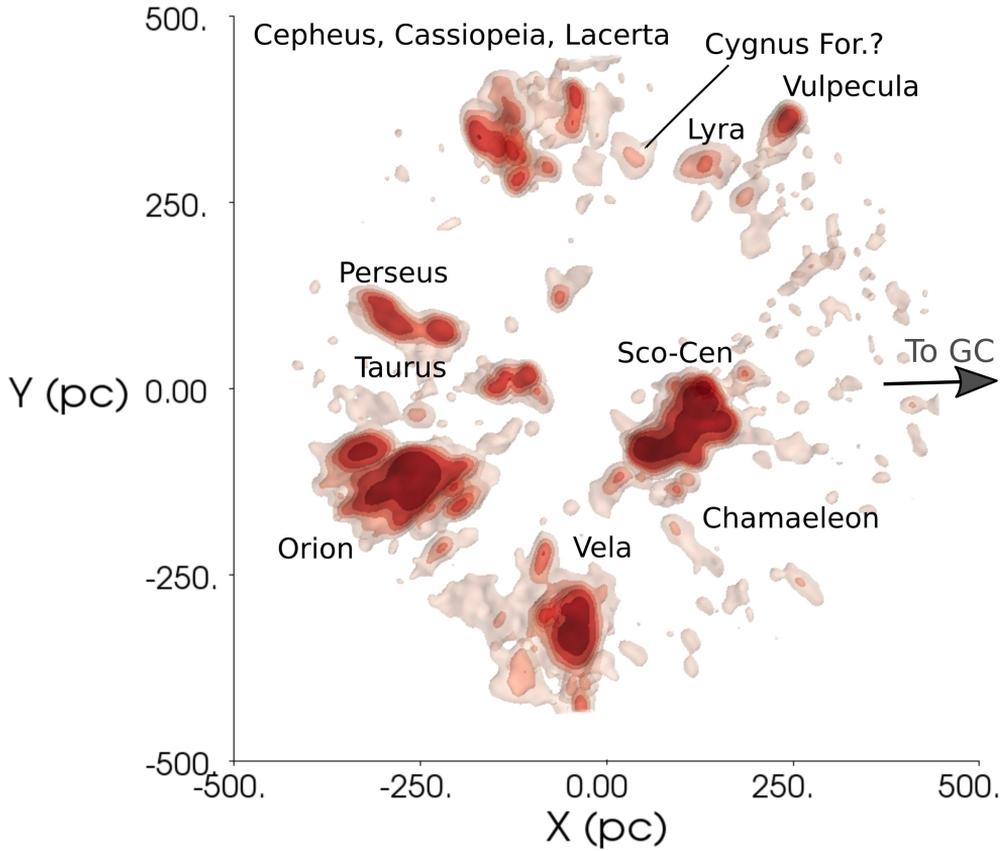
Figure 5.12: Density distribution of pre-main sequence sources younger than 20 Myr in the galactic plane. The Sun is at the centre, in $(X, Y) = (0, 0)$, the x-axis is directed towards the galactic centre (whose direction is indicated by the arrow), and the y-axis towards the direction of the galactic rotation.

Conclusions

The main conclusion of this thesis is that large scale star formation events that lead to the formation of OB associations are complex, and not well understood. There is not a general star formation theory that completely explains the sub-structure (in space, kinematics, and ages) observed in Orion and in other OB associations in the solar neighbourhood. The origin of OB associations remains somewhat mysterious. The structure of the solar neighbourhood is undeniably different than what was thought in the pre-*Gaia* era. These findings call for a revision of the theories of propagation and triggering of star formation. Data from the future releases of the *Gaia* satellite and from upcoming spectroscopic surveys will contribute in exploring in more detail the kinematic and physical sub-structure of large star-forming complexes.

Nederlandse samenvatting

Orion de Jager is een bekend sterrenbeeld dat zichtbaar is in de winter van het noordelijk halfrond. Het complete gebied, welke getoond is in het grote paneel van Fig. 5.13, is een gebied waar extreem veel nieuwe sterren worden gevormd en wordt daarom intensief bestudeerd. Orion bevat veel jonge sterclusters van verschillende leeftijden. Het gedetailleerd bestuderen van de volgorde van ster formatie (de zogenoemde *ster formatie geschiedenis*) welke verantwoordelijk is boor de vorming of zulke aantallen sterpopulaties is een van de belangrijkste onderwerpen van dit proefschrift.

Bijna alle heldere blauwe sterren die zichtbaar zijn in Fig. 5.13 behoren tot het zogenoemde "Orion OB associatie". OB associaties waren voor het eerst geidentificeerd als losse groepen van jonge, massieve sterren. Deze grote structuren, met fysieke afmetingen die overeenkomen met honderden parsecs [4], zijn de laatste fase van het proces waarin massieve sterren worden gevormd. De huidige ster formatie in de Orion wolk (zie 5.13, onder) is gelinkt aan de eerdere generaties van massieve sterren in de aangrenzende groepen in de Orion associatie.

De formatie van enkele OB associaties is gerelateerd aan grotere schaal sterformatie gebeurtenissen, welke op schalen van honderden parsecs plaatsvinden. In de omgeving van de zon, binnen een straal van 500 parsec van de zon, worden tijdens zulke gebeurtenissen grote aantallen associaties gevormd, van welke historisch verwacht waren dat een ringachtige structuur werd gevormd, welke de *Gouden Riem* wordt genoemd. De associaties en clusters die samen de Gouden Riem vormen zijn heel jong in vergelijking met bijvoorbeeld de zon, die ongevee 5 miljard jaar oud is, en zijn ongeveer ontstaan op het moment dat dinosaurussen uitstierven (ongeveer 66 miljoen jaar geleden). De *Homo habilis* en *Homo erectus* ontstonden rond 4,5 en 2 miljoen jaar geleden: dit is ongeveer de leefdtijd van de jongste clusters in de Gouden Riem, zoals de Orion Wolk Cluster, zie het onderste paneel van Fig. 5.13.

Als deze eerste mensen naar de sterren zouden kijken tijdens een heldere nacht, zouden ze een andere hemel zien dan wat wij vandaag waarnemen. Sterren bewegen over de hemel. Ze draaien rond het centrum van ons sterrenstelsel, de Melkweg, volgens een vast patroon hoewel er ook lokale snelheidsverschillen zijn. De leden van OB associaties waren waarschijnlijk geboren uit dezelfde wolken van gas en stof, en delen daarom niet allen dezelfde rotatiesnelheid maar ook dezelfde lokale snelheid. Dit kenmerk is vaak gebruikt om leden van verschillende OB associaties te scheiden van andere sterren.

Een deel van de O en B type sterren beweegt met snelheden die hoger zijn dan men zou verwachten: deze worden *wegrensterren* genoemd. Wegrensterren krijgen hun snelheid niet mee tijdens hun geboorte, maar tijdens hun leven. Het ontstaan van zulke hoge snelheden valt uit te leggen volgens twee verschillende mechanismen. Het eerste mechanisme voorspelt dat een wegrenstel voorheen lid was van een binair stersysteem, bestaande uit twee zware sterren. Als een van de sterren explodeert als een supernova, wordt de ander plotseling uit het systeem gegooid en met grote snelheid gelanceerd. Het tweede mechanisme voorspelt dat een wegrenster is gevormd is een

---

[4]1 parsec komt overeen met ongeveer 3×10^{16} m, of 3,28 lichtjaren

dichtbevolkte jonge stercluster, zoals de Orion Nevel Cluster. In zo een omgeving kunnen twee binaire stersystemen elkaar dichtbij passeren en elkaar uit balans brengen.

Gaia

Om de structuur en sterformatie geschiedenis van jonge associaties te begrijpen hebben we nauwkeurige informatie over de afstanden tussen sterren, hun bewegingen en leeftijden nodig. De data van de ESA *Gaia* ruimteschip is cruciaal hiervoor. Het hoofddoel van de *Gaia* missie is om de grootste en meest precieze driedimensionale plattegrond van ons sterrenstelsel te maken, door het meten van de posities en bewegingen van meer dan een miljard sterren in ons sterrenstelsel en daarbuiten. De positie van een ster op de hemel verandert met tijd doordat het zich relatief van de zon beweegt, en doordat de aarde rond de zon draait. De projectie van de beweging van een ster relatief tot de zon word de *eigenbeweging* genoemd. Dit is een hoeksnelheid (hoek per tijdseenheid). De overeenkomende snelheid is de *tangentile snelheid*. Wanneer de snelheid tot een object bekend is, kan met de hoeksnelheid de tangentile snelheid berekend worden. Wanneer de radiale snelheid ook gemeten is, kan de *totale snelheid* verkregen worden door het optellen van de tangentile en radiale snelheden. De zichtbare beweging van een ster door de rotatie van de aarde ron de zon wordt de *parallactische beweging* genoemd, and is gecorreleerd met de afstand tot de ster. Hoe kleiner de parallactische beweging is, hoe groter de afstand tot de ster is.

Gaia is niet de eerste missie met dit doel. In de jaren negentig produceerde de *Hipparcos* satelliet een catalogus van posities en bewegingen voor ongeveer honderdduizend sterren, waardoor sterren in de dichtstbijzijnde OB associaties konden worden bestudeerd. Deze data was echter niet nauwkeurig genoeg om de driedimensionale structuur te verkrijgen van zelfs de dichtstbijzijnde associatie, Scorpius-Centaurus, of om Orion nauwkeurig in kaart te brengen. *Gaia* verbetert Hipparcos significant voor verschillende redenen. *Gaia* meet bijvoorbeeld the sterposities en bewegingen 200 keer nauwkeuriger dan *Hipparcos*. Ter vergelijking, Gaia's precisie is zo hoog dat het de hoek kan meten die overeenkomt met de diameter van een munt op de maan, terwijl *Hipparcos* alleen de lengte van een astronaut kan meten.

Dit proefschrift

Dit proefschrift gebruikt de eerste *Gaia* data (*Gaia* DR1 and DR2) om een gedetailleerde telling van jonge sterpopulaties in de buurt van de zon te verkrijgen. Figuur 5.14 laat zien waar de zware sterren en pre hoofdreeks sterren zich bevinden in een *kleurmagnitude diagram* van de Orion regio. Een kleur-magnitude diagram is een manier om sterren te classificeren door middel van hun helderheid, temperatuur, massa en hun evolutionaire fase. Als een ster waterstof begint te verbranden verplaatst het zich naar de hoofdreeks (gestreepte lijn in Fig. 5.14). De positie van een ster op de hoofdreeks hangt van zijn massa af: de hoge hoofdreeks (grijze rechthoek links boven in het diagram) is waar massieve, vroegtype sterren zich bevinden. Net voordat sterren op de hoofdreeks uitkomen, zijn ze gelokaliseerd op de pre-hoofdreeks (grijze ellips): in deze fase trekken sterren zich samen waardoor hun temperatuur stijgt, maar waterstofverbranding is nog niet begonnen. Nieuwe sterren komen op de pre-hoofdreeks als zij bijna al hun massa hebben verkregen en hun envelop van gas en stof weg hebben geblazen. De pre-hoofdreeks fase duurt ongeveer een paar miljoen jaar tot tientallen
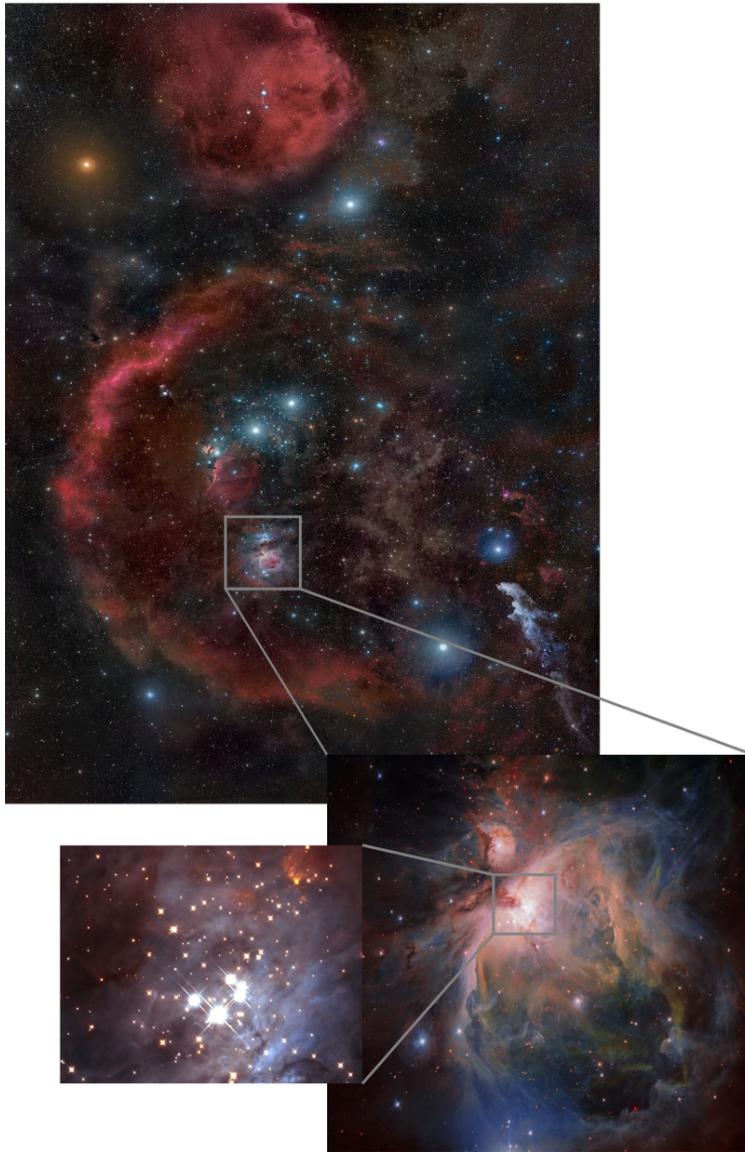
Figuur 5.13: Boven: sterren en gas in het Orion gebied. De heldere rode gebied is de Barnards's loop (Rogelio Bernal Andreo, *DeepSkyColors.com*). Rechts onder: de Orion Nevel, waar op dit moment nieuwe sterren worden gevormd (ESO/G. Beccari). Links onder: centrale gedeelte van de Orion Nevel Cluster (K.L. Luhman (Harvard-Smithsonian Center for Astrophysics, Cambridge, Mass.); and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thompson (Steward Observatory, University of Arizona, Tucson, Ariz.) and NASA/ESA).
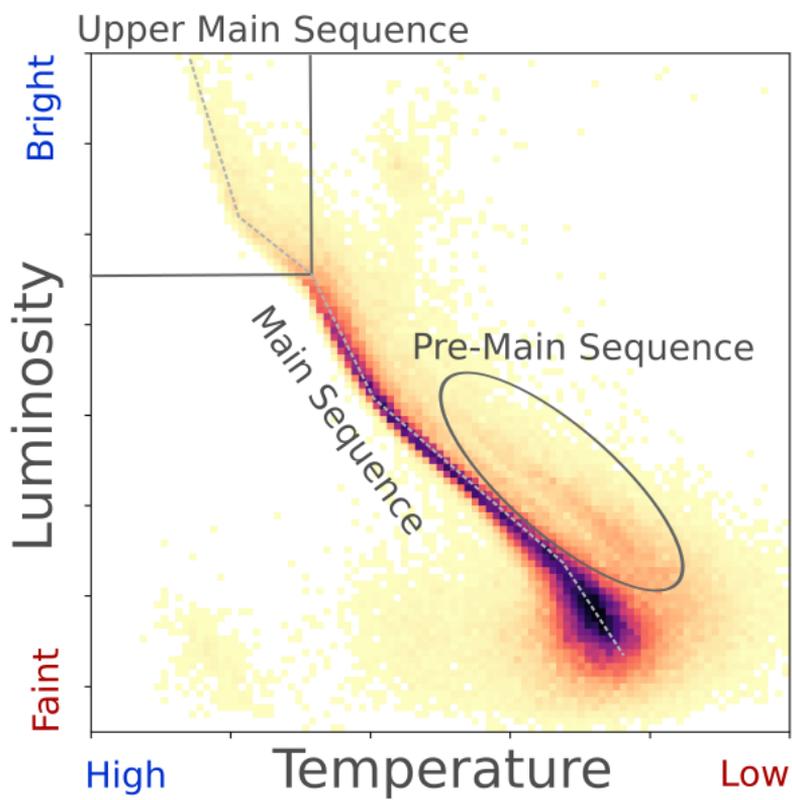
Figuur 5.14: Kleur-magnitude diagram van de sterren in de Orion regio. De gestreepte grijze lijn laat de hoofdreeks zien; het grijze rechthoek geeft een schatting van de grenzen van de hoge hoofdreeks; de grijze ellipse laat zien waar sterren voor de hoofdreeks zich bevinden.

miljoenen jaren afhangende van de stermassa: dit is erg kort vergeleken met de gemiddelde tijd dat een ster zich op de hoofdreeks bevindt.

Het doel van dit proefschrift is om de vorming en verspreiding van OB associaties te begrijpen, wat de eigenschappen van de sterpopulaties binnen een enkele associatie zijn, wat de eigenschappen van OB associaties samen zijn in termen van hun positie aan de hemel, en hoe dit zich vergelijkt met wat is waargenomen voor andere sterrenstelsels. Hiervoor bestuderen we de Orion OB associatie en andere gebieden waarin sterren worden gevormd in de buurt van de zon. De focus ligt op het beantwoorden van de volgende vragen:

- Wat zijn de sterpopulaties in de Orion OB associatie?

- Wat is de stervormingsgeschiedenis van de Orion OB association?

- Wat is de structuur van de omgeving van de zon zoals waargenomen door jonge sterren?

- Hoe veel wegrensterren zijn er in de omgeving van de zon?

Het onderzoek dat gepresenteerd is in **Hoofdstuk 2** is de eerste stap tot het ontrafelen van de complexiteit van de stervormingsgeschiedenis van Orion, in termen van de verschillende sterformatie episodes, hun duur, en hun effect op de omliggende interstellaire medium. De Gaia DR1 data geeft het bewijs van de aanwezigheid van jonge sterpopulaties die zich bevinden rondom bekende clusters. De geschatte leeftijden van de leden van deze populatie suggereren de aanwezigheid van een leeftijdsvolgorde in de associatie.

Deze conclusies zijn deels herzien in **Hoofdstuk 3**. De hogere nauwkeurigheid van Gaia DR2 in vergelijking met Gaia DR1 and het hogere aantal sterren voor welke het mogelijk is de afstanden en snelheden te bepalen maakt het mogelijk om de driedimensionale configuratie van de stergroepen die samen de Orion OB associaties vormen en hun bewegingen en leeftijden te bestuderen. Het belangrijkste resultaat van dit hoofdstuk is dat de sterformatie processen in Orion een complexe geschiedenis volgen, welke veroorzaakt zijn door kinematische en fysische substructuren.

De focus van **Hoofdstuk 4** is de complete omgeving van de zon. Driedimensionale plattegronden van de verspreiding van pre-hoofdreeks sterren en hoge hoofdreeks sterren laten deze prominente structure zijn, Scorpius-Centaurus, Orion, en Vela (zie Fig. 5.15). De distributie van de pre-hoofdreeks sterren als een functie van hun leeftijd laat zien dat jongere clusters in dichte, compacte hopen, and zijn omgeven door oudere bronnen, welke meer uitgespreid zijn. Deze plattegronden geven geen bewijs van het bestaan van de ring-achtige structuur welke normaal gesproken gerefereerd wordt als de Gouden Riem.

Hoofdstuk 5 presenteert een onderzoek naar wegrensterren in een straal van 1 kiloparsec van de zon. Kandidaat wegrensterren zijn geselecteerd uit hoge hoofdreekssterren, and zijn geselecteerd door hun tangentile snelheid, en, wanneer mogelijk, hun totale snelheid. Dit bevat sterren met een tangentile snelheid die significant afwijkt van de rest van de populatie of met totale snelheden hoger dan $30\,\mathrm{km\,s}^{-1}$. De analyse is gefocust op wegrensterren in de Orion en Scorpius-Centaurus (Sco-Cen) gebieden. Zes nieuwe wegrensterren zijn toegevoegd aan de reeds bekende wegrensterren in Orion. In Sco-Cen zijn twee wegrensterren gedentificeerd die waarschijnlijk dezelfde oorsprong hebben.
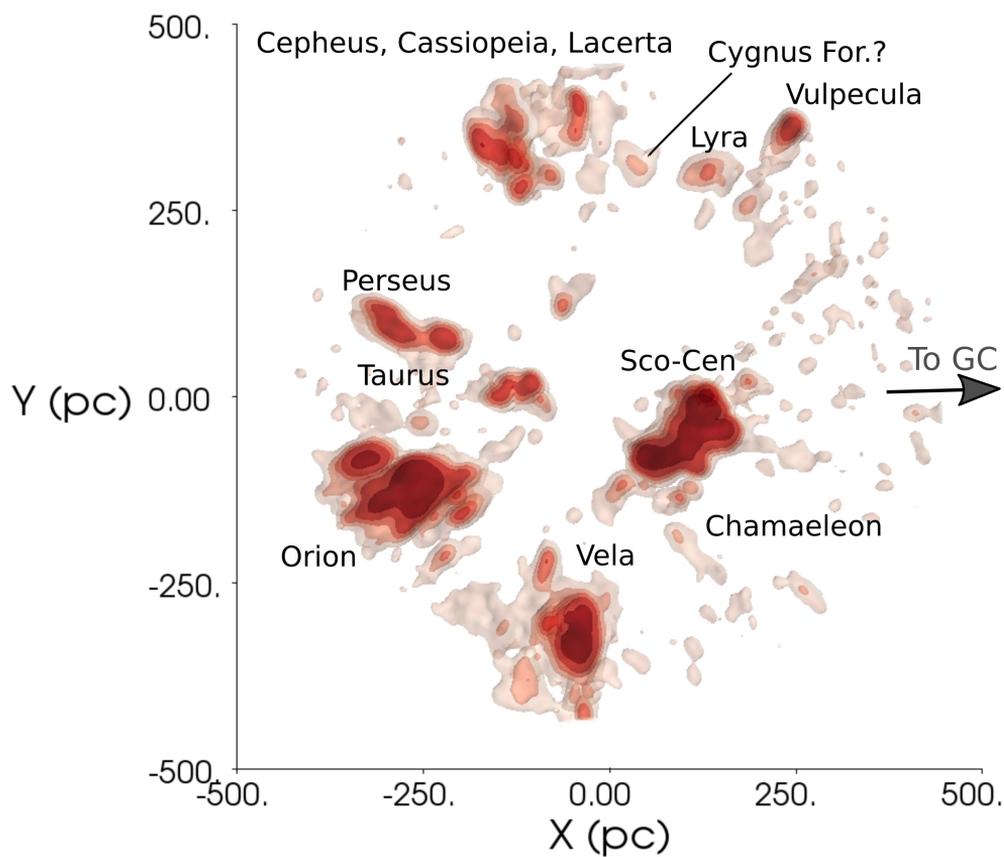
Figuur 5.15: Dichtheidsdistributie van pre-hoofdreeks bronnen jonger dan 20 miljoen jaar. De zon is in het centrum, in $(X, Y) = (0, 0)$, de x-as is richting het centrum van de melkweg (deze richting is aangegeven door de pijl), en de y-as richting de galactische rotatie.

Conclusies

De belangrijkste conclusie van dit proefschrift is dat sterformatie processen op grote schaal, welke leiden tot de vorming van OB associaties, complex zijn en niet goed begrepen. Er is geen algemene theorie over stervorming dat alle substructuren (ruimte, kinematica, en leeftijden) kan verklaren in Orion en in andere OB association in de buurt van de zon. De oorsprong van OB associaties blijft mysterieus. De structuur van de omgeving van de zon is zeker anders dan wat gedacht was voor het Gaia tijdperk. Deze bevindingen vragen om een herziening van theorien omtrent de propagatie en de totstandbrenging van sterformatie. Toekomstige nieuwe waarnemingen door de Gaia satelliet en van aankomende spectroscopie missies zullen bijdragen aan gedetailleerd verkennen van de kinematica en de fysische substructuren van grote stervorming complexen.

Riepilogo

La figura di Orione il cacciatore è una vista familiare nel cielo invernale dell'emisfero Boreale. L'intera area della costellazione, che è mostrata nel pannello grande della Fig. 5.16, è un sito di formazione stellare straordinariamente attivo ed è stato oggetto di numerosi studi astronomici. Orione ospita numerosi giovani ammassi stellari, sovrapposti lungo la linea di vista, e a diversi stadi evolutivi. La spiegazione della dettagliata sequenza di eventi (la cosiddetta storia della formazione stellare) che hanno causato la formazione di una popolazione tanto numerosa è uno degli argomenti principali di questa tesi.

Quasi tutte le stelle blu e brillanti visibili nella Fig. 5.16 appartengono alla "associazione OB di Orione". Le associazioni OB sono state inizialmente identificate come gruppi sparsi di stelle giovani e massive. Queste strutture, le cui dimensioni fisiche sono nell'ordine di centinaia di parsec [5], sono l'ultimo stadio del processo di formazione stellare di stelle massive e il contesto in cui nascono nuove stelle. Per esempio, l'attuale formazione stellare nella nebulosa di Orione (5.16, pannello in basso) è connessa alle precedenti generazioni di stelle massive nei gruppi adiacenti tra loro dell'associazione.

Analogamente, la formazione di singole associazioni OB può essere relazionata a eventi di formazione stellare su larga scala, che avvengono su scale di centinaia di parsec. Nelle vicinanze del Sole (entro 500 parsec), questi eventi hanno dato origine a numerose associazioni. Storicamente si ritiene che le associazioni formino una struttura ad anello, chiamata "Gould Belt" (la cintura di Gould). Le associazioni e gli ammassi stellari che compongono la Gould Belt sono molto giovani se confrontati con il Sole, che ha circa cinque miliardi di anni, e hanno iniziato a formarsi all'incirca quando i dinosauri si sono estinti (circa 66 milioni di anni fa). I più antichi resti di Homo habilis e Homo erectus sono datati tra i 4.5 e i 2 milioni di anni fa: questa è circa l'età degli ammassi più giovani della Gould Belt, come l'ammasso della nebulosa di Orione mostrato nella Fig. 5.16, in basso.

Se questi uomini primitivi avessero osservato le stelle in una notte serena, avrebbero visto un cielo leggermente diverso da quello che osserviamo in questo momento. Le stelle infatti si muovono. Esse orbitano in modo ordinato intorno al centro della nostra galassia, la Via Lattea, ma ci sono anche degli andamenti di velocità locali che sono differenti da regione a regione. I membri delle associazioni OB nascono probabilmente dalle stesse nubi di gas e polvere, e pertanto condividono non solo la stessa velocità di rotazione, ma anche quella locale. Questa proprietà è spesso sfruttata per separare le stelle che appartengono a una associazione OB dalle altre stelle nella stessa regione.

Una frazione delle stelle di tipo O e B si muove a velocità più alte della norma: queste stelle sono chiamate "stelle fuggitive" (runaway stars). Le stelle fuggitive non acquisiscono le loro velocità alla nascita, ma durante la loro vita. Per spiegare l'origine di tali velocità, sono stati suggeriti due meccanismi. Il primo prevede che la stella fuggitiva sia in origine parte di in un sistema binario, costituito da due stelle massive.

---

[5]1 parsec corrisponde circa a 3×10^{16} m, o 3.28 anni luce.

Se una delle stelle esplode come supernova, l'altra viene improvvisamente rilasciata dall'energia gravitazionale che lega il sistema, e lanciata ad alta velocità su una nuova traiettoia. Il secondo meccanismo prevede invece che una stella fuggitiva si possa essere formata in un giovane ammasso stellare molto denso, come l'ammasso della nebulosa di Orione. In questo ambiente, due sistemi binari possono avvicinarsi e interagire gravitazionalmente. Una interazione di questo tipo può distruggere entrambi i sistemi, e una o più stelle possono essere espulse ad alta velocità.

Gaia

Dipanare la struttura e la storia della formazione stellare delle giovani associazioni richiede una conoscenza accurata delle distanze, delle velocità e delle età delle stelle che ne fanno parte. I dati della missione spaziale ESA *Gaia* sono cruciali sotto questo punto di vista. L'obiettivo principale di *Gaia* è di comporre la più grande e precisa mappa tridimensionale della nostra galassia, misurando le posizioni e i moti di più di un miliardo di stelle al suo interno e oltre. La posizione di una stella in cielo cambia nel tempo in seguito al suo moto relativo rispetto al Sole e al moto della Terra intorno al Sole. La proiezione del moto della stella rispetto al Sole sulla sfera celeste è chiamato "moto proprio". Il moto proprio è una velocità angolare (l'arco sul cielo percorso in un certo tempo). La corrispondente velocità è la "velocità tangenziale". La velocità tangenziale può essere calcolata dalla velocità angolare se la distanza della stella è nota. Se la proiezione della velocità della stella lungo la linea di vista (la "velocità radiale") può essere misurata, si ottiene la "velocità totale" sommando la velocità tangenziale a quella radiale. Il moto apparente di una stella dovuto alla rotazione della Terra intorno al Sole si chiama "moto parallattico", e dipende dalla distanza della stella. Più il moto parallattico è piccolo, più la distanza della stella è grande.

Gaia non è la prima missione del suo genere. Negli anni novanta, il satellite *Hipparcos* ha prodotto un catalogo di posizioni e moti di circa centomila stelle. *Hipparcos* ha permesso di effettuare un censo estensivo delle associazioni OB vicine al Sole. Tuttavia i dati non erano sufficientemente precisi per determinare la struttura tridimensionale nemmeno della associazione più vicina, Scorpione-Centauro, o per permettere di compiere progressi significativi su Orione. *Gaia* migliora significativamente *Hipparcos* per una serie di motivi. Per esempio, *Gaia* misura le posizioni e i moti delle stelle circa 200 volte più accuratamente di *Hipparcos*. La precisione di *Gaia* è tanto alta da poter misurare l'angolo che corrisponde al diametro di una moneta sulla Luna, mentre *Hipparcos* riusciva a misurare solo quello corrispondente all'altezza dell'astronauta che la teneva in mano.

Il contributo di questa tesi

Questa tesi fa uso delle prime pubblicazioni dei dati di *Gaia* (*Gaia* DR1 e DR2) per ottenere un censo dettagliato delle popolazioni stellari giovani nelle vicinanze del Sole, esplorando la distribuzione e le proprietà delle stelle giovani e massive e di quelle poco massive, ancora in pre-sequenza principale. La figura 5.17 mostra la collocazione delle stelle massive e in pre-sequenza principale nel "diagramma colore-magnitudine" della regione di Orione. Un diagramma colore-magnitudine è uno strumento usato per classificare le stelle in base alla loro luminosità, temperatura, massa e stadio evolutivo. Quando nel nucleo di una stella iniziano le reazioni nucleari di fusione di idrogeno in elio, la stella entra nella sequenza principale (indicata dalla la linea tratteggiata nel-
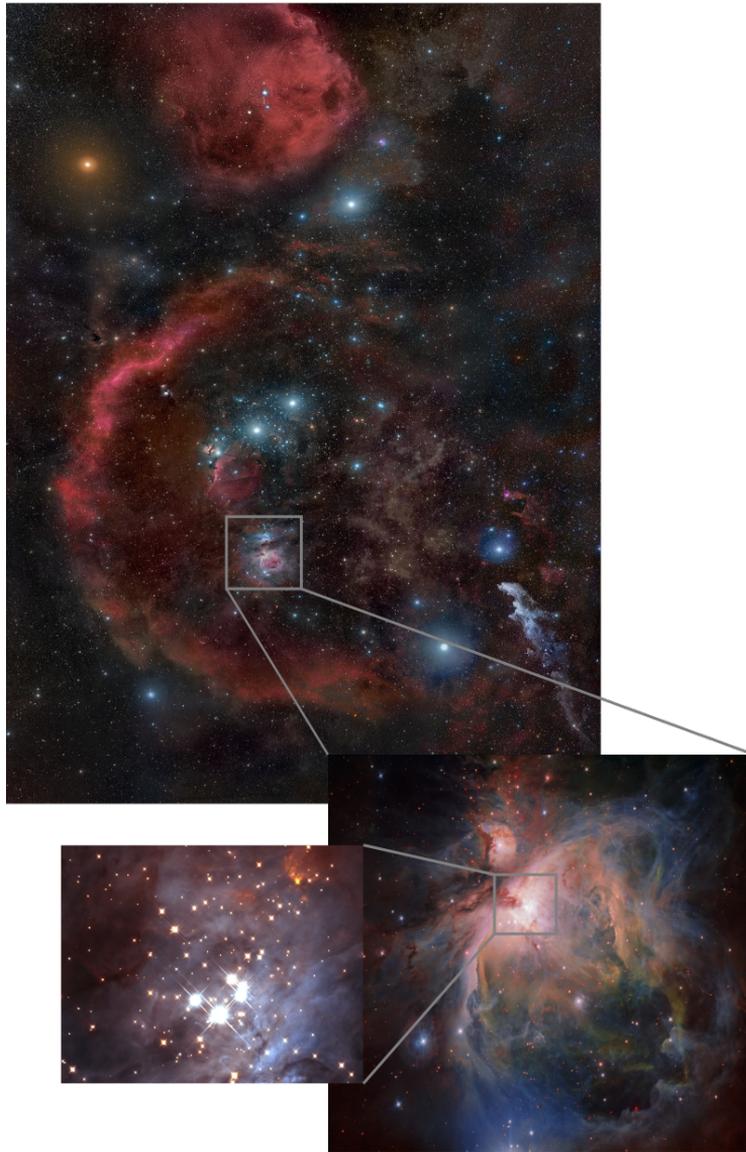
Figura 5.16: Pannello in alto: stelle e gas nella regione di Orione. Il filamento rosso e brillante che crea un arco al centro dell'immagine è il "Barnard's Loop" (Rogelio Bernal Andreo, *DeepSkyColors.com*). Pannello in basso a destra: la nebulosa di Orione, dove stanno avvenendo fenomeni di formazione stellare (ESO/G. Beccari). Pannello in basso a sinistra: regione centrale dell'ammasso della nebulosa di Orione (K.L. Luhman (Harvard-Smithsonian Center for Astrophysics, Cambridge, Mass.); and G. Schneider, E. Young, G. Rieke, A. Cotera, H. Chen, M. Rieke, R. Thompson (Steward Observatory, University of Arizona, Tucson, Ariz.) and NASA/ESA).
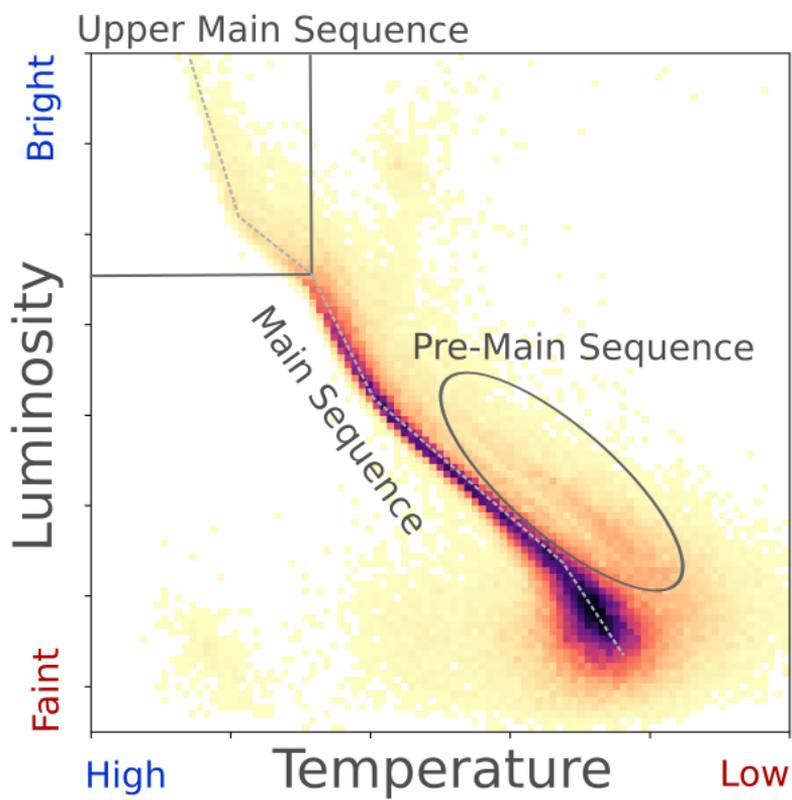
Figura 5.17: Diagramma colore-magnitudine delle stelle nella regione di Orione. La linea grigia tratteggiata indica la sequenza principale (main sequence); il rettangolo grigio delimita l'alta sequenza principale (upper main sequence); l'ellisse grigia evidenzia la pre-sequenza principale (pre-main sequence).

la Fig. 5.17). La posizione di una stella nella sequenza principale dipende dalla sua massa: le stelle massive e giovani si collocano nella sequenza principale alta (rettangolo grigio in alto a sinistra della figura). Prima di entrare nella sequenza principale, le stelle sono collocate nella pre-sequenza principale (ellisse grigia): in questo stadio, le stelle si stanno contraendo, e la loro temperatura si sta innalzando, ma la fusione dell'idrogeno non è ancora partita. Le stelle giovani entrano nella pre-sequenza principale dopo aver acquisito quasi tutta la loro massa, e dopo aver spazzato via l'involucro di gas e polvere che le avvolgeva. La fase di pre-sequenza dura da qualche milione a qualche decina di milioni di anni, a seconda della massa della stella: un tempo molto corto rispetto a quello che, in media, la stella passa in sequenza principale.

Gli obiettivi di questa tesi sono comprendere come si formano e disperdono le associazioni OB, quali sono le caratteristiche delle popolazioni stellari nelle singole associazioni, quali sono le proprietà dell'insieme delle associazioni OB in termine della loro disposizione spaziale, e paragonare queste informazioni a ciò che viene osservato in altre galassie. Questi argomenti sono affrontati usando la associazione OB di Orione come banco di prova per studiare i meccanismi che portano alla formazione di una associazione OB, studiando la disposizione delle associazioni OB e delle regioni di formazione stellare entro 500 pc dal Sole, e infine analizzando le proprietà cinematiche delle stelle massive di tipo O e B entro 1000 pc dal Sole.
Il fulcro è, in particolare, rispondere alle seguenti domande:

- Quali sono le popolazioni stellari della associazione di Orione?

- Qual è la storia della formazione stellare dell'associazione di Orione?

- Qual è la struttura della regione vicino al Sole, tracciata dalle stelle giovani?

- Quante stelle fuggitive ci sono nelle vicinanze del Sole?

Lo studio presentato nel **Capitolo 2** rappresenta il primo passo per svelare la complessità della storia della formazione stellare di Orione, dal punto di vista dei vari episodi di formazione stellare, della loro durata, e dei loro effetti sul mezzo interstellare. I dati di *Gaia* DR1 dimostrano la presenza di una popolazione stellare giovane, distribuita intorno ad ammassi noti. Le età stimate per i membri di questa popolazione suggeriscono la presenza di una precisa sequenza di età.

Questo conclusioni sono state in parte rivisitate nel **Capitolo 3**. La maggiore accuratezza di *Gaia* DR2 rispetto a *Gaia* DR1 e il più alto numero di stelle per cui è possibile determinare distanze e velocità tangenziali hanno permesso uno studio meticoloso della configurazione spaziale dei gruppi stellari che compongono la associazione OB e delle loro velocità ed età. La conclusione principale di questo capitolo è che gli eventi di formazione stellare in Orione seguono una sequenza complessa, che ha causato la presenza di sotto-strutture sia spaziali che cinematiche.

Il **Capitolo 4** si concentra sull'intera regione nelle vicinanze del Sole (la regione entro 500 parsec dal Sole). Mappe in tre dimensioni della distribuzione spaziale delle stelle in pre-sequenza principale e nella parte alta della sequenza principale mostrano tre strutture prominenti, Scorpione-Centauro, Orione e Vela (si veda la Fig. 5.18). La distribuzione delle stelle in pre-sequenza principale in funzione della loro età mostra che le stelle più giovani si raggruppano in ammassi compatti, e sono circondati dalla stelle più vecchie, la cui distribuzione spaziale è invece più diffusa. Notevolmente, le mappe non mostrano alcuna evidenza della presenza della Gould Belt.
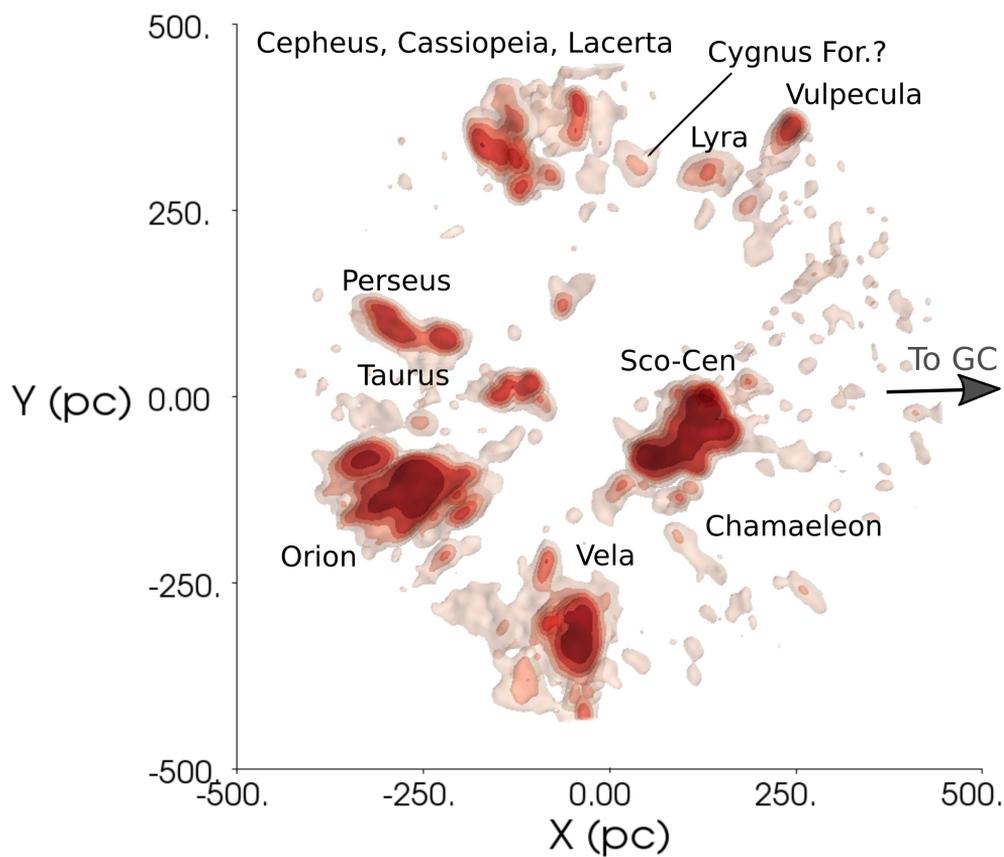
Figura 5.18: Distribuzione della densità delle stelle in pre-sequenza principale più giovani di 20 milioni di anni nel piano galattico. Il Sole è al centro, a $(X, Y) = (0, 0)$, l'asse x è diretto verso il centro galattico (la cui direzione è indicata dalla freccia), e l'asse y segue la direzione della rotazione galattica.

Il **Capitolo 5** presenta una ricerca delle stelle fuggitive entro 1000 parsec dal Sole. Le candidate stelle fuggitive sono selezionate tra le stelle in alta sequenza principale, e classificate come fuggitive usando la loro velocità tangenziale e, quando possibile, la loro velocità totale. In particolare, le candidate stelle fuggitive sono definite come le stelle che hanno velocità tangenziale significativamente differente dal resto della popolazione, o velocità totali più alte di $30\,\mathrm{km\,s^{-1}}$. L'analisi si concentra sulle stelle fuggitive candidate nelle regioni di Orione e Scorpione-Centauro. In Orione, sono state aggiunte sei nuove candidate all'insieme di stelle fuggitive note. In Scorpione-Centauro, sono state identificate due candidate che probabilmente condividono la stessa origine.

Conclusioni

La conclusione principale di questa tesi è che gli eventi di formazione stellare che conducono alla formazione delle associazioni OB sono complessi e non compresi appieno. Non c'è una teoria della formazione stellare che spieghi completamente la sotto-struttura (spaziale, cinematica e evolutiva) osservata in Orione e in altre associazioni OB. La stessa origine delle associazioni OB rimane in qualche modo misteriosa. La struttura della regione vicino al Sole è innegabilmente diversa da quello che si pensava prima di *Gaia*. Queste scoperte richiedono una revisione delle teorie che spiegano l'innesco e la propagazione della formazione stellare. Le future pubblicazioni dei dati del satellite *Gaia* e di future indagini spettroscopiche contribuiranno all'esplorazione più dettagliata delle sotto-strutture fisiche e cinematiche dei grandi complessi di formazione stellare.

List of publications

First author

Searching for runaway stars in Gaia DR2
E. Zari, T. Marchetti, A. G. A. Brown, and P.T. de Zeeuw, A&A, in prep.

Structure, kinematics, and ages of the young stellar populations in the Orion region
E. Zari, A. G. A. Brown, and P.T. de Zeeuw, 2019, A&A, in press, `arXiv:1906.07002`

3D mapping of young stars in the solar neighbourhood with Gaia DR2
E. Zari, H. Hashemi, A. G. A. Brown, K. Jardine, and P.T. de Zeeuw, 2018 A&A, 620, 172

Mapping young stellar populations toward Orion with Gaia DR1
E. Zari, A.G.A. Brown, J. de Bruijne, , C.F. Manara, and P.T. de Zeeuw, 2017, A&A, 608, 148

Herschel-Planck dust optical depth and column density maps. II. Perseus
E. Zari, M. Lombardi, J. Alves, C.J. Lada, and H. Bouy, 2016, A&A, 587, 106

Contributing author

3D shape of Orion A from Gaia DR2
J. E. Grossschedl, J. Alves, S. Meingast, C. Ackerl,
J. Ascenso, et al. (incl. **E. Zari**), A&A, 619, 106

The Gaia-ESO Survey: a kinematical and dynamical study of four young open clusters
L. Bravi, **E. Zari**, G.G. Sacco, S. Randich, R. D. Jeffries, et al., 2018, A&A, 615, 37

A Gaia study of the Hyades open cluster
S. Reino, J. de Bruijne, **E. Zari**, F. d'Antona, P. Ventura, 2018, MNRAS, 477, 319

The Gaia-ESO Survey: open clusters in Gaia-DR1 . A way forward to stellar age calibration
S. Randich, E. Tognelli, R. Jackson, R. D. Jeffries, S. Degl'Innocenti, at al. (incl. **E. Zari**), 2018, A&A, 612, 99

Curriculum Vitae

I was born in Milano, Italy, on 9 February 1989. Like other kids, I first wanted to become an astronaut, then a paleontologist. At some point towards the end of middle school I was set to become a journalist, so I decided to focus my high school studies on humanities, and I enrolled the Liceo Classico G. Berchet in Milano. After having successfully learned to translate Latin and ancient Greek texts, in 2008 I found myself attending Physics at Università Statale di Milano, where I completed both my Bachelor's and Master's degree.

My first project related to Astrophysics was my Bachelor's thesis, which I completed in 2012 under the supervision of Prof. dr. Marco Lombardi. In the thesis, I analysed the data of the *Herschel* satellite, and obtained dust maps of the Pipe Nebula. I continued to work with Marco when, one year later, I started my Master's thesis. The project was based on the analysis of the *Herschel* and *WISE* data in the Perseus molecular cloud. The goal was to study the local Schmidt law, that is the relation between gas density and star formation rate within single molecular clouds. A significant fraction of the thesis was carried out at the University of Vienna, under the supervision of Prof. dr. J. Alves.

After my graduation, I started to look for a PhD in Italy. In October 2014, I won a position in Milano, although not in Astrophysics. I thus decided to resign and I started working as a supplementary physics and maths teacher at the Istituto Orsoline di Saronno. At the same time, I applied for PhD positions in Germany and the Netherlands. As a result, in April 2015 I landed in Schipol with a 32 kg suitcase, and took the first of many trains to Leiden. Under the guidance of Dr. Anthony Brown and Prof. dr. P.T. de Zeeuw, I used the first *Gaia* data releases to study the star formation history of the Orion region, the three dimensional structure of the solar neighbourhood, and the kinematic properties of young, early-type stars, looking for runaway star candidates. I presented my work in conferences in Europe and in the US. In 2016 I also attended the 11th Heidelberg IMPRS Summer School on astrostatistics and data mining. Besides my research work, during my years at the Sterrewacht I served as a teaching assistant for the course "Project management for scientists", taught by Prof. dr. C. Keller, and organised the master's student colloquia with Dr. J. Hodge. I also had the opportunity to organise the PhD talks in 2017 as well as the Borrel in 2017/2018.

In early 2019 I accepted a post-doc position in Heidelberg, to work with Hans-Walter Rix on the preparation and exploitation of the SDSS-V survey of young stars.

Acknowledgements

In these many years towards the end of my education, I met many people who supported, inspired, and motivated me. I am grateful to all of them.

I would like to start by thanking my family: without them, none of this would have been possible. Grazie per avermi fatto scoprire l'astronomia, e in generale le scienze, fin da bambina: ricordo ancora la prima conferenza al Planetario di Milano, il primo libro di Astronomia letto al mare, il film Apollo 13, e come non menzionarle, tutte le puntate di Super Quark, con Piero e Alberto Angela, Paco Lanciano, e Danilo Mainardi. Grazie per tutti i vostri sacrifici, che mi hanno permesso di seguire sempre i miei desideri. Grazie per aver accettato tutte le mie scelte, anche quando non le capivate, e per avermi incoraggiata e spronata anche a distanza.

My research career in Astronomy started in Milano and Vienna, where I was supervised by Marco Lombardi and João Alves. Marco and João, thank you for teaching me what research is about, for motivating me, and for being always available for me. In my university years I shared memorable moments with some special friends: Alessia, Amedeo, Benedetta, Carlo, Charlotte, Chiara, Chiara, Claudia, Enrico, Ettore, Federico, Francesca, Guglielmo, Luca, Marco, Maria Giulia, Michela, Valeria: grazie per i pranzi, le pause, i laboratori, le ore di studio in biblioteca, le corse, le vacanze e i viaggi! Gulielmo, abbiamo condiviso non solo gli anni universitari, ma anche quelli di dottorato: grazie per le passeggiate ad Amsterdam, per le intense sere di scrittura della tesi, e per le tante risate insieme. Thanks also to my friends in Vienna, in particular Stefan and Josefa. I fondly remember the time I spent in Vienna. Thank you for welcoming me as part of the group since the very first day!

I feel very privileged to have spent the last years as a PhD student in Leiden. Living in such a welcoming country, working in a vibrant department, and meeting so many people from all over the world, truly opened my mind. I would like to first thank the computer and secretarial staff, for keeping everything running smoothly. Thank you in particular to Leonardo Lenoci, who helped me installing the software required for creating some of the figures of Chapter 4. I am grateful to Xander Tielens, for the support in the final part of my PhD and for his encouraging words. I would also like to thank Tim de Zeeuw for the discussions and advice on different topics, and for having introduced me to the Spaghetti Western genre. I will never be able to fully express my gratitude to Anthony Brown for giving me the opportunity to embark on my PhD. Anthony, I feel very lucky for having experienced your balanced guidance, your care, and your expertise throughout my entire PhD.

Thank you to all the participants of the Leiden-ESTEC *Gaia* group meetings, and in particular: Timo Prusti, Jos de Bruijne, Daniel Michalik, and Alice Zocchi. You have always made the meeting active and insightful with your comments. I would also like to thank Carlo Manara, for his support in the first steps of my PhD, and for giving me the opportunity to join him in an observing trip to Paranal. Thanks to Arthur Bosman, Merel van 't Hoff, Sierk van Terwisga, Lukasz Tychoniec, and all

the participants to the Star Formation coffees, for sharing with me their expertise on disc chemistry, jets, and proto-stars, and in particular to Daniel Harsono, for the many interesting conversations. Special thanks also to my office-mates, Alvaro Hacar, Nadia Murillo, and Giovanni Rosotti. Alvaro, it has been a pleasure to share an office with you in Leiden (and Vienna as well!). Thank you for the scientific and non-scientific discussions, and for the infinite amount of cookies you provided me during the long winter afternoons. Thanks also to all the participants and the organisers of the *Gaia* sprints, which have always been an incredibly useful learning experience. I would like to thank in particular Sven Buder, Christina Eilers, Keith Hawkins, Eloisa Poggio, and Sara Rezaei for the many fun discussions.

During these years in Leiden, Amsterdam, and Padova, I shared lunches, coffee breaks, borrels, dinners, and drinks with many people who always provided topics for interesting and stimulating conversations: Alberto, Alice, Allison, Andrej, Andrew, Anna, Anna Grazia, Ann-Sofie, Aurelien, Christian, Chris, Clement, Emanuele, Eva, Francisca, Giada, Heather, Hiddo, Igone, Javi, Jorrit, Josh, Kirsty, Leah, Luis Henry, Luke, Macla, Mattia, Marcelo, Matus, Mike, Nico, Niels, Omar, Pedro, Salvatore, Stella, Vincenzo, Ylva... and many others. Thanks everyone! Special thanks to the PhD students, past and present, who volunteered to bring us the Friday borrels. David, Mason, Allison, Christian, Ann-Sofie, Chris, Nico, Pedro, Santiago, Kim, Maria Cristina, Francisca, Yours Truly, Kirsty, Hiddo, Stijn, Fraser, Anna, Mantas, Martijn, Sanjana: thank you for helping everyone relax every Friday!

With a small group of colleagues I have a special bond. During these four years and half, they made my life special, and certainly their support and our friendship contributed to the positive outcome of this PhD. Dilovan and Fraser, thank you for being the extra office-mates of my final year: the chats and coffee breaks with you have helped me to unwind during some of the toughest times of my PhD. Gabriella, grazie per i fantastici wine and cheese e per le incredibili performances canore! Santiago, I admire your relaxed attitude in all the things you do. I am glad that we have organised the *Gaia* meetings together! Valeriya, your achievements are inspirational to me, both as a scientist and as a pastry chef. Thank you for all your delicious cakes! Kim, your motivation and your passion for science are really amazing! Grazie per le corse, i borrel, e le ricette condivise! Jit and Christos, the times I have spent with you have always been fun. Thanks for all the playing and singing together, and for sharing with me all the ups and downs of the PhD life. Francesco, sei tra i protagonisti e artefici di alcuni dei miei migliori ricordi a Leiden. Grazie per le serate pre-elettorali, i film di Fantozzi, le canzoni di Battiato e gli inni partigiani. Marco e Gabi, l'ultimo anno ad Amsterdam è stato fantastico anche grazie a voi. Grazie per i brunch, per le cene del sabato sera, e per tutti i momenti trascorsi insieme. Gabi, grazie per tutto quello che abbiamo condiviso in questi quattro anni, specialmente nel periodo delle applications. Le lunghe discussioni sul futuro mi hanno sempre aiutata a fare chiarezza sui miei prossimi passi. Mieke, you are one of the first people I met here at the sterrewacht, and I am so glad we kept our friendship through the years. You are a woman of action, always busy in your many activities, but you have always found time to help me. You are one of the reasons I have felt so welcome in this country! Aayush, I am thankful for all our discussion about life and work. I deeply cherish the many memories of our times together (they are so many I can not decide which one to mention), but most importantly, thank you for trying to teach me the difference between "hair" and "air".