This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

The dynamics of star formation in the local Milky Way as traced by OB associations

Alexis Laurent Pierre-Henri Quintana Isasi

Doctor of Philosophy

Astrophysics, Keele University

October 2023

Abstract

OB associations are gravitationally unbound, low-density groups of young stars delineated by bright OB stars. At the intersection of star formation and Galactic structure, they are important across many areas of astronomy. However, most historical OB associations had their membership defined several decades ago. Recent data shows a lack of kinematic coherence amongst them, alongside many contaminants, which encourages us to revisit them.

I have developed a code that fits a model spectral energy distribution (SED), based on stellar atmosphere and evolutionary models and an extinction map, to an observed SED, based on photometry from several surveys and *Gaia* parallaxes. This process derives stellar parameters that are used to identify OB stars. I used this to identify OB stars across Cygnus, Auriga and Cassiopeia, identifying thousands of OB stars. I applied clustering algorithms to identify kinematically-coherent OB associations, which I characterized physically and kinematically.

In Cygnus, the OB associations form two groups exhibiting a large-scale expansion pattern. A kinematic traceback revealed that they reached their most compact state ∼8 Myr ago. This expansion could either be attributed to feedback or turbulence within their primordial molecular cloud. In Auriga, the OB associations are connected to the surrounding open clusters and star-forming regions. They follow an age gradient, up to ∼20 Myr ago, coinciding with the motion of the Perseus spiral arm. In Cassiopeia, the kinematics of OB stars suggest an interesting large-scale pattern similar to that seen in Cygnus, though further investigating suggests it may be due to Galactic rotation. Despite this, the OB stars in Cassiopeia can be traced back to a compact configuration ∼17 Myr ago.

This thesis highlights the importance of identifying OB associations with recent data and modern techniques. Not only are they tracers of early stellar evolution, helping to grasp the physical condition driving the expansion of stellar groups, but they can also be used to reconstruct the motion of the spiral arms in the recent past.

Acknowledgements

When Frodo Baggins had the insurmountable task of carrying the One Ring into Mordor, he said: ”I cannot do this alone”. While my own quest was obviously not as strenuous, it was nonetheless the longest and biggest project of my life so far, for which I received a much valuable help.

I would like to express all my gratitude to my supervisor, Nick Wright, for his constant support throughout the ∼4 years of my PhD. I was always keen to meet him every week, as I had thrilling scientific discussions that has made me feel part of the academic world. He gave me the everyday motivation to carry on my projects no matter the arising challenges on the way. He provided me insightful and regular feedback, which undoubtedly improved my work, despite my tendency of writing (very) long email updates about my progress. Most of all, he encouraged me to pursue a research career and to never let my ambitions down.

I would also like to thanks my co-supervisor, Rob Jeffries, for his complementary comments on my work, particularly during the weekly group meetings. His contribution for the study of the Auriga region was especially useful to improve the quality of this project.

My appreciation also goes for Sara Berlanas, not only for her valuable career advice as a research fellow, but also for the ongoing study of the Cassiopeia region, offering me the opportunity to conduct another exciting and ambitious research project.

Many thanks for my colleagues who made my time at Keele University very pleasant, making it easier to live in another country for the first time in my life. I will remember the regular lunches at the KPA and the drinks at the pubs for the years to come, although sometimes I was hoping they would appreciate more my subtle sense of humour.

Finally, I wouldn’t have accomplished such a long project without the steady support of my family and friends in Belgium. I express my gratitude for my girlfriend, Mezoon Zakiuddin, whom I met during my PhD and who has always motivated to me to do my best and was lifting me up when I was feeling down. Also to my parents and my grandmother, who supported me each in their own way, and who shaped the person I am today.

This thesis is dedicated to my grandfather, Francis Timmermans. He offered me my very first astronomy book when I was six years old, which started my life passion for this subject, so much that already as a child I would turn my eyes into the splendours of the universe. Twenty years later, here I am, completing a PhD degree in astronomy. I wish he was still with us, I would have been glad to discuss my research with him. I am forever grateful for him.

Contents

List of Figures

List of Tables

1 Introduction

The main subject of this thesis is OB associations, which represent the transition between star-forming regions and the galactic field population of stars. This first chapter summarizes our knowledge about star formation, star clusters and associations. It especially focuses on massive OB stars whose role in galactic evolution is important: notably, they shape the structure of spiral arms, affect their environment through feedback processes and bring new chemical elements into the interstellar medium as they evolve and when they die. I also introduce the different regions studied in this thesis and their individual OB associations.

1.1 Star formation

Star formation is a vital phenomenon in astronomy, shaping the universe in all its complexity. This takes place within giant molecular clouds (GMCs) (Mooney & Solomon, 1988; Froebrich & Rowles, 2010; Lada et al., 2010; Kennicutt & Evans, 2012). GMCs are cold, dense and filamentary entities of molecular gas in the interstellar medium, composed of cores and clumps (Chevance et al., 2020). Their life lies in the balance of magnetism, thermal pressure, turbulence and gravitation (see e.g. Girichidis et al. 2020). Some GMCs collapse due to gravity and give birth to protostars (Prialnik, 2000). Protostars then grow through accretion, either core (Bondi, 1952) or competitive (Bonnell et al., 2001), before contracting along the Hayashi tracks in the Hertzsprung-Russell diagram and joining the main-sequence (Henyey et al., 1955; Hayashi, 1961). This succession of steps is detailed below.

1.1.1 GMCs

Giant molecular clouds stand out as overdensities of gas in the interstellar medium (Chevance et al., 2022), with a highly filamentary structure and compact regions known as clumps (Dobbs et al., 2014). GMCs contain the majority of molecular gas in the interstellar medium (see e.g. Ferrière 2001).

GMCs are cold (10-50 K) and dense (up to a few hundreds cm^{-3}) physical entities, evolving across scales of ∼10-100 pc (Stark & Blitz, 1978; Sanders et al., 1985; Chevance et al., 2020). They also are the most obscured regions of the ISM due to their large dust content (Dobbs et al., 2014). GMCs are characterized by different masses and brightnesses depending on their location in galaxies (Heyer & Dame, 2015). Whilst they can reach up to $10^7\ M_\odot$ in the central parts of galaxies (Oka et al., 2001), they can be as low as $100\ M_\odot$ at their edges and at high galactic latitudes (Magnani et al., 1985; Heyer et al., 2001). Similarly, the cloud surface brightness (^{12}CO emission, radio domain) decreases in the outer parts of galaxies (see Heyer & Dame 2015 and references therein).

Numerous mechanisms have been invoked to explain the formation of GMCs, such as gravitational instabilities (Goldreich & Lynden-Bell, 1965), local turbulence (Dobbs et al., 2014) and collision between clouds (Oort, 1954). GMCs are thought to arise from converging flows. Fragmentation, cooling and compression of the diffuse ionized and neutral hydrogen of the ISM occur (Dobbs et al., 2014; Chevance et al., 2022). Once enough material is accreted, atomic gas is converted to molecular gas (especially H_2). These molecules are then protected against photodissociation from the ambient stellar UV radiation thanks to molecular self-shielding and dust absorption in the far-UV (Sternberg et al., 2014). CO molecules are also produced although the process is more difficult due to the lower abundance of carbon and oxygen atoms (Chevance et al., 2022).

Star formation occurs in GMCs (Mooney & Solomon, 1988; Froebrich & Rowles, 2010; Lada et al., 2010; Kennicutt & Evans, 2012). More than 10 Myr are required for a GMC to accumulate a significant amount of molecular gas (Heyer & Dame, 2015), a fraction of which is subsequently converted into stars. This phenomenon is quantified by the star formation efficiency parameter ϵ_* (Dobbs et al., 2014) which is typically a few percent per free-fall time (Evans et al., 2009; Murray, 2011; Chevance et al., 2020). This implies that only a small fraction of the total GMC mass will be converted into stars.

GMCs live between 10 and 50 Myr (see e.g. Engargiola et al. 2003; Dobbs & Pringle 2013; Corbelli et al. 2017; Schinnerer et al. 2019; Chevance et al. 2020), albeit some studies have argued that lifetimes of 100 Myr could be necessary due to the presence of GMCs in interarms regions (see e.g. Scoville & Hersh 1979; Scoville & Wilson 2004; Koda et al. 2009). Instead of star formation converting the entire mass of GMCs into stars, they are believed to be disrupted through turbulent flows (which typically occurs only when they have a low self-gravity), or stellar feedback (Krumholz, 2014; Dale, 2015; Chevance et al., 2020, 2022). The latter will be further discussed in Section 1.3.

GMCs are subject to various forces throughout their lifetime, mainly turbulence, gravity, magnetic fields and thermal pressure (McKee & Ostriker, 2007). Gravitational forces from the galactic disk (towards the midplane) act strongly upon molecular clouds whilst self-gravity acquires importance solely at their densest regions (Girichidis et al., 2020). A GMC resists gravitational collapse thanks to the combined effects of thermal pressure (although it loses its efficiency when the GMC undergoes cooling, see e.g. Girichidis et al. 2020) and magnetic fields (Fletcher et al., 2011; Beck, 2015). This is physically expressed by the Jeans length, the critical scale at which gravity is held in equilibrium, described as:

$$\lambda_J = \sqrt{\frac{\pi\, c_s^2}{G\, \rho_0}} \tag{1.1}$$

where a uniform cloud of density ρ_0 is considered, while G and c_s respectively stand for the gravitational constant and the sound speed (Jeans, 1902; Lequeux, 2005; Girichidis et al., 2020).

Eq. (1.1) sets the balance between the cloud self-gravity, engendering a contraction at a rate of $\frac{1}{G\, \rho_0}$, and the thermal pressure, resisting the contraction at a rate of $\frac{R}{c_s}$, where R is the cloud radius. When the latter term becomes smaller than the former, gravitation starts to dominate over thermal pressure and causes the cloud to collapse (Jeans, 1902; Lequeux, 2005; Girichidis et al., 2020). Therefore, when $L > \lambda_J$ (where L is the cloud length), the region collapses due to gravity.

This balance is also expressed by the Jeans mass; representing the minimum mass that can resist collapse:

$$M_J = \frac{4}{3}\,\rho_0\,\frac{\lambda_J^3}{8} \tag{1.2}$$

Thus, Eq. (1.2) corresponds to the mass that a volume of radius $\frac{\lambda_J}{2}$ would encompass.

1.1.2 Turbulence

GMCs are dynamically active structures (Heitsch et al., 2006) whose internal velocity dispersion can reach a few km s^{-1} (Larson, 1981; Heyer & Dame, 2015). It is therefore not surprising that they experience turbulence (McKee & Ostriker, 2007; Dobbs & Pringle, 2013) which, among other phenomena, shape their internal structure (Chevance et al., 2022). Turbulence occurs when the velocity of fluid flows have temporal and spatial fluctuations (Girichidis et al., 2020). In molecular clouds, its drivers range from the gravitational instabilities of the disk (whether global or local), to magnetorotational instability and stellar feedback (McKee & Ostriker, 2007; Miville-Deschênes et al., 2017).

Turbulence translates into an intrinsic velocity dispersion within molecular clouds. It therefore sets the initial kinematics of the young stars and is even likely to dominate their kinematics on large (10 - 100 pc) scales. The velocity dispersion within molecular clouds was observed by Larson (1981) who established a power-law relationship between the size of molecular clouds and their linewidth (or 1D velocity dispersion), expressed as:

$$\sigma_v\,(\mathrm{km\,s^{-1}}) = 1.1\,L(\mathrm{pc})^{0.38} \tag{1.3}$$

where σ_v is the intrinsic 1D velocity dispersion (a combination of large-scale, small-scale and thermal motions) and L is again defined as the length of the molecular cloud (Larson, 1981). Subsequent studies confirmed and refined this relation (Solomon et al., 1987; Bolatto et al., 2008; Miville-Deschênes et al., 2017), whereas others proposed a more complex relationship (Heyer & Dame, 2015; Cen, 2021). Most analyses agree on an index ∼0.5, such that:

$$\sigma_v\,(\mathrm{km\,s^{-1}}) = \sigma_0\, L(\mathrm{pc})^{0.5} \tag{1.4}$$

where σ_0 is a constant whose value changes depending on the study.

The original index from Larson (1981) is consistent with an incompressible turbulence interpretation, where thermal speed dominates over all velocities, in line with the classical view from Kolmogorov (1941). On the other hand, the most common index of 0.5 first formulated by Solomon et al. (1987) is consistent with a compressible turbulence where large-scale velocities are supersonic, a more realistic case for molecular clouds (Brunt & Mac Low, 2004; McKee & Ostriker, 2007).

1.1.3 Internal structure of molecular clouds

As mentioned in Section 1.1.1, GMCs are characterized by a filamentary structure (see Fig. 1.1). Within them prevail a supersonic turbulence: turbulent flows converge and create sheets of gas, where filaments appear at their intersections (McKee & Ostriker, 2007).

Ostriker (1964) introduced these filaments as hydrostatic, isolated and isothermal cylinders whose self-graviting gas is in equilibrium between thermal pressure and gravity. This model states that filaments fragment only when a critical mass is reached, but some observations failed to observe the mass threshold where fragmentation is avoided (see e.g. Fischera & Martin 2012). Models from Gómez & Vázquez-Semadeni (2014) and Chira et al. (2018) rather presented filaments as parts of hierarchical collapsing clouds. Furthermore, filaments are also subject to fragmentation due to accretion, turbulence and magnetism (see e.g. Lee et al. 1999; Hartmann 2002). As such, they live for a short time, and recent models have showed that their fragmentation is clustered (Chira et al., 2018), in contrast with the periodical fragmentation model (e.g. Jackson et al. 2010) that would actually occur only for solely compressive and supersonic turbulent motions (Enoch et al., 2006; Gutermuth et al., 2009; Seifried & Walch, 2015;

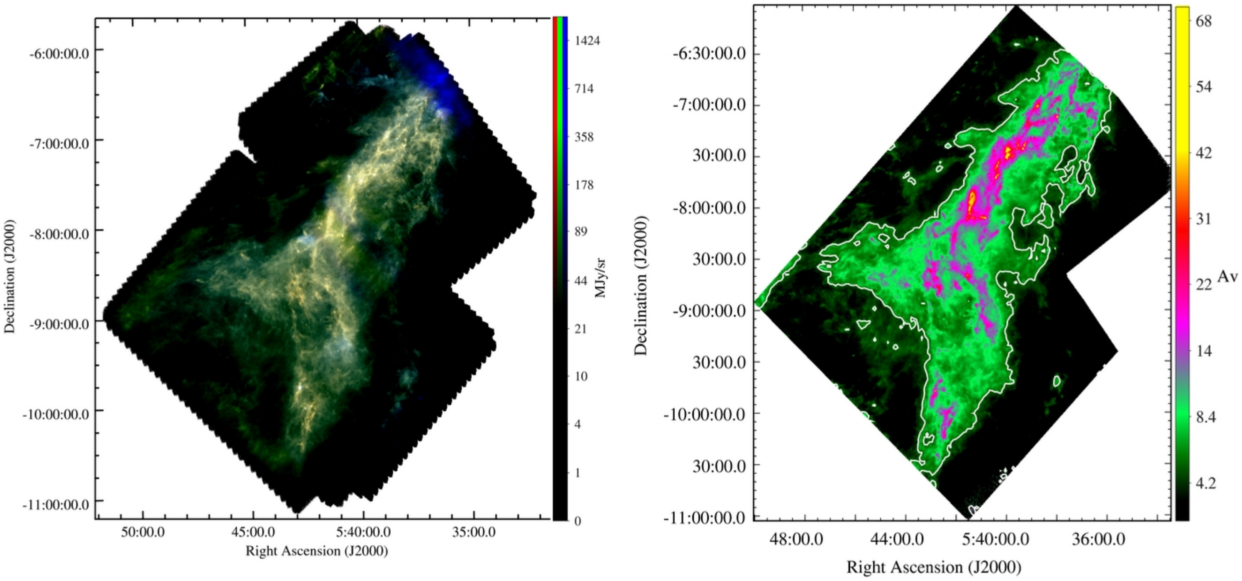


Figure 1.1: Map of the molecular cloud L 1641 in Orion A. The filamentary structure is highlighted in both the left panel (RGB map from the Herschel telescope) and the right panel (extinction map). This figure was taken from Polychroni et al. (2013).

Clarke et al., 2017).

GMCs are composed of overdensities known as star-forming clumps and pre-stellar cores. While individual stars and binaries arise from the gravitationally bound cores, stellar clusters originate from massive clumps, which are thought to be gravitationally bound (McKee & Ostriker, 2007; Ward-Thompson et al., 2007; Girichidis et al., 2020).

Star-forming clumps contain the bulk of a GMC's mass (Bertoldi & McKee, 1992), with masses typically ranging from Jupiter masses to several thousands of solar masses (Heithausen et al., 1998; Urquhart et al., 2018), temperatures of a few tens of K (see e.g. Tang et al. 2021) and size of ∼1 pc (Urquhart et al., 2018). Whether these clumps follow the size-linewidth relationship introduced in Section 1.1.2 remains open to debate (Falgarone et al., 1992; Plume et al., 1997; Heyer et al., 2001). Pre-stellar cores, on the other hand, reach densities higher than 10^5 cm^{-3}, temperatures lower than 10 K, and scales up to a few pc with a prolate or oblate shape (Basu, 2000; McKee &

Williams, 1997; Chacón-Tanarro et al., 2019). Benson & Myers (1989) measured a 1D velocity dispersion of 0.11 km s^{-1} for starless cores. Their lifetime ranges from 0.1 - 10 Myr depending on the effects of turbulence, clumping and outflows (Langer et al., 2000; Goldsmith & Li, 2005; Glover & Mac Low, 2007).

1.1.4 Protostar formation and accretion

GMCs are doomed to collapse. This happens when they accumulate too much gas and exceed the Jeans mass (see Eq. (1.2)), such that the cloud cannot resist gravity anymore (Prialnik, 2000). Alternatively, clouds can collapse because of a collision (from another cloud or even another galaxy), a nearby supernova explosion or other mechanisms able to compress them. This latter scenario is called the triggered star formation or 'compress and collapse' method (Elmegreen & Lada, 1977; Elmegreen, 2011; Li et al., 2014).

The first step of the classical collapse process is the formation of a protostar from the gravitational collapse of a pre-stellar core (Larson, 2003). When this happens, these cores will concentrate in a region with a radius equal to the Jeans length, and with a density profile $\rho \propto r^{-2}$ (Larson, 1969). Waves of collapse subsequently progress inside the region and make it unstable, with $\dot{m}_{\rm in} \propto \frac{c_s^3}{G}$ (Larson, 2003; McKee & Ostriker, 2007), where $\dot{m}_{\rm in}$ stands for the infall rate. Turbulent pressure may also engender gravitational collapse at more local scales (Bonazzola et al., 1987, 1992) and, a core collapse may start once turbulence dissipates (as turbulence can support GMCs, Myers & Lazarian 1998).

When rotation is included in the model, the gas collapse will occur at a constant angular momentum, resulting in disk formation (McKee & Ostriker, 2007). The slow-rotating core with constant density will be characterized by a singular surface density profile (Norman & Silk, 1980; Narita et al., 1984). From this configuration will be created a thin disk nearly at equilibrium, whose infall rate will be inferior to the non-rotating case (Saigo & Hanawa, 1998). An opposite scenario of slow collapse takes place when quasi-equilibrium and centrally concentrated state is attained prior to the collapse. This will result in a non-Keplerian disk, with an inner disk close to equilibrium and an outer disk undergoing dynamical contraction (Stahler et al., 1994).

Magnetic fields can favour core contraction as well where they are weak: this is the ‘supercritical core’ case. Alternatively, when strong magnetic fields resist gravitational collapse, this is a subcritical core (McKee & Ostriker, 2007). Ambipolar diffusion can lead to a transition between these two states (Fiedler & Mouschovias, 1993). In the model from Krasnopolsky & Königl (2002), including rotation and magnetism, magnetic braking is able to withold accretion on the protostar.

The first accreting model for the resulting protostar is the Bondi-Hoyle accretion (or core accretion), referring to the accretion done by a moving source (Hoyle & Lyttleton, 1939; Bondi, 1952). Ignoring the magnetic field, tidal gravitational field and gas self-gravity, the accretion rate can be written as (Bonnell et al., 2001; McKee & Ostriker, 2007):

$$\dot{M}_{BH} = \frac{4\,\pi\,\rho\,G^2\,m_*^2\,\phi_{BH}}{(1+\mathcal{M}_0^2)^{3/2}\,c_s^3} \tag{1.5}$$

where is m_* the star mass, $\mathcal{M}_0$ the Mach number and ϕ_{BH} is a number, close to unity, whose value varies depending on the instabilities within the flow (Ruffert & Arnett, 1994).

Another scenario is the competitive accretion model. It suggests that as stars begin to form groups and clusters, the resulting gravitational potential will enhance the flow of gas towards the proto-cluster of stars. This causes a preferential accretion of mass onto the stars at the centre of the potential (Zinnecker, 1982; Bonnell et al., 1998, 2001).

The newly-born stars can be found in multiple systems, especially binaries, with a multiplicity fraction higher for more massive stars (Duquennoy & Mayor, 1991; Reid & Gizis, 1997; Lada, 2006; Preibisch et al., 2001; Sana & Evans, 2011; Sana, 2017). These multiple systems stem from the fragmentation of pre-stellar cores (see e.g. Bodenheimer et al. 2000), with the bulk of the cores producing 2 to 3 fragments (Goodwin & Kroupa, 2005).

After their growth through core or competitive accretion, protostars become pre-

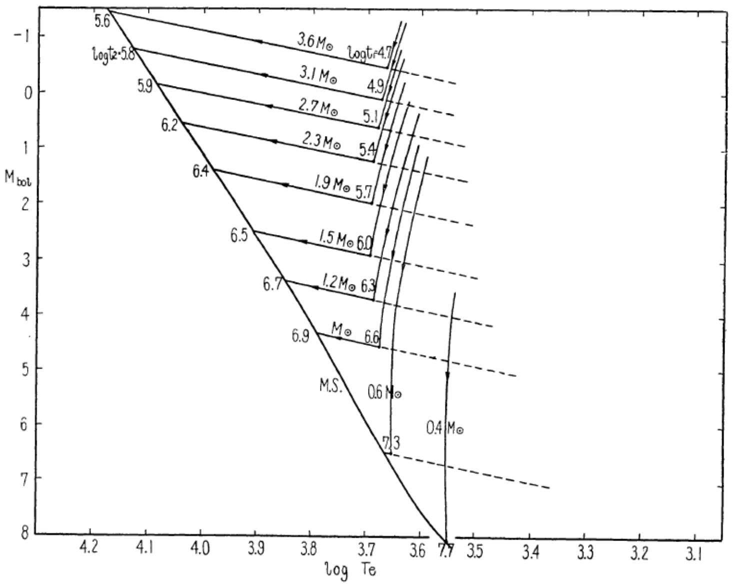


Figure 1.2: Hayashi tracks for stars of various initial masses, where t_1 and t_2 respectively denote the age a star begins following the Hayashi track and the age it reaches the main-sequence. The ages are in logarithmic scale with their units in years. This plot was taken from Hayashi (1961).

main-sequence stars (PMS). The contraction timescale for low-mass PMS stars is a lot longer than it is for high-mass stars. Low and intermediate-mass stars (typically $<$ 3 $M_{\odot}$) will follow the so-called Hayashi tracks (illustrated in Fig. 1.2), with a path depending on their mass and chemical composition. They will become T-Tauri stars, a phase during which their contraction will slow down and their luminosity decreases whilst keeping the same temperature. They eventually join the main-sequence once nuclear fusion is initiated in their core (Henyey et al., 1955; Hayashi, 1961; Prialnik, 2000; Girichidis et al., 2020).

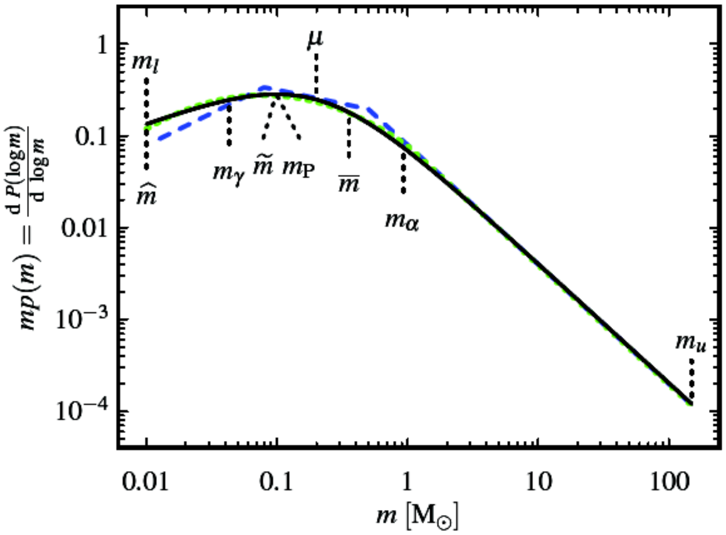


Figure 1.3: Probability density function of the IMF from Maschberger (2013), compared with the IMF from Chabrier (2003) in green and the IMF from Kroupa (2001, 2002) in blue. In this model, the lower and upper mass limits are respectively equal to 0.01 $M_\odot$ and 150 $M_\odot$, and the peak and median masses are close (around 0.2 $M_\odot$). This plot was taken from Maschberger (2013).

1.1.5 The Initial Mass Function (IMF) and massive star formation

How new-born stars are distributed in mass is described by the initial mass function (IMF), first defined by Salpeter (1955) as a power-law relating the number of stars as a function of their initial mass:

$$\frac{dN}{dM} \propto M^{-2.3} \tag{1.6}$$

The IMF peaks around 0.2 - 0.5 $M_\odot$ (see Fig. 1.3) and has been found to have a universal log-normal distribution (see e.g. Kroupa 2001, 2002; Chabrier 2003; Maschberger 2013). It thus favours the formation of low-mass stars. In particular, the upper-mass IMF is characterized by a power-law with a slope of -2.35, which is constant regardless of the studied massive star population (Kroupa, 2002), and can be explained by the competitive accretion model (Zinnecker & Yorke, 2007).

The smaller fraction of massive stars, as illustrated by the IMF (Fig. 1.3), stems from the more constrained initial conditions for their formation to take place. Indeed, high-mass stars must form inside very dense clumps, whose H_2 column density equates 10^{23} - 10^{24} cm^{-2} (Zinnecker & Yorke, 2007). During the compression phase, pressure and turbulence should be abundant enough to ensure there is a sufficient amount of material at the cores of GMCs. Only in these conditions can high-mass star formation occur (McKee & Tan, 2003).

Important to massive star formation is the fact that they initiate hydrogen burning and have strong stellar winds that begin whilst the forming star is still accreting matter (Kudritzki, 2002). Thus, their growth can stop when there is not remaining material to capture, but it can be also halted by stellar wind opposing accretion, hence the complexity of high-mass star formation (Zinnecker & Yorke, 2007). For the core accretion model, a massive star can only be produced if the accretion rate is superior to the outflow rate. This condition is fulfilled when the radiative acceleration of the accreting core is more important than the acceleration due to gravity (Zinnecker & Yorke, 2007).

It is however thought that high-mass stars preferentially form through competitive accretion. In this model, they are rare because accretion occurs at each early stage of their formation, and thereby requires a more favourable accretion than low-mass stars (see e.g. Bonnell et al. 2001). Krumholz et al. (2005) opposed this theory due to the difficulty of increasing protostellar masses in a turbulent medium. This would prevent the Bondi-Hoyle accretion to take place due to the large relative velocities between the turbulent gas and the high-mass protostar. Furthermore, it has been proposed that radiative feedback can act against Bondi-Hoyle accretion too (Edgar & Clarke, 2004). But Bonnell & Bate (2006) argued that the scale of supersonic turbulence as predicted by Larson (1981) results in an actual low relative velocity, allowing a tremendous growth of the protostar and therefore the formation of massive stars.

This view is consistent for early B-type stars for which surrounding disks have been observed (Zhang, 2005), but there is a lack of evidence for progenitors of O-type stars with masses higher than 20 $M_\odot$ (Cesaroni et al., 2007). For these stars, it has been proposed instead that either their disk has been eroded during their stellar growth (Zinnecker & Yorke, 2007), or they form through protostellar or stellar collisions (Bonnell et al., 1998; Stahler et al., 2000).

1.2 OB stars

O and B-type stars correspond to the hottest and most massive of all stars, being early-type according to the Harvard spectral classification scheme (Payne, 1925). Most recent models show that these stars have an effective surface temperature greater than 10,000 K, an initial stellar mass greater than 2.5 $M_\odot$, and luminosity from several hundreds to several hundreds of thousands solar luminosities (Martins et al., 2005; Pecaut & Mamajek, 2013). B-type stars account for only $\sim$0.1 % of all main-sequence stars while O-type stars are even rarer with an estimated fraction lower than 0.0001 % (Ledrew, 2001). This stems from both the preferential formation of low-mass dictated by the IMF (see Section 1.1.5), but also from their short lifetimes. OB stars lifetimes typically span from a few million years to a few tens of million years, much shorter than their low-mass counterparts (Crowther, 2012).

It is frequent to refer to O and early B-type stars (B2 and earlier) as massive stars (though the exact definition can vary between studies). These are stars whose mass is high enough for a collapsing core to form at the end of their life (Zinnecker & Yorke, 2007; Langer, 2012). The exact minimum initial mass value for this to happen remains up to debate, but studies have narrowed down the interval to 8-12 $M_\odot$, depending on metallicity, with a mean around 9 $M_\odot$ (Poelarends et al., 2008; Jones et al., 2013; Doherty et al., 2017).

1.2.1 Multiplicity amongst massive stars

As mentioned in Section 1.1.4, stars are often found in multiple systems, with a multiplicity factor larger for massive stars (see e.g. Preibisch et al. 2001; Sana & Evans

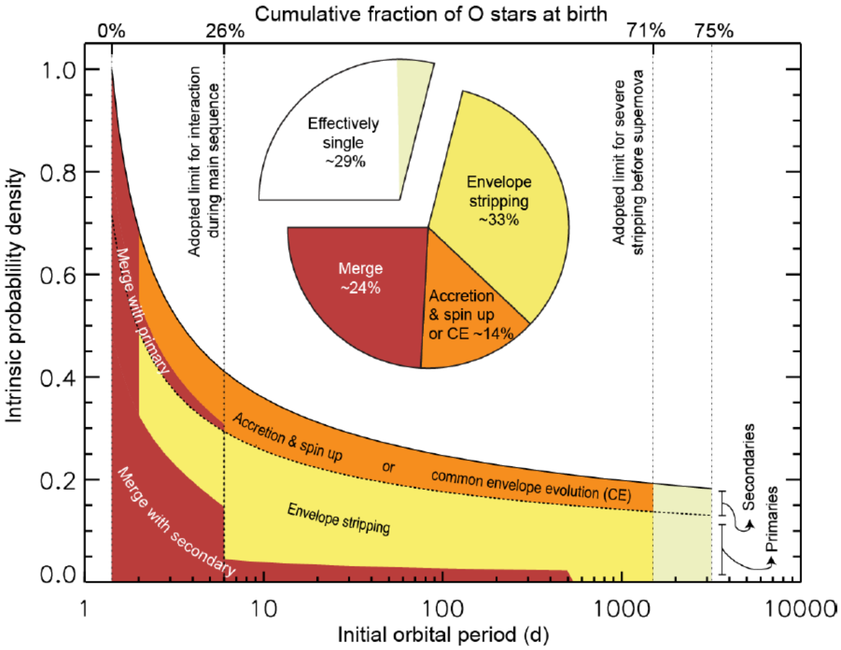


Figure 1.4: Probability density of the diffrent interactions for O-type binary systems as a function of their orbital period. This figure was taken from Sana et al. (2012).

2011; Sana 2017). A close binary companion will influence the evolution of massive stars (see e.g. Claeys et al. 2011; Farrell et al. 2022), through interactions that can affect the mass and evolution of a star (Eldridge et al., 2017). More than 70 % of massive stars will exchange matter with a companion in a binary (or multiple) system during their life, often leading to envelope striping or a common envelope (Fig. 1.4). This even engenders a binary merger one third of the time (Sana et al., 2012). When the companion is compact (neutron star or black hole), it can capture the stellar wind released by the massive star, thus emitting X-rays: this is a high-mass X-ray binary (HMXB, Tauris & van den Heuvel 2006). It is therefore unlikely that OB stars reach unperturbed their predicted end of life (Sana et al., 2012).

1.2.2 Post main-sequence evolution and supernovae

As soon as massive stars have consumed the hydrogen in their core, they leave the main-sequence and start burning helium. This is the red supergiant (RSG) phase, during which the outer layers of the stars undergo expansion and cooling due to the increased energy output from the core. More massive stars ($> 20\ M_\odot$) can turn into blue supergiants (BSGs) where these same outer layers contract and become hotter, whereas stars heavier than $\sim$45 $M_\odot$ evolve into luminous blue variables (LBVs) (Maeder & Meynet, 2010; Sander et al., 2012). The next step for massive stars is the Wolf-Rayet phase (WR), characterized by intense stellar winds (up to a few thousands of km s^{-1}), significant mass loss rates (up to $10^4\ M_\odot\ \mathrm{yr}^{-1}$), and broad emission lines in their spectra (van der Hucht, 2001; Crowther, 2007; Langer, 2012). Massive stars at different stages of their evolution are shown in a HR diagram in Fig. 1.5.

Following the gravitational core collapse of an evolved massive star, a supernova explosion occurs (type II and type Ib/c, Burrows & Vartanyan 2021). An intense radiation often follows the supernova explosion (particularly the most luminous ones) under the form of a long gamma-ray burst (Woosley & Bloom, 2006; Nieva & Przybilla, 2014). A supernova explosion releases of the order of about 10^{51} erg of energy, such that supernova explosions account for 80 % of the kinetic energy injection into the ISM (de Geus, 1991). Supernovae feed the ISM with heavy elements (such as carbon and oxygen) subsequently used to form new stellar systems, and are consequently important for the chemical evolution of galaxies (de Rossi et al., 2010). The nature of the remaining compact core depends on the initial mass, metallicity, and evolution of the star, i.e. neutron stars ($> 8\ M_\odot$) or stellar black holes ($> 20\ M_\odot$), although these initial mass thresholds apply only for solar metallicities and will be smaller at lower metallicities. On the other hand, the expelled outer layers will shine as a supernova remnant for a few tens of thousand of years (Heger et al., 2003; Dubner & Giacani, 2015).

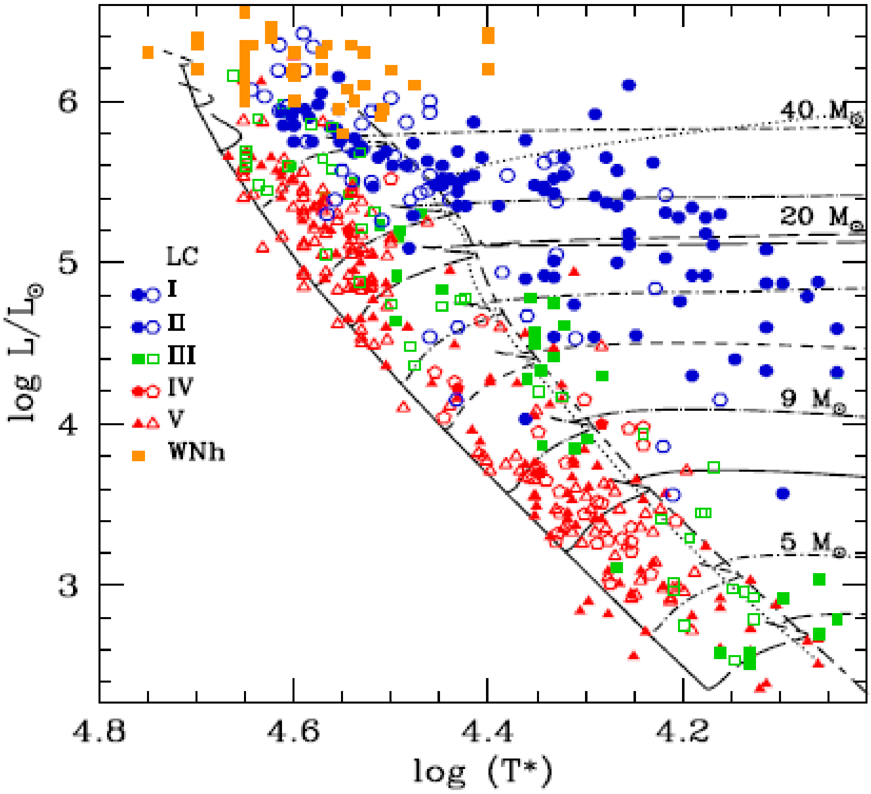


Figure 1.5: Evolutionary tracks from Ekström et al. (2012) plotted alongside observed OB stars (see Martins & Palacios 2013 and references therein), whose stellar parameters have been calibrated from Martins et al. (2005).

1.3 Feedback and the end of star formation

Stellar feedback is essential to comprehend since the newly formed stars (especially massive ones) will drastically affect their environment (Krumholz, 2014; Dale, 2015). Numerical simulations have shown that feedback processes bring high amounts of momentum and energy into the ISM (Geen et al., 2015; Kim & Ostriker, 2015; Walch & Naab, 2015). They include a wide range of phenomena acting on different spatial and time scales, but the main processes are outlined as follows:

1. *Protostellar outflows*: these outflows inject energy and angular momentum into

their surroundings, directly affecting the star-forming regions in which they occur (Norman & Silk, 1980; McKee, 1989; Shu et al., 1999). They happen when young stars of all masses undergo collimated mass-loss through the effect of rotating gas energizing magnetic field lines (Bai et al., 2016; Bally, 2016). In young embedded stars, the molecular outflows are typically slow and cold, whereas in older and less embedded YSOs, they can reach velocities up to 200 km s^{-1} (highly collimated optical jets, Reipurth & Bally 2001). They play an important role in star formation and turbulence, particularly on cluster-forming clumps where they are able to maintain stellar formation for longer, impeding the global collapse of the parent cloud during massive star formation (see e.g. Matzner 2007; Nakamura & Li 2007; Wang et al. 2010; Hansen et al. 2012; Offner & Chaban 2017).

2. *Radiation pressure*: the radiation field produced by stars can inject momentum and energy into the encompassing gas. Anywhere massive stars are present, radiation pressure can be the dominant feedback mechanism (Krumholz, 2014). This process may take place during star formation, when heat converted from the kinetic energy of the falling gas is radiated away (Stahler et al., 1980; Offner & McKee, 2011). Radiation can also occur when the protostar evolves along the Hayashi track, with energy generated through internal processes such as gravitational contraction and deuterium burning (Girichidis et al., 2020). This radiation then heats the surrounding gas, which could stabilize the accreting disk, and thus strongly influences the resulting IMF (Offner et al., 2014). Radiation pressure may also blow away some of the surrounding molecular cloud and thus affect star formation that way (Murray et al., 2010).

3. *Stellar winds*: while every star loses some mass through high-speed winds, this phenomenon is stronger for massive stars, whose winds inject large amounts of momentum and energy into the surrounding molecular cloud (Churchwell et al., 2006). Only stars with $T_{\rm eff} > 25,000$ K will produce intense winds (Vink et al., 2000). Stellar winds are described by their terminal velocity v_∞ (defined

as the velocity of the wind far from the star) and their mass-loss rate $\dot{M}_w$. In particular, early-type stars (OBA), giants and supergiants generate a radiation pressure driven wind (i.e. carried by optically thick lines), with $v_\infty \sim 2000$ km s^{-1} and $\dot{M}$ from 10^{-9} to 10^{-4} $M_\odot$ yr^{-1} (Vink et al., 2001).

4. *Photoionization*: together with their wind, massive stars release large quantities of ionizing photons and thereby engender ionized bubbles that will become HII regions (Anderson et al., 2009). Their constant temperatures of ~10,000 K represent the outcome of the balance between cooling and heating in an ionized gas (Osterbrock & Ferland, 2006). These bubbles have a huge pressure that causes the region to expand at a rate of the magnitude of the sound speed (Spitzer, 1978). Photoionization is thought to be the dominant feedback mechanisms within the lifetime of GMCs, except for the most massive, luminous and earliest evolutionary stages clusters, where radiation pressure exceeds the thermal pressure and dominates instead (Chevance et al., 2022).

5. *Supernova explosions*: as mentioned in Section 1.2, these explosive feedback events release a tremendous amount of energy and momentum into the ISM. They occur between a few Myr and a few tens of Myr after the beginning of star formation (Matzner, 2002; Leitherer et al., 2014). Even if they do not represent the most important feedback process in low-mass clusters, where the crossing time will be achieved before the first supernova occurs (Tan et al., 2006; Krumholz & Matzner, 2009; Fall et al., 2010), they will have greater effects on the scale of more massive clusters, spiral arms or entire galaxies (Krumholz, 2014).

All the above mechanisms can cause the expulsion of gas from the GMCs previously used to form stars (Whitworth, 1979; Krumholz et al., 2019). Consequently, stellar feedback is able to disrupt the molecular cloud the stars emerged from (Dobbs & Pringle, 2013). Chevance et al. (2020) estimated the timescale for important feedback was in the range of 1-5 Myr, implying that these processes rule over other destruction causes for GMCs. This also results in a short duration for star formation, with star formation brutally ended through the effects of feedback.

1.4 Clusters and associations

In Lada & Lada (2003), star clusters are defined as groups of physically related stars with a density higher than 1 $M_\odot$ pc^{-3}, large enough to resist tidal disruption and the crossing of interstellar clouds (Bok, 1934; Spitzer, 1958). Lada & Lada (2003) stated that clusters should contain at least 35 stars, such that their evaporation time, i.e. the time-scale over which all members are ejected through internal stellar encounters, exceeds 100 Myr (Binney & Tremaine, 1987; Adams & Myers, 2001).

1.4.1 Types of stellar groups

The main types of stellar groups are defined below:

- *Embedded clusters:* They are clusters partially or entirely embedded (hidden) within giant molecular clouds, thereby requiring IR observations to analyse them (Lada & Lada, 2003). This means that embedded clusters are usually very young ($<$ few Myr), sometimes referred to as 'protoclusters', and may represent the initial level of clustering of recently-formed stars. Their internal structure can be either centrally-condensed (where the distribution of their surface density is highly concentrated, characterized by smooth radial profiles) or hierarchical (where this distribution is instead characterized by multiple peaks, with evidence of important structure across large scales), inherited from their formation process (Lada & Lada, 2003). Evidence for mass segregation within embedded clusters have been mixed, though some observations showed more massive stars preferentially located near their centre (e.g. Hillenbrand & Hartmann 1998; Nürnberger & Petr-Gotzens 2002).

- *Open clusters:* They are gravitationally bound clusters, i.e. the sum of their potential and kinetic energy is negative. This allows them to reach lifetimes of 100 Myr at least (Lada & Lada, 2003). Several hundreds to several thousands of stars can be contained within a single open cluster, whose core generally extends to a few pc and their corona to a few tens of pc (Nilakshi et al., 2002). There have been evidence of mass segregation in open clusters, engendered by mutual exchange of energy and momentum between cluster members, resulting in equipartition (Elmegreen et al., 2000). Open clusters can be found in spiral and irregular galaxies where star formation is active (Janes & Phelps, 1994; Hunter, 1997), and as such, constitute tracers of both the galactic spiral arms and recent star formation (Lada & Lada, 2003). They are thought to be one of the products of the star formation process and evolve from embedded clusters.

- *Globular clusters:* They are also bound but tend to be older (up to several Gyr old) and include more members (up to a few million stars) than open clusters (Gratton et al., 2019). Their density can reach 1000 stars per cubic parsec in their core (Abell et al., 1987). Observations reveal a clear lack of dust and gas within them, implying that the star formation has long been halted (Bastian & Strader, 2014). Globular clusters can be found in every type of galaxy (see e.g. Harris 1991); in spiral galaxies like the Milky Way, they are mostly located in the halo (Yoon & Lee, 2002). Their population of low-metallicity old stars make them ideal probes to study previous generations of stars (Lada & Lada, 2003; Gratton et al., 2019). They are thought to have formed during an early period of star formation in the Milky Way, before the galaxy formed a disk shape (hence their presence in the halo, see Harris 1976). Being in the halo and away from dense molecular clouds, spiral arms and the Galactic disk, they are less prone to tidal disruption and thus have been able to survive for many Gyrs (Bose et al., 2018).

- *OB associations:* Stellar associations have a lower density than star clusters (Blaauw, 1964). Specifically, OB associations are gravitationally unbound groups of stars including many OB stars, whose low-density ($< 0.1\ M_\odot\ \mathrm{pc}^{-3}$) make them vulnerable against galactic tidal forces (Bok, 1934; Ambartsumian, 1947). Their spatial and kinematic concentration in spite of this implies that they must be young: as such, they constitute a transition phase between the higher density and prominent clustering present in star-forming regions and the field star population (Wright, 2020). As the main subject of this thesis, OB associations are discussed in more details below.

1.4.2 Simulations of star, star cluster and OB association formation

Hydrodynamical simulations of star cluster formation are able to reproduce stellar properties similar to those observed (see e.g. Bate et al. 2003; Bate 2009). Some simulations of massive star formation support the monolithic star formation scenario (e.g., Kuiper et al. 2011; Klassen et al. 2016). By contrast, besides emphasizing the likely role of competitive accretion to explain the emergence of massive stars (Bonnell et al., 2001; Bonnell & Bate, 2006), some simulations show that turbulent molecular clouds tend to undergo hierarchical fragmentation resulting in small subclusters. These subclusters would subsequently merge to form bigger clusters (Bonnell et al., 2003; Fujii et al., 2012; Chen et al., 2021), although more recent simulations unveiled that clusters can also split (Dobbs et al., 2022).

The evolution of the formed clusters can be modelled through N-body simulations. It was initially thought that this evolution was dictated by two- or few-body relaxation (see e.g. Spitzer 1987; Portegies Zwart et al. 2010). However, N-body simulations simply assuming a Plummer-profile (i.e. a flat density profile) cannot well capture the observed substructure within star clusters, because they do not include the stellar ejections, binary captures and dissolutions that characterize their rapid dynamical evolution in the earliest stages of their life (Goodwin & Whitworth, 2004). Furthermore, dynamical mass segregation cannot be explained by the energy equipartition resulting from two-body relaxation (Parker et al., 2016). Instead it has been proposed that fractal initial conditions are crucial to produce the clumpiness observed in star clusters (Parker et al., 2014; Daffern-Powell & Parker, 2020), with two-body relaxation acting on greater timescales (Ballone et al., 2020). Recent hydrodynamical simulations from Cournoyer-Cloutier et al. (2023) have shown that the gravitational potential from the star-forming region rules over two-body relaxation to affect the evolution of young and low-mass ($\leq 1000\,M_\odot$) embedded clusters.

N-body simulations also focused on the effect of residual gas expulsion in star clusters (see e.g. Goodwin 1997), whose effect would be to destroy the majority of star clusters, or at least would eject most of their members. If the gas removal occurs on a slower timescale than the initial crossing time, this may explain the anisotropic velocity dispersions (Baumgardt & Kroupa, 2007). Such phenomenon is thought to be engendered by stellar feedback, acting as a regulator of star formation (Dobbs & Pringle, 2013; Dale, 2015), consistent with the *clustered* star formation model where OB associations arise as the unbound remnants of embedded clusters after residual gas expulsion (Lada & Lada, 2003). In particular, simulations from Dobbs et al. (2022) with a strong initial dynamics (i.e. less dominated by gravity) and small density were able to produce stellar groups with properties similar to OB associations, highlighting the importance of feedback.

1.4.3 Properties of OB associations

OB associations typically extend from several tens to several hundreds of parsecs (see e.g. Blaauw 1964; Blaha & Humphreys 1989; Garmany & Stencel 1992; Mel'Nik & Efremov 1995; Gouliermis 2018a), whereas their mass ranges from thousands to tens of thousands of solar masses (Wright, 2020). Their ages span usually between 1 Myr and 20 Myr (Blaauw, 1964), between their emergence from the embedded environment of their primordial molecular cloud and the disappearance of their more massive members (Wright, 2020). It is however difficult to assign a single age to individual OB associations since they have been shown to exhibit a significant level of internal substructure, with subgroups of various kinematics and ages (see e.g. Ambartsumian 1949;

Blaauw 1964; Garmany & Stencel 1992; Wright et al. 2014; Pecaut & Mamajek 2016; Kounkel et al. 2018; Cantat-Gaudin et al. 2019). Such substructures likely arise from the kinematic substructure present during their formation process (Elmegreen, 2008; Wright et al., 2016). Velocity dispersions from OB associations often equate to several km s^{-1} (Mel'nik & Dambis, 2017; Ward & Kruijssen, 2018; Melnik & Dambis, 2020), and exhibit non-negligible anisotropy along different axes (see e.g. Wright et al. 2016; Wright & Mamajek 2018; Cantat-Gaudin et al. 2019). This asymmetry could originate from the formation process itself (Wright et al., 2019) or their later evolution, such as the effects of tidal forces, making them unbound and engendering an asymmetric expansion (Elmegreen & Hunter, 2010; Kruijssen, 2011).

Most OB stars can be found in OB associations (McKee & Williams, 1997) and as such are easy targets for identifying them (Morgan et al., 1953). Their bright and young members are not only useful to analyse the IMF (see e.g. Massey et al. 1995; Wright et al. 2015), but can also help to understand early stellar evolution, the evolution of proto-planetary disks and, potentially, planet formation. Kruijssen et al. (2012) estimated in their simulations that more than two thirds of stars form in associations.

Another impact of prominent members of OB associations is their feedback processes (Dove & Shull 1994; Shull & Saken 1995, see Section 1.3). OB associations are very efficient at sweeping the gas from their primordial molecular clouds (Blaauw, 1964). They can undergo several supernova explosions within their lifetime that will generate hot, small density and large-scale cavities referred to as super-bubbles (McCray & Snow, 1979; Higdon & Lingenfelter, 2013), which shape the chemo-dynamical evolution of the Galaxy (Wright, 2020). Observations often showed super-bubbles encompassing OB associations (see e.g. Cash et al. 1980; Sahu 1992; Robitaille et al. 2018).

1.4.4 Formation of OB associations and relation between stellar groups

The exact origin of OB associations remains up to debate. The *clustered* model is tied to the embedded clusters. Indeed, it is thought that 70 to 90 % of stars emerging from GMCs originate from dense, embedded and bound clusters (Lada & Lada 2003, see Section 1.4). Embedded clusters however suffer from a high mortality rate, with only 4 to 7 % of them able to evolve into bound open clusters (Lada & Lada, 1991). Only the most massive of them ($> 500\ M_\odot$) can resist against the dispersion of the residual gas caused by feedback from new-born stars (Hills, 1980; Lada et al., 1984; Kroupa, 2001). This process is known as *residual gas expulsion*. Embedded clusters unable to turn into bound open clusters will become unbound and therefore expand. They will briefly be observable as a low-density, gravitationally unbound OB associations, and will subsequently become part of the field star population (see e.g. Brown et al. 1997).

Alternatively, in the *hierarchical* model, stars form over a wide range of scales and densities (see e.g. Elmegreen 2008; Kruijssen et al. 2012). Bound clusters then form in the regions of higher density where the star formation efficiency is the highest, potentially helped by mergers between sub-clusters (Kruijssen et al., 2012; Vázquez-Semadeni et al., 2017)). By contrast, OB associations form from the region of lower density and are therefore born gravitationally unbound. This model does not require residual gas expulsion to unbind the star clusters, nor does it necessitate bound GMCs (Heyer et al., 2001).

The reality is likely between these two extremes (Wright, 2020). Detecting signatures of expansion from OB associations would favour the first model, while observations of young stars at various densities would favour the second one (see e.g. Bressert et al. 2010).

1.4.5 Expansion of OB associations

Since their discovery it was generally assumed that OB associations should be expanding due to their youth and unbound state (Ambartsumian, 1947, 1949; Blaauw, 1964). Nonetheless, measuring expansion in OB associations turned out to be complex, with different levels of success.

Wright et al. (2016) failed to detect expansion in Cyg OB2 whilst Ward & Kruijssen (2018) could not find any evidence for expansion in their study of OB associations. Although Wright & Mamajek (2018) were able to measure expansion in the direction of Galactic rotation for Sco-Cen, there was no evidence for a fully coherent 3D pattern expansion as would be predicted under the *clustered* model of OB association formation. Mel'nik & Dambis (2017) and Melnik & Dambis (2020) measured expansion in an handful of their catalogue of OB associations. Kounkel et al. (2018) identified an expansion pattern inside the subgroups of Orion whereas both Cantat-Gaudin & Anders (2020) and Armstrong et al. (2020) discovered something similar inside the Vela-Puppis region.

Measurements of expansion in OB associations have therefore been mixed, although the success rates have been higher with better kinematically-defined subgroups (Wright, 2020). Furthermore, the bulk of OB associations that have been measured to be expanding are expanding anisotropically, against the classical radial expansion pattern predicted by Blaauw (1964).

1.4.6 Catalogues of OB associations

A map of known OB associations within 2 kpc is shown in Fig. 1.6. Humphreys (1978) established a list of Galactic OB associations. To do so, she compiled the spectroscopic observations of known OB stars, and assessed their membership to a given OB association through their galactic coordinates and photometric distances. Their catalogue contains ∼50 OB associations, including up to 20 O-type stars and 60 B-type stars in each association.

Since then, this catalogue has been revisited and expanded several times (Blaha & Humphreys, 1989; Mel'Nik & Efremov, 1995; de Zeeuw et al., 1999; Melnik & Dambis, 2020). de Zeeuw et al. (1999) notably improved the study of OB associations thanks to the proper motions from HIPPARCOS, but significant improvements were limited to the nearest (< 500 pc) systems. The original membership of OB associations was mostly based on the spatial concentration of their brightest stars (Humphreys, 1978), requiring to possibly revisit them in terms of kinematic coherence (Wright et al., 2023).

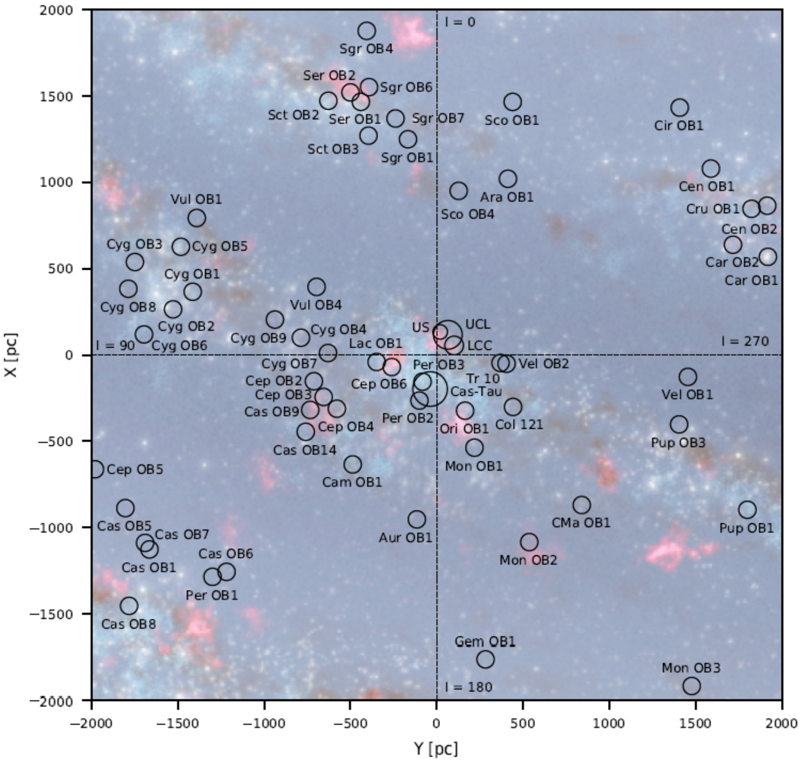


Figure 1.6: Distribution of known OB associations in the X-Y plane (Galactic cartesian coordinates), with their distance originating from *Gaia* DR2 (Gaia Collaboration et al., 2018) The background image corresponds to the Milky Way spiral arms, respectively the Perseus, local and Sagittarius arms, from an artist's depiction. This figure has been taken from Wright (2020).

1.4.7 Extragalactic OB associations

Beyond the Milky Way, OB associations have been observed in the Magellanic Clouds (Hodge & Wright, 1967, 1977), in the Andromeda Galaxy (Hodge, 1981), and the Local Group in general (Hodge et al., 2002). In star-forming disk galaxies, they sit within the thin disks and along the spiral arms, whilst in starbust and irregular galaxies, they serve as complex centres of star formation (Gouliermis, 2018b).

Distinguishing unbound stellar associations from the field population turned out to be easier in extragalactic environments. Contrary to the Galactic OB associations, kinematic information is not required: instead, they stand out as over-densities of bright and young stars (Ivanov, 1987; Gouliermis et al., 2000; Gouliermis, 2018b), although these may be limited to the most compact of extragalactic OB associations. Nonetheless, historically, the definition of an extragalactic OB association showed variations depending on the applied techniques and identification methods, notably due to the difficulty of determining their physical boundaries (Gouliermis, 2011).

The arrival of the Hubble Space Telescope (HST), alongside automated algorithms (see e.g. Battinelli 1991; Wilson 1991), led to improved and more homogeneous methods, sufficient to characterize extragalactic OB associations. Whereas the number density of OB associations has been observed to vary in different galaxies, their size has been found to be similar, with an average of $\sim$80 pc thought to constitute a universal scale for OB associations (Bresolin et al., 1998; Gouliermis et al., 2003). Furthermore, extragalactic OB associations exhibit large structures, shaped like filaments, which are typically located at the edges of HI shells (Gouliermis et al., 2003; Werk et al., 2008).

1.5 Large-scale distribution of OB stars

Star formation occurs primarily in the spiral arms of a galaxy, where molecular clouds are also more abundant (Elmegreen, 2011). OB associations, and by extension OB stars, preferentially occupy the spiral arms, as hinted by Fig. 1.6 (see e.g. Russeil 2003;

Chen et al. 2019b; Pantaleoni González et al. 2021; Zari et al. 2021). Being short-lived, OB stars tend to remain close to their birth environment (Sparke & Gallagher, 2000), and together with GMCs, HII regions and dust distribution, constitute excellent tracers of galactic structure (see e.g. Hou et al. 2009; Vallée 2014; Lallement et al. 2019; Green et al. 2019).

OB associations tend to form larger structures of several hundreds of pc (see e.g. Elmegreen & Efremov 1997). Bouy & Alves (2015) notably showed the existence of single age sequence structures shaped by OB associations in the solar neighbourhood. They referred to these as blue streams and hinted that they could represent the large-scale progression of star formation.

In this thesis I study the populations and properties of OB associations in certain constellations. These regions were chosen as being both regions of current and recent active star formation as well as containing multiple, often poorly-studied and parameterized, OB associations.

1.5.1 The Cygnus region

The Cygnus region harbours active and recent star formation. It includes the massive star-forming region Cygnus X where the historical Cygnus OB associations have been discovered (including the massive and widely analysed Cyg OB2). This region therefore constitutes an ideal first target to study OB associations (Humphreys, 1978; Reipurth & Schneider, 2008).

1.5.1.1 Cygnus X

The main component of interest inside the Cygnus region is Cygnus X, at a distance of ∼1.4 kpc (Rygl et al., 2012). It is considered as one of the most abundant regions of star formation in the Milky Way, with a molecular cloud complex of a total mass of 3-4 $\times\ 10^6\ M_\odot$ (Schneider et al., 2006), thus carrying a high level of extinction (Dickel & Wendker, 1978). The northern part of Cygnus X, at higher latitudes, is characterized by a filamentary structure, whilst the emission at the southern part (lower latitudes) is more diffuse (Schneider et al., 2006, 2007). In this region, thousands of OB stars accompany massive molecular clouds complexes, HII regions, supernova remnants and protostars (Downes & Rinehart, 1966; Comerón & Torra, 2001; Uyanıker et al., 2001; Hora et al., 2008). The Cygnus X region is illustrated in Fig. 1.7.

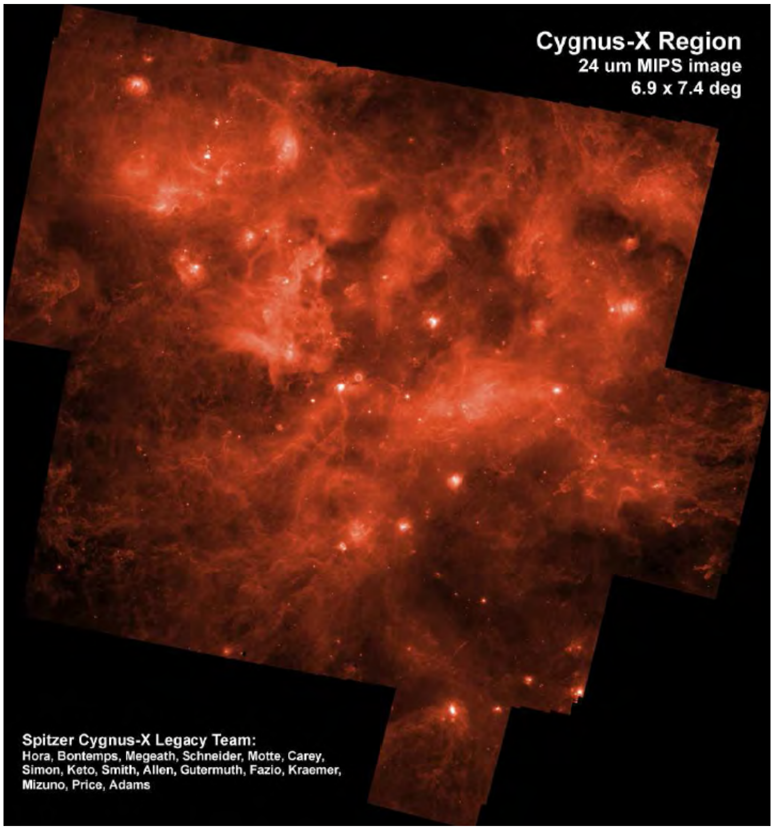
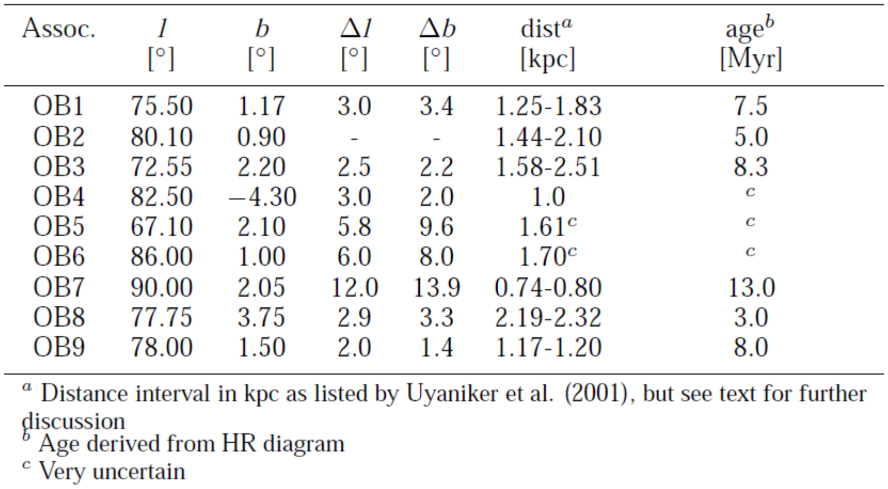


Figure 1.7: Cygnus X as appearing from the Spitzer telescope at 24 μm. This picture was taken from Hora et al. (2008).

Supernova explosions and stellar winds from massive stars create an expanding and hot medium ($\sim 10^6$ K), reaching a thickness comparable to the Galactic disk (Reipurth & Schneider, 2008). This Cygnus super-bubble covers 18° × 13° in the sky and surrounds the association Cyg OB2 (at the centre of Cygnus X, see e.g. Massey & Thompson 1991): in fact, this strong signature served as an early proxy for the

Table 1.1: Position, spatial extent in galactic coordinates, estimated age and distance of all the OB associations in Cygnus. This table originates from Uyanıker et al. (2001) and has been adapted in Reipurth & Schneider (2008).



Assoc.	l [°]	b [°]	Δl [°]	Δb [°]	dist[a] [kpc]	age[b] [Myr]
OB1	75.50	1.17	3.0	3.4	1.25-1.83	7.5
OB2	80.10	0.90	-	-	1.44-2.10	5.0
OB3	72.55	2.20	2.5	2.2	1.58-2.51	8.3
OB4	82.50	-4.30	3.0	2.0	1.0	[c]
OB5	67.10	2.10	5.8	9.6	1.61[c]	[c]
OB6	86.00	1.00	6.0	8.0	1.70[c]	[c]
OB7	90.00	2.05	12.0	13.9	0.74-0.80	13.0
OB8	77.75	3.75	2.9	3.3	2.19-2.32	3.0
OB9	78.00	1.50	2.0	1.4	1.17-1.20	8.0

[a] Distance interval in kpc as listed by Uyaniker et al. (2001), but see text for further discussion

[b] Age derived from HR diagram

[c] Very uncertain

presence of OB associations in Cygnus (Cash et al., 1980; Bochkarev & Sitnik, 1985; Reipurth & Schneider, 2008). Nevertheless, Uyanıker et al. (2001) argued that this super-bubble was merely a projection effect due to the superposition of close emission features along the line-of-sight.

1.5.1.2 The Cygnus OB associations

Nine OB associations have previously been catalogued in the Cygnus region. Humphreys (1978) was among the first to quote the positions, spectral types, extinctions and distances of stars in these associations. Blaha & Humphreys (1989) then produced a more complete list, refined and used since by Mel'Nik & Efremov (1995), Mel'Nik & Dambis (2009) and Melnik & Dambis (2020). On the other hand, Uyanıker et al. (2001) listed some of their physical properties, as shown in Table 1.1.

Fig. 1.8 shows the spatial distribution of these associations. Most of them are found at distances from 1 - 2.5 kpc, with the sole exception being Cyg OB7 in the foreground, consistent with its larger angular size. They are mostly located at low

Galactic latitudes, thus close to the Galactic plane, but at different levels of extinction. They are also young (typically aged below 10 Myr).

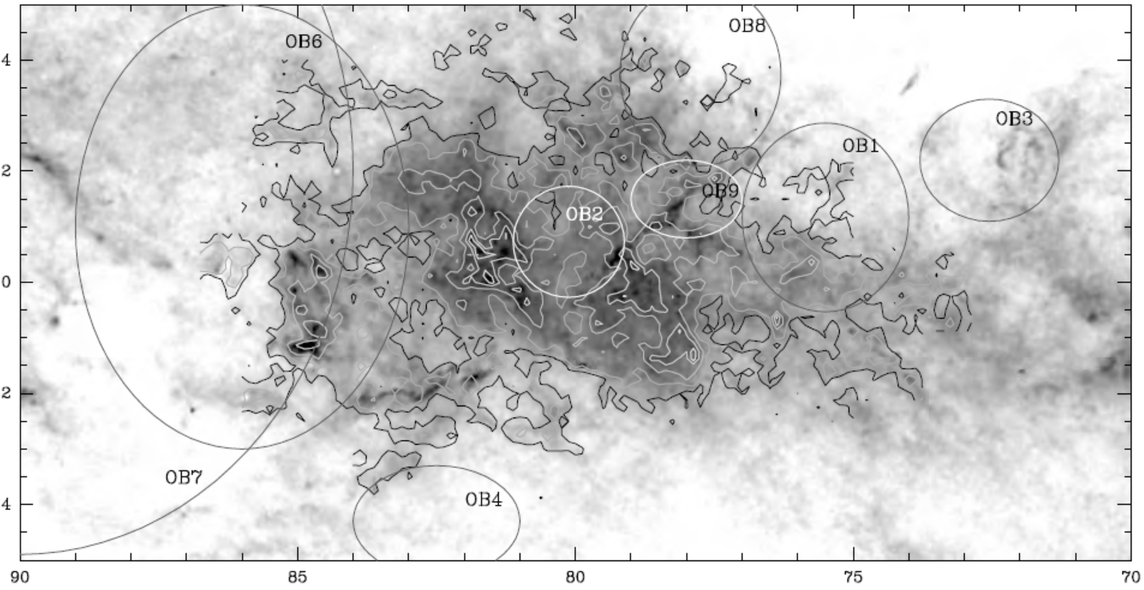


Figure 1.8: Visual extinction map of the Cygnus region, produced with data from the 2MASS survey, such that darker regions correspond to more obscured parts. OB associations are represented as grey ellipses. This figure has been taken from Reipurth & Schneider (2008) (c.f. Table 1.1).

In Fig. 1.8 contrast the levels of extinction of the Cygnus OB associations. None of them carry high level of reddening, except Cyg OB2 which spans typically from 4 to 8 mag of visual extinction (and whose most extreme objects can even go up to 20 mag, Wright et al. 2015).

1.5.1.3 Cyg OB2

By far the most studied Cygnus OB association is Cyg OB2. It was first identified by Münch & Morgan (1953) but labelled as an association in Johnson & Morgan (1954). Only a few OB stars were identified by Humphreys (1978) but the census increased with

Massey & Thompson (1991), who found a rich reddened population of stars, including the blue supergiant VI Cyg No 12, one of the most luminous objects ever observed in the Milky Way (Clark et al., 2012). Knödlseder (2000) estimated a half-light radius of 6.4 pc for the association, and also found that it could contain as many as 100 O-type stars. While some subsequent studies spectroscopically confirmed their findings (see e.g. Comerón et al. 2002; Hanson 2003; Comerón et al. 2008), the count has not been brought at this date to such a number.

Wright et al. (2015) listed 169 OB stars in Cyg OB2, including 52 O-type stars and 3 Wolf-Rayet stars. They estimated a total stellar mass of about 16,500 $M_{\odot}$, making it the most massive OB association ever studied (Wright, 2020), as well as a young age in the range 1-7 Myr. Such a census was further increased by Berlanas et al. (2018a), and is represented alongside the stars from Wright et al. (2015) in Fig. 1.9.

Cyg OB2 was first thought to have a spherically symmetric shape (Knödlseder, 2000). Wright et al. (2014) performed a structural study of the association and found considerable physical substructure and a distinct lack of mass segregation, which led them to argue that Cyg OB2 is not dynamically evolved, as confirmed in the kinematic study of Wright et al. (2016). Berlanas et al. (2019) divided Cyg OB2 in two substructures along the line of sight. The main group is located around 1760 pc and a smaller, foreground group is at 1350 pc.

Kiminki et al. (2007) conducted a radial velocity survey of 146 of the OB stars in Cyg OB2 and estimated a radial velocity dispersion of 8.03 ± 0.26 km s^{-1} (Kiminki et al., 2008). Wright et al. (2016) reported a 3D velocity dispersion of 17.8 ± 0.6 km s^{-1} from a proper motion study of about 900 low-mass stars in Cyg OB2. With data from Gaia DR2 (Gaia Collaboration et al., 2018), Orellana et al. (2021) identified 2767 new proper motion members in Cyg OB2 at an estimated distance of 1683 ± 5 pc.

Such a widely analysed OB association constitutes an ideal target to study the origin of OB associations. Wright et al. (2016) could not find evidence for expansion of Cyg OB2, and a similar finding was reached in Arnold et al. (2020).

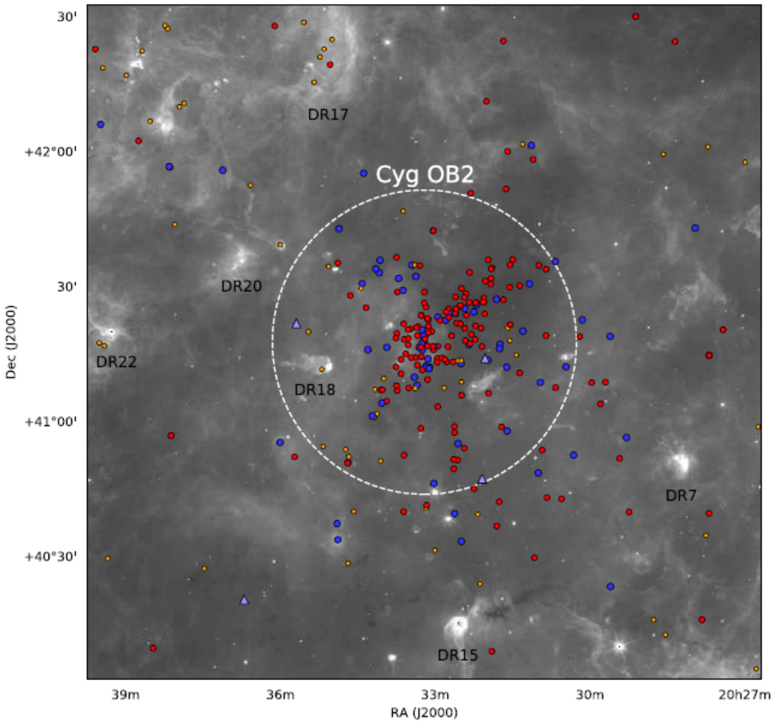


Figure 1.9: Cyg OB2 as seen in equatorial coordinates, and centered inside the dashed white ellipse representing the region of study in Wright et al. (2015). B-type stars are displayed as red dots whilst blue dots correspond to O-type stars and triangles to Wolf-Rayet stars. This figure has been taken from Wright (2020).

1.5.1.4 Other Cygnus OB associations

Apart from Cyg OB2, there is little knowledge about the other Cygnus OB associations, encouraging a deeper investigation of the region.

Cyg OB1 forms together with Cyg OB3 a supershell in which are embedded Wolf-Rayet stars and supernova remnants (Lozinskaya & Sitnik, 1988; Lozinskaya, 1998; Sitnik et al., 2019, 2020).

Little is known about Cyg OB8 and no open cluster is included within its association. It could however be linked to Cyg OB9 through the star-forming region IRAS 20126+4104 with their peculiar motion (Nagayama et al., 2015).

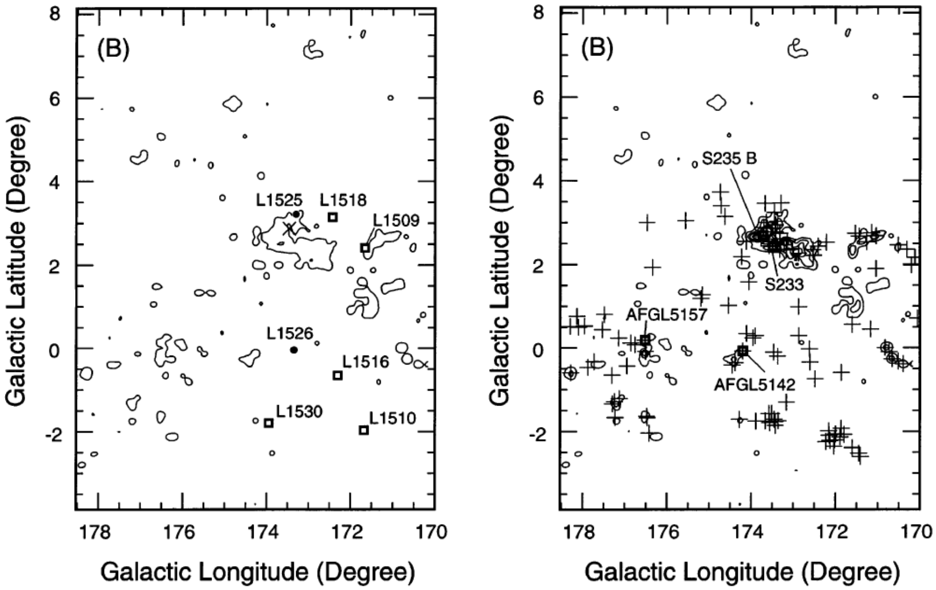


Figure 1.10: A ^{13}CO molecular cloud map of the Auriga region. The left panel shows the dark clouds while the right panels shows YSOs as plus signs along with star-forming regions. This figure was taken from Kawamura et al. (1998).

Cyg OB9 could be part of a larger structure containing Cyg OB1, OB8 and OB9, as hinted in Mel'Nik & Efremov (1995) and Reipurth & Schneider (2008).

1.5.2 The Auriga region

Auriga constitutes another region of active star formation (see e.g. Kirsanova et al. 2008; Gyulbudaghian 2011; Pandey et al. 2013), as indicated by a large population of young stars (Pandey et al., 2020), HII regions, and with a previous supernova explosion (Kang et al., 2012). A map of the Auriga region is shown in Fig. 1.10.

The Auriga constellation intercepts the local arm and the Perseus spiral arm (Marco & Negueruela, 2016). Specifically, star clusters, OB associations and intense star-forming regions have been used to trace the structure of the Perseus spiral arm between 100° and 140° (see e.g. Vázquez et al. 2008; Choi et al. 2014), but few studies have focused on Galactic longitudes between 140° and 180° where the Auriga constellation spans (Marco & Negueruela, 2016). Results from Negueruela & Marco (2003) suggest it is a less populated part of the spiral arm, in spite of its star-forming regions.

Two OB associations have been identified in Auriga (Humphreys, 1978), and are shown in Fig. 1.11. Their exact membership and distribution evolved with time (Tovmassian et al., 1994; Kalloghlian, 2009), though the updated catalogue of Melnik & Dambis (2020) provides a clear picture. Information about these associations is summarized in Table 1.2.

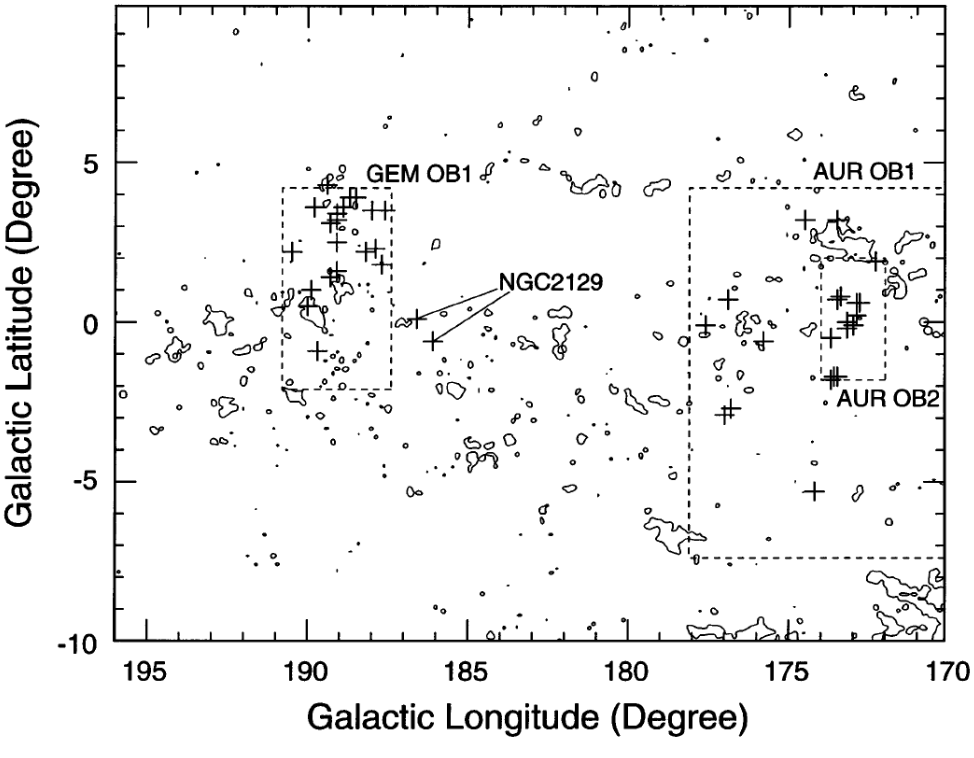


Figure 1.11: Auriga OB associations in galactic coordinates as appearing in the Auriga and Gemini regions, where the plus symbols are their members from Humphreys (1978). This ^{13}CO molecular cloud map has been taken from Kawamura et al. (1998).

Together with the open cluster NGC 1960, located at a distance of 1.17±0.06 kpc and with an estimated age of about 18-26 Myr (Jeffries et al., 2013; Joshi et al., 2020),

Table 1.2: Some physical properties of the two OB associations in Auriga. Their mass, distance and radial velocity come from Mel'Nik & Dambis (2009) and Melnik & Dambis (2020). References for the open clusters are Marco & Negueruela (2016) and Joshi et al. (2020). A11, A12 and S8 respectively stand for Alicante 11, 12 and Stock 8.

Assoc.	Mass ($M_\odot$)	d (kpc)	RV (km s^{-1})	OCs
Aur OB1	3700	1.06	-1.9	NGC 1960
Aur OB2	-	2.42	-2.7	A11, A12, S8, NGC 1893

Aur OB1 is connected to the dark cloud LDN 1525 located at 1.2-1.3 kpc (Straižys et al., 2010). Many prominent massive star binaries have been observed in Aur OB1 (Harries et al., 1998; Mayer et al., 2013) and supergiants (Dicenzo & Levesque, 2019), consistent with the view of it being a more evolved OB association.

Aur OB2 has been historically thought to include the open clusters Stock 8 and NGC 1893, which illuminate the HII regions IC 417 and 410, respectively. Aur OB2 is probably younger than Aur OB1, since Stock 8 has an age between 4 and 6 Myr. It was initially assumed that Aur OB2 stretches between Stock 8 and NGC 1893, but NGC 1893 is more distant than Stock 8 and therefore probably unrelated (Marco & Negueruela, 2016). While a distance of 3790^{+700}_{-510} pc has been estimated for NGC 1893 (Kuhn et al., 2019), Marco & Negueruela (2016) calculated a distance of 2800^{+270}_{-240} pc, ~1731 pc by Dib et al. (2018) and 2107^{+563}_{-367} pc from Cantat-Gaudin & Anders (2020) for Stock 8. This suggests that the classical definition of Aur OB2 should be revisited (Negueruela & Marco, 2003; Marco & Negueruela, 2016).

1.5.3 The Cassiopeia region

Cassiopeia constitutes another region of massive star formation. It is filled with several prominent open clusters (see e.g. Frinchaboy & Majewski 2008; Wu et al. 2009), along with emission nebulae and star-forming regions (see e.g. Kun 2008). Notably, the Cassiopeia region is home to 3C 10 and Cassiopeia A, two of the most famous and

Table 1.3: Some physical properties of the OB associations in Cassiopeia. Their mass, distance and radial velocity come from Mel'Nik & Dambis (2009), Kharchenko et al. (2013) and Melnik & Dambis (2020). References for the OCs are: Humphreys (1978), Lozinskaia et al. (1986), Fabregat & Capilla (2005), Kusuno et al. (2013), Castro-Ginard et al. (2019).

Assoc.	Mass ($M_\odot$)	d (kpc)	RV (km s^{-1})	OCs
Cas OB1	-	2.01	-42.0	UBC 84
Cas OB2	9600	2.10	-50.1	NGC 7510
Cas OB4	5400	2.30	-37.0	NGC 103
Cas OB5	9600	2.01	-45.8	King 12, NGC 7788 & 7790
Cas OB7	6700	2.01	-50.0	-
Cas OB8	8300	2.30	-34.6	NGC 581, 654, 659 & 663
Cas OB14	-	0.88	-15.0	-

well studied supernova remnants (see e.g. Ruiz-Lapuente et al. 2019; Vink et al. 2022). Chen et al. (2022) identified in this region a large molecular filament with an extent of ∼400 pc, that they thought was formed by the Galactic shear forces and appears as a spur of the local arm.

Ten OB associations in Cassiopeia were catalogued by Humphreys (1978). The list was updated in Blaha & Humphreys (1989) and Mel'Nik & Efremov (1995), with the most recent version shown in Melnik & Dambis (2020). Properties of the OB associations of interest are displayed in Table 1.3. With the exception of the foreground Cas OB14, they are all located at distances of 2 - 2.5 kpc, and thus intercept the Perseus spiral arm (de la Fuente Marcos & de la Fuente Marcos, 2009; Marco & Negueruela, 2016). They are all also very young with ages ranging from 6 to 10 Myr (Cazzolato & Pineault, 2003; Tetzlaff et al., 2010). They contain both runaway stars (Maíz Apellániz et al., 2018) and supergiants (see e.g. Levesque et al. 2005; Dicenzo & Levesque 2019).

The population of Cas OB2 comprises pulsating, solar-metallicity and relatively evolved stars (Usenko et al., 2001). Two dust shells surround this association with its related open cluster NGC 7510, which are thought to ionize the emission nebula Sh 157 (Lozinskaia et al., 1986). Christopoulou et al. (1995) also found that Cas OB2 created the nebula NGC 7635 inside the Sh 2-162 complex.

The population of Cas OB5 comprises two hypergiants (Bartaya et al., 1994). This association is encompassed by a supershell of mass $7.5 \times 10^5\ M_\odot$ (Moór & Kiss, 2003). Suad et al. (2016) detected another supershell, although they argued it is a distinct one, that they named GS 118+01-44. Even if Cas OB5 contains stars inside this supershell, Suad et al. (2016) calculated that the released energy of its members through their stellar wind was too low to create it. Cas OB5 could have however created the HII region Sh 2-173, which is very young (0.6 - 1 Myr, see Cichowolski et al. 2009). It has been shown with *Gaia* data that Cas OB5 can actually be divided into two groups, one at ∼3.3 and one at ∼4.3 kpc: this highlights the need to revisit this association (Chentsov, 2020).

The population of Cas OB7 comprises many YSOs (Cazzolato & Pineault, 2002). Cazzolato & Pineault (2003) studied the HI shell of mass $3 \times 10^4\ M_\odot$ surrounding this association and outlined a strong interaction between the stars and the gas. Chen et al. (2022) also identified a giant shell coinciding with the position of Cas OB7, however they found that Cas OB7 was too distant to be related to it (Cazzolato & Pineault, 2003).

Cas OB14 includes the runaway star κ Cassiopeiae, an α Cygni variable star (Walborn, 1971). The kinematics of Cas OB14 was analysed in Ward & Kruijssen (2018), where they did not measure any expansion signature and thus favoured the *hierarchical* model for its emergence.

Both Cas OB1 and Cas OB7 are thought to belong to the ‘Cassiopeia-Perseus family’, a large structure traced by OB associations and open clusters, included within the spiral arms at ∼2 kpc and with a diameter of ∼600 pc (de la Fuente Marcos & de la Fuente Marcos, 2009). Furthermore, Cas OB7 may undergo sequential star formation together with Cas OB4 and Cas OB5. It was even proposed that they could form a single larger OB association (Velasco et al., 2013).

A map of the Cassiopeia region, where the Cassiopeia OB associations are located, is shown in Fig. 1.12.

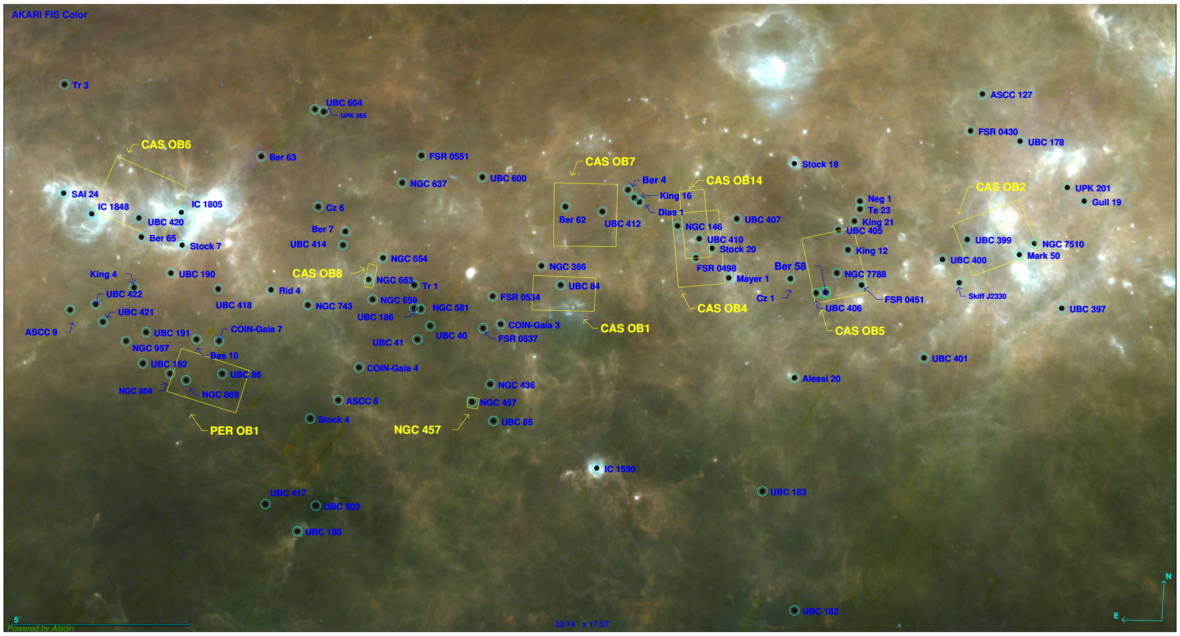


Figure 1.12: Map of the Cassiopeia region produced in Aladdin. The OB associations have been outlined in yellow and come from Humphreys (1978), while the open clusters stem from Cantat-Gaudin & Anders (2020). Credits: Sara Berlanas.

2 SED fitting

This second chapter is dedicated to a description of the SED fitting code that will be used throughout this thesis to derive physical stellar parameters and to identify candidate OB stars. This code fits an observed SED, using a combination of astrometry from *Gaia* EDR3 and photometry from optical and NIR surveys, to a reddened model SED made with stellar atmospheric and evolutionary models.

The fitting process works by forward modelling an observed SED from a set of model parameters. The model parameters are the stellar initial mass and age, and the distance to the star. The mass and age are used, in combination with stellar evolutionary models, to derive effective temperature and luminosity, which in turn provide, in combination with stellar atmosphere models, an unreddened absolute SED. The extinction to the modelled SED is estimated from the distance parameter and a 3D extinction map, and these are combined to convert the absolute SED to an observed SED that can then be compared to the observed SED. This process is discussed in full details below.

2.1 Observed SEDs

An SED (spectral energy distribution) of a star corresponds to its emitted flux or magnitude per unit wavelength. For this work, I consider four photometric surveys to build an observed SED: *Gaia*, IGAPS (IPHAS+UVEX), 2MASS and UKIDSS, whose respective descriptions can be found below.

2.1.1 *Gaia*

Gaia is an ESA spacecraft launched in December 2013, dedicated to an astrometric and a photometric survey of the entire sky, including some spectroscopic data obtained from its radial velocity spectrometer. Thanks to its high sensitivity and accuracy, it is working towards a myriad of scientific goals, from the dynamics and stellar formation of the Milky Way to the investigation of the Local Group (Gaia Collaboration et al., 2016). *Gaia* data provides photometry in the optical domain along with parallaxes and proper motions (Jordi et al., 2010).

For this work I primarily use data from *Gaia* EDR3, published in December 2020, with the most up-to-date photometry and astrometry. The data included in EDR3 was obtained during the first 34 months of the mission (Gaia Collaboration et al., 2021). A total number of 1,811,709,771 sources is contained in this catalogue, including 1,467,744,818 with either a 5 or 6-parameter astrometric solution, that is with at least the equatorial coordinates, parallax and proper motions (RA, DEC, μ_{RA}, μ_{DEC}, ϖ) (Lindegren et al., 2021a). I also use the three main photometric bands G, G_{BP} and G_{RP}. They have a respective central wavelength of 6390.7 Å, 5182.6 Å and 7825.1 Å and their full-width at half-maximum (FWHM) are equal to 4584.2 Å, 2659.0 Å and 2927.5 Å with the G band thus overlapping the two others (see Fig. 2.1). About 1.806 billion sources possess measured G-band photometry, and this number reduces to about 1.542 and 1.555 billion sources for those with, respectively, G_{BP} and G_{RP} photometry. Finally, the G band covers a wide range of magnitudes, saturating around 3 mag and with a limiting magnitude of 21 mag (Riello et al., 2021).

2.1.2 IGAPS

The Isaac Newton Telescope (INT) Photometric Hα Survey (IPHAS) of the Northern Galactic Plane operated between 2003 and 2012 (Drew et al., 2005). It spans Galactic latitudes (b) between -5° and 5° and Galactic longitudes (l) between 30° and 215°, covering an area of 1860 deg^2. The full catalogue presents a total number of 218,991,524 sources in its latest version (Barentsen et al., 2014; Monguió et al., 2020). This survey observed the northern Galactic plane in the Hα narrow band, but also in the r and i broad band Sloan filters. The r filter possesses a central wavelength of 6240 Å, whilst the i filter has a central wavelength of 7743 Å as shown in Fig. 2.1.

The UltraViolet EXcess survey (UVEX) is a survey complementary to IPHAS, covering the same area (Groot et al., 2009). The main difference lies in the filters, namely He I, U, g, r, the latter band already included in IPHAS, but observed again to provide coeval photometry with the g and U filters. The g band uses a Sloan filter while the U band uses an RGO filter, with respective central wavelengths of 4846 Å and 3640 Å (Stroe et al., 2017).

IPHAS and UVEX have been combined into one larger survey called IGAPS (Monguió et al., 2020). From this I select the bands g, r and i, whose transmissions curves are shown in Fig. 2.1[1]. These bands have respectively limiting magnitudes of 22.4, 21.5 and 20.4 mag, while they saturate at magnitudes of 14, 13 and 12 mag, though these are mean values and their saturation varies according to the position in the sky (Monguió et al., 2020).

2.1.3 2MASS

The Two Micron All Sky Survey (2MASS) operated between 1997 and 2001 and reached a photometric precision better than 0.03 mag for bright sources and an astrometric accuracy better than 100 mas. It investigated nearly all the sky (99.998 %) in the near-infrared domain, with a total number of 470,992,970 detected sources. Its three main bands are J, H and K_s, whose respective central wavelengths are 1.235 μm, 1.995 μm and 2.159 μm. They saturate at $\sim$ 4.5, 4 and 3.5 mag whereas their magnitude limits are respectively equal to 15.0 - 15.8, 14.3 - 15.1 and 13.5 - 14.3 mag, depending whether the sources are extended or not (Cutri et al., 2003).

2.1.4 UKIDSS

The UKIDSS (UKIRT Deep Infrared Survey) Galactic Plane Survey (GPS) survey covers an area of 1868 deg^2 in the northern Galactic plane (b between -5° and 5° and irregular coverage in l). Specifically, I work with data from DR6 which contains

---

[1] At first, the U band was utilised, but its lack of calibration decreased its accuracy (Monguió et al., 2020).

557,877,373 unique sources (Lucas et al., 2008). From this catalogue I use the J, H and K bands, with respective central wavelengths of 1.248 μm, 1.631 μm and 2.201 μm (Hewett et al., 2006). UKIDSS reaches fainter magnitudes than 2MASS, with completeness limits of ∼19.50, 18.75 and 18.00 mag, but saturates at brighter magnitudes too, respectively at 13.25, 12.75 and 12.00 mag (Lucas et al., 2008).

Throughout the thesis the subscript '2M' is chosen for 2MASS and 'U' for UKIDSS, to distinguish between near-IR photometry obtained from the two surveys.

2.1.5 Transmission curves

Each of the filters is displayed in Fig. 2.1, where the transmission designates the fraction of total light that enters the filters at this specific wavelength. In total, 12 photometric bands will compose the observed SED.

2.2 Model SEDs

The model SEDs are derived from the convolution of reddened stellar atmospheric models and the filter profiles described above. The parameters for the stellar atmosphere models were selected from stellar evolutionary models. This process is described below.

2.2.1 Stellar atmospheric models

The model SEDs are based on two stellar atmosphere models:

1. The Kurucz synthetic stellar library provides a collection of theoretical stellar spectra whose wavelengths extend from 1300 Å to 100,000 Å with a grid of $T_{\rm eff}$ between 3000 K and 25,000 K, with a variety of $\log g$ values (expressing the surface gravity in cm s^{-2}) and metallicities (Coelho, 2014). For this work I extracted a sample of spectra between 4000 K and 20,000 K with a grid spacing

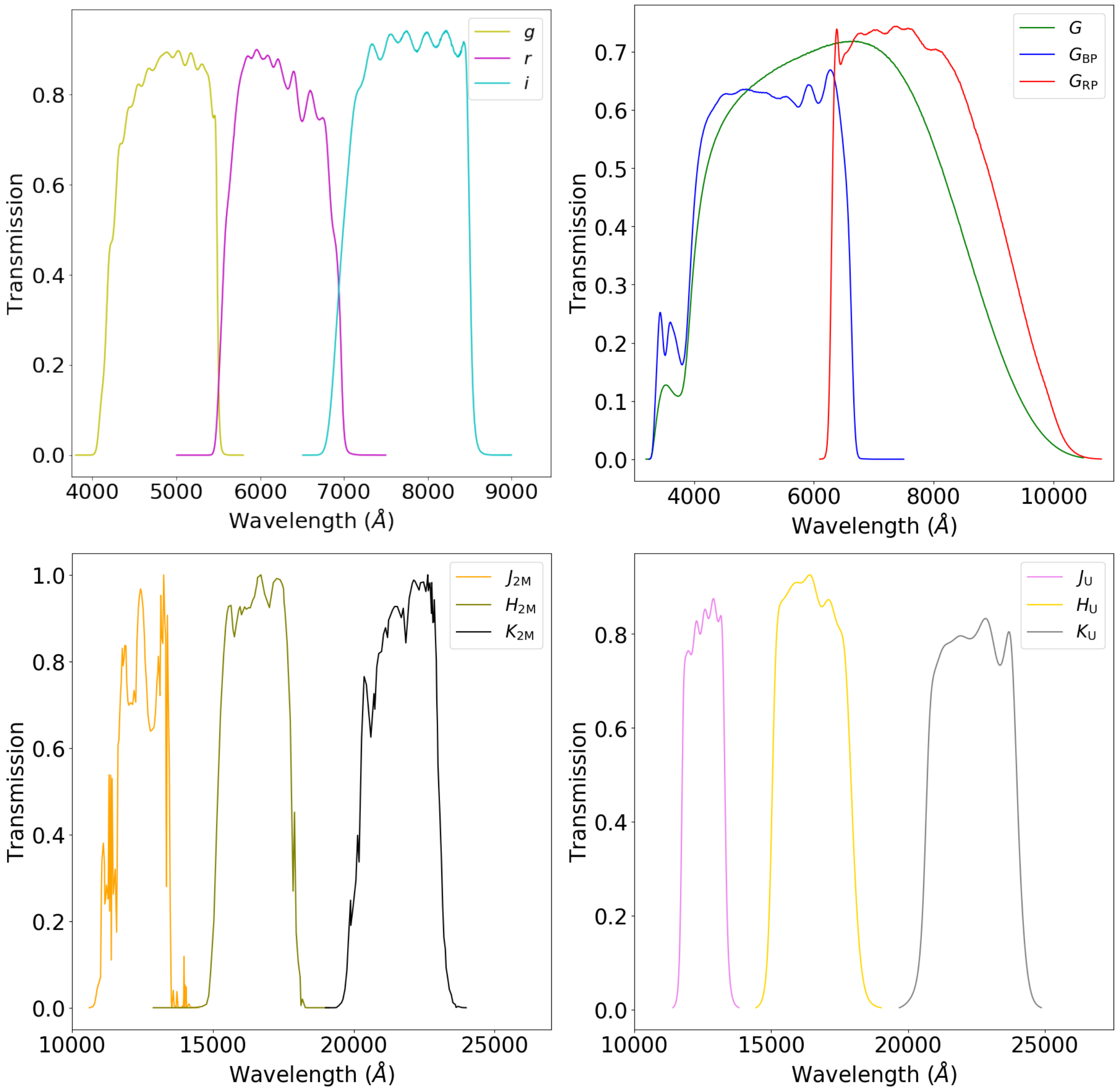


Figure 2.1: Transmission as a function of wavelength of the IGAPS photometric filters at upper left, Gaia at upper right, 2MASS at lower left and UKIDSS at lower right.

of 1000 K, each with a $\log(g) = 4$ (except for the model at 4000 K because it was unavailable, I had to choose a model with $\log(g) = 4.5$ instead) and a metallicity ([Fe/H]) equal to 0.

2. The Tubingen NLTE Model Atmosphere Package (TMAP Grid 2) contains models with wavelengths extending from 4 to 400,000 Å, with $T_{\rm eff}$ between 20,000K and 150,000 K (Werner & Dreizler, 1999; Rauch & Deetjen, 2003; Werner et al., 2003). Here I choose again spectra with $\log g = 4$, a solar metallicity and with a $T_{\rm eff}$ step of 1000 K. These models cover the temperature range from 21,000 to 50,000 K.

The choice of a constant surface gravity is constrained by the availability of the models and therefore constitutes a simplification whose impact will be discussed in Section 2.5. An example model for each stellar library is displayed in Fig. 2.2.

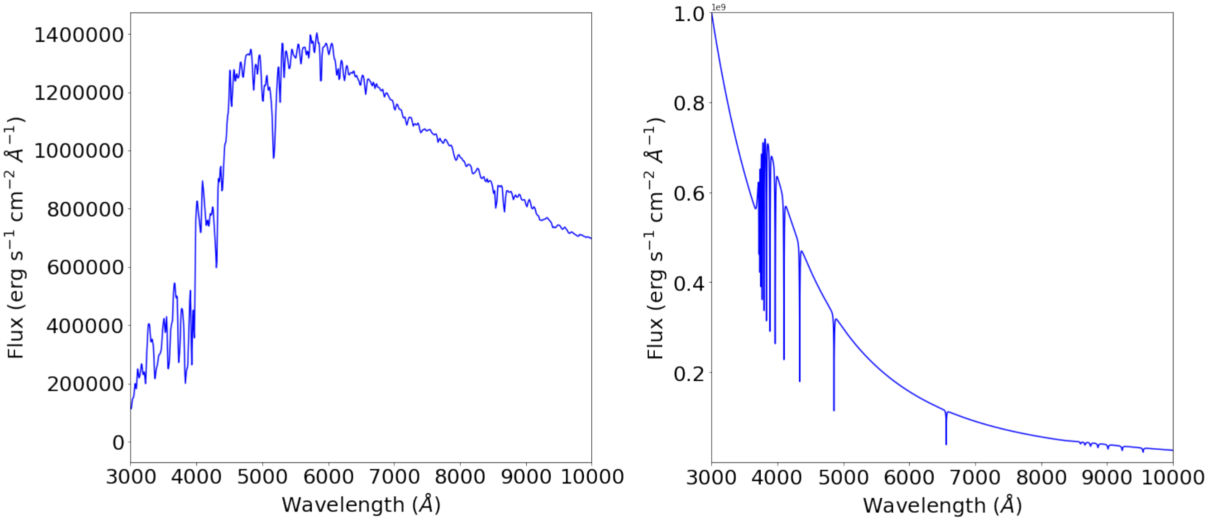


Figure 2.2: Stellar atmosphere models for the Kurucz model at 5000 K (left panel) and the TMAP 2 model at 21,000 K (right panel), both with $\log g = 4$. Each spectrum has been represented from 3000 to 10,000 Å. For the right panel, the flux units have been divided by 10^9 for readability.

2.2.2 Reddening the spectra

Extinction affects light travelling along the path from the source to the observer. Dust grains in the interstellar medium absorb, transmit and redirect electromagnetic waves, with an intensity heavily dependent upon wavelength. It is especially strong in the blue and UV part of the spectrum while weaker in the infrared (Fitzpatrick, 1999, 2004). The stellar spectra models consequently have to be reddened, and to that end I utilised the extinction laws from Fitzpatrick et al. (2019), which are detailed below.

The total extinction at a wavelength λ, A_λ, is expressed as:

$$k(\lambda - V) = \frac{A_\lambda - A_V}{A_B - A_V} \tag{2.1}$$

which is also written specifically for the V and B Johnson filters. The extinction is expressed in units of magnitude. For instance, $A_V = 1$ mag will increase the magnitude by one unit in the V band.

The ratio between the total and the selective extinction, R_V, is then written:

$$R_V = \frac{A_V}{A_B - A_V} \tag{2.2}$$

This allows me to rewrite A_λ as:

$$A_\lambda = A_V \left[a(\lambda) + \frac{b(\lambda)}{R_V} \right] \tag{2.3}$$

where $a(\lambda)$ and $b(\lambda)$ are equal to:

$$a(\lambda) = s(\lambda - 55) + 1 \tag{2.4}$$

$$b(\lambda) = -s(\lambda - 55) \left[\frac{R(55)_0 + \beta}{\alpha} \right] + \left[\frac{k(\lambda - 55)_0 - \beta}{\alpha} \right] \tag{2.5}$$

Eq. (2.4) and Eq. (2.5) introduce the following new parameters:

- The unitless parameters α and β are linked to R_V through the equation $R(55) = \alpha\, R_V - \beta$ where the symbol '55' stands for the monochromatic measurement of total to selective extinction at 5500 Å, specifically applied here to replace the V filter. In this work, $\alpha = 0.990$ and $\beta = 0.049$, since the notation '0' stands for an average Milky Way value (Fitzpatrick et al., 2019). Consequently, with $R_V(0) = 3.10$, this gives $R_{55}(0) = 3.02$.

- $s(\lambda - 55)$ and $k(\lambda - 55)_0$ are provided in a table from Fitzpatrick et al. (2019), for a grid of wavelengths from 1150 to 10,000 Å (also including the central J, H and K_s wavelengths from 2MASS). It is a slope of a fit described through $s(\lambda - 55) = \frac{\Delta k(\lambda - 55)}{\Delta R(55)}$. These functions are both displayed in Fig. 2.3.

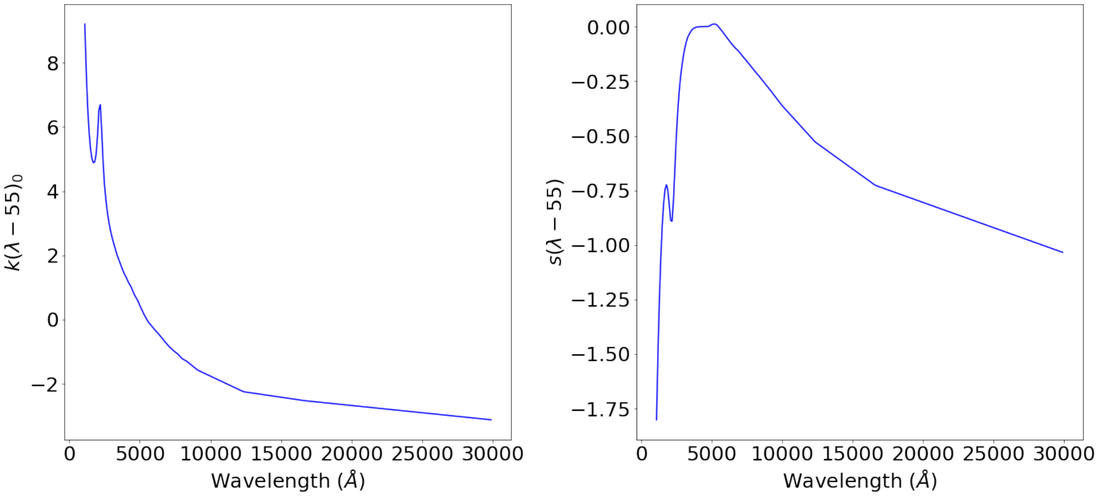


Figure 2.3: Variation of $k(\lambda - 55)_0$ (left panel) and $s(\lambda - 55)$ (right panel), as a function of wavelength. The original data from Fitzpatrick et al. (2019) has been interpolated linearly along the wavelength grid of the stellar atmosphere models. Limits at infinity are equal to -3.020 for $k(\lambda - 55)_0$ and -1.000 for $s(\lambda - 55)$, therefore the interpolation is an approximation that is satisfactory for the range of interest.

To apply the reddening process, both $s(\lambda - 55)$ and $k(\lambda - 55)_0$ are interpolated over the range of wavelengths of the stellar spectra. A_V is sampled in a set of values between 0 and 15, covering a grid of 0.1 mag, with the aim of exploring a wide parameter space (using the same upper limit as Mohr-Smith et al. 2015). R_V is set to 3.1 as the mean Galactic value (Fitzpatrick, 2004). The reddened fluxes are computed through $F_{red}(\lambda) = F_0\,10^{-0.4\,A_\lambda}$, where F_0 is the unreddened flux of the spectra and $F_{red}(\lambda)$ stands for the reddened flux.

2.2.3 Convolution of the reddened fluxes

All stellar spectra were convolved with the filters introduced in Section 2.1. The equation used to derive synthetic magnitudes is the same as used by Smith (2016), and is expressed as:

$$m_x = -2.5 \log \left[\frac{\int \lambda \, F_{red} \, T_x \, d\lambda}{\int \lambda \, F_V \, T_x \, d\lambda} \right] \tag{2.6}$$

where T_x is the transmission curve for filter x and F_V corresponds to the Vega flux, a reference flux to normalize the integral, with its synthetic spectrum originating from Bohlin (2007). A factor λ is included in the integral because the energy flux has to be converted to photon flux for photon-counting devices like CCDs.

Eq. (2.6) was applied to the Johnson V band filter, from which I calculated a set of colours for each filter, as written in Eq. (2.7):

$$V - x = m_V - m_x \tag{2.7}$$

2.2.4 Model SED at ZAMS

Using Eq. (2.7), I calculated the relationship between the V magnitude and the magnitude at any given filter, which gives the basic shape of the model SEDs. Normalising these model SEDs however required to determine the absolute V magnitude. To that end, I selected the zero-age main-sequence (ZAMS) models from Ekström et al. (2012) to interpolate the magnitude with luminosity.

A direct relation between the luminosity and the absolute bolometric magnitude $M_{\rm bol}$ (i.e. the absolute magnitude over all wavelengths) is given by:

$$M_{\rm bol} = 4.75 - 2.5 \log \left(\frac{L}{L_\odot} \right) \tag{2.8}$$

where $L_\odot$ is the luminosity of the Sun.

The absolute magnitude in the V band can then be derived from M_{bol} with the bolometric correction. I performed this conversion using the table from Pecaut & Mamajek (2013), applying:

$$M_V = M_{\rm bol} - BC_V \tag{2.9}$$

where M_V designates the absolute magnitude in the V band, and is reddened through:

$$m_V(10\,\mathrm{pc}) = M_V + A_V \tag{2.10}$$

where $m_V(10\,\mathrm{pc})$ corresponds to the apparent magnitude in the V band at a distance of 10 pc.

To produce a model SED suitable for the fitting of the available observations, I calculated the magnitude for each band for all combinations of T_{eff} and A_V. At this step, I included distance in the model SED. The apparent magnitude for a filter x is given by:

$$m_x(d) = m_V(10\,pc) - (m_V - m_x) + 5\,\log(d) - 5 \tag{2.11}$$

where d is the distance in parsecs.

The effects of reddening and distance now both shape the SED, the latter incorporated through the distance modulus as shown in Eq. (2.11).

2.2.5 Quality control of model SEDs

2.2.5.1 Dependence on parameters

T_{eff} and A_V characterize a model SED and affect the photometry in different ways, as displayed in Fig. 2.4.

- Increasing T_{eff} leads to brighter magnitudes and a general steepening of the SED in the blue for the hotter stars, while cool stars will be brighter in the red and near-IR part of the spectrum.

- Increasing A_V causes the magnitude to increase as starlight is absorbed. Nevertheless, this general effect influences the various bands differently. As can be observed in the lower left panel of Fig. 2.4, the magnitudes increase more at shorter wavelengths. This is because the effect of reddening is stronger in the UV and visible bands of the spectrum and weaker in the infrared (see Fig. 2.3 and more generally Fitzpatrick 2004 and Fitzpatrick et al. 2019).

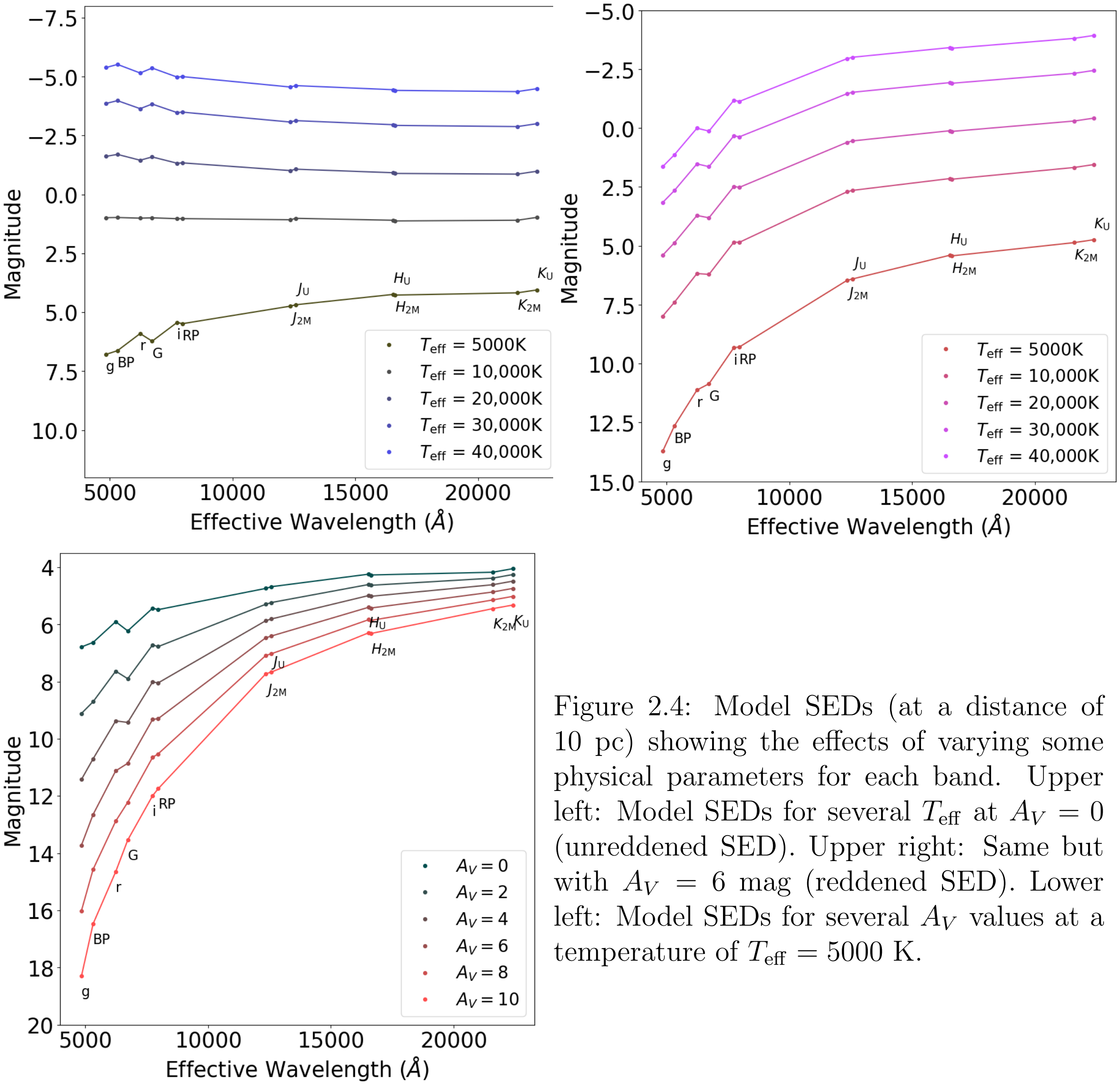


Figure 2.4: Model SEDs (at a distance of 10 pc) showing the effects of varying some physical parameters for each band. Upper left: Model SEDs for several $T_{\rm eff}$ at $A_V = 0$ (unreddened SED). Upper right: Same but with $A_V = 6$ mag (reddened SED). Lower left: Model SEDs for several A_V values at a temperature of $T_{\rm eff} = 5000$ K.

2.2.5.2 Comparison with literature colours

To validate the photometric colours from the model SEDs, I compared my estimated colours with the empirical colours from Pecaut & Mamajek (2013). In the most recent tabulation of the table from Pecaut & Mamajek (2013) (from April 2022, see `http://`

`www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt`), there is no near-IR colour for spectral types earlier than O9V ($T_{\text{eff}} > 32{,}500$ K) and no *Gaia* colours for spectral types earlier than B9V ($T_{\text{eff}} > 10{,}700$ K), so the comparison is limited to temperatures cooler than these limits for those filters.

Plots for J-H and H-K are displayed in Fig. 2.5 and show a reasonable agreement. At first I noticed an offset, so I added 0.0375 to the model H apparent magnitude as a calibration. This brought them in line with empirical values and are reflected in the good agreements displayed in Fig. 2.5. I attribute this to either uncertainties in the modelled OH lines in stellar spectra or the H-band filter transmission function.

Plots for G-RP and BP-RP are displayed in Fig. 2.6. They also show a reasonable agreement, albeit stronger for the BP-RP colour, while the G-RP models tend to be slightly underestimated at higher values. This comparison is valid for cool stars due to the limitations of the table.

Finally, since no IGAPS colour is provided in Pecaut & Mamajek (2013), I used the table from Verbeek et al. (2012) instead. g-r colours are listed for spectral type up to O5V (or $T_{\text{eff}} = 41,000$ K), and the comparison is also shown in Fig. 2.6. There is a slight deviation from the smallest values before reaching a better agreement for the highest values.

The median absolute difference between literature colours and model colours were respectively (g-r, BP-RP, G-RP, J-H, H-K) = (0.025, 0.023, 0.033, 0.003, 0.011). Unreddened model colours are thus close to the literature colours, giving an early verification of the models.

2.2.6 Evolutionary models

So far the model SEDs only represent stars on the ZAMS. A more realistic model should represent all types of stars and notably the evolved stars. To do so I extracted the evolutionary tracks from Ekström et al. (2012). This gives a set of initial masses, from 0.8 $M_{\odot}$ to 120 $M_{\odot}$. The evolutionary tracks include the model stars at different stages of their evolution at a given age. It means that a unique value of initial mass

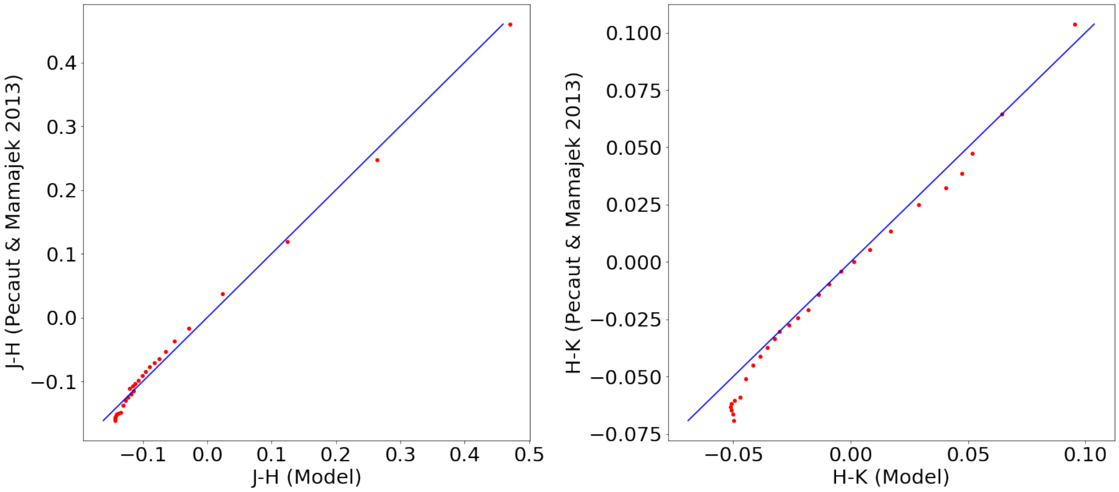


Figure 2.5: Comparison of 2MASS colours (J-H and H-K) from our unreddened model SED and Pecaut & Mamajek (2013), from 5000 K to 32,000 K. The blue line shows a 1:1 correlation.

and age will correspond to a unique value of $\log(T_{\text{eff}})$ and $\log(L/L_\odot)$.

The SED fitter treats initial mass and age as the critical model parameters, using the evolutionary models to infer $\log(T_{\text{eff}})$ and $\log(L/L_\odot)$, which are then used to build a model spectrum. The process is outlined as follows:

1. I sample the logarithm of initial mass, labelled $\log(M/M_\odot)$: it varies between -1 (0.1 $M_\odot$) and 2 (100 $M_\odot$), with a step size of 0.1. Models with masses lower than 1 $M_\odot$ and down to 0.1 $M_\odot$ have only been extrapolated along the ZAMS[2]. An initial mass of 100 $M_\odot$ is chosen as upper limit for symmetry. While the models from Ekström et al. (2012) can reach 120 $M_\odot$, a O3V star with an initial mass of 59 $M_\odot$ in Pecaut & Mamajek (2013) is characterized by $\log(T_{\text{eff}}/K) = 4.65$, and the maximum temperature reached by my model

---

[2] This model is thus not suitable for an exploration of evolved stars cooler than the Sun, but since this work is dedicated to hot and luminous stars, and such stars are rare due to the time required for them to have evolved off the main sequence, this approximation can be considered valid.

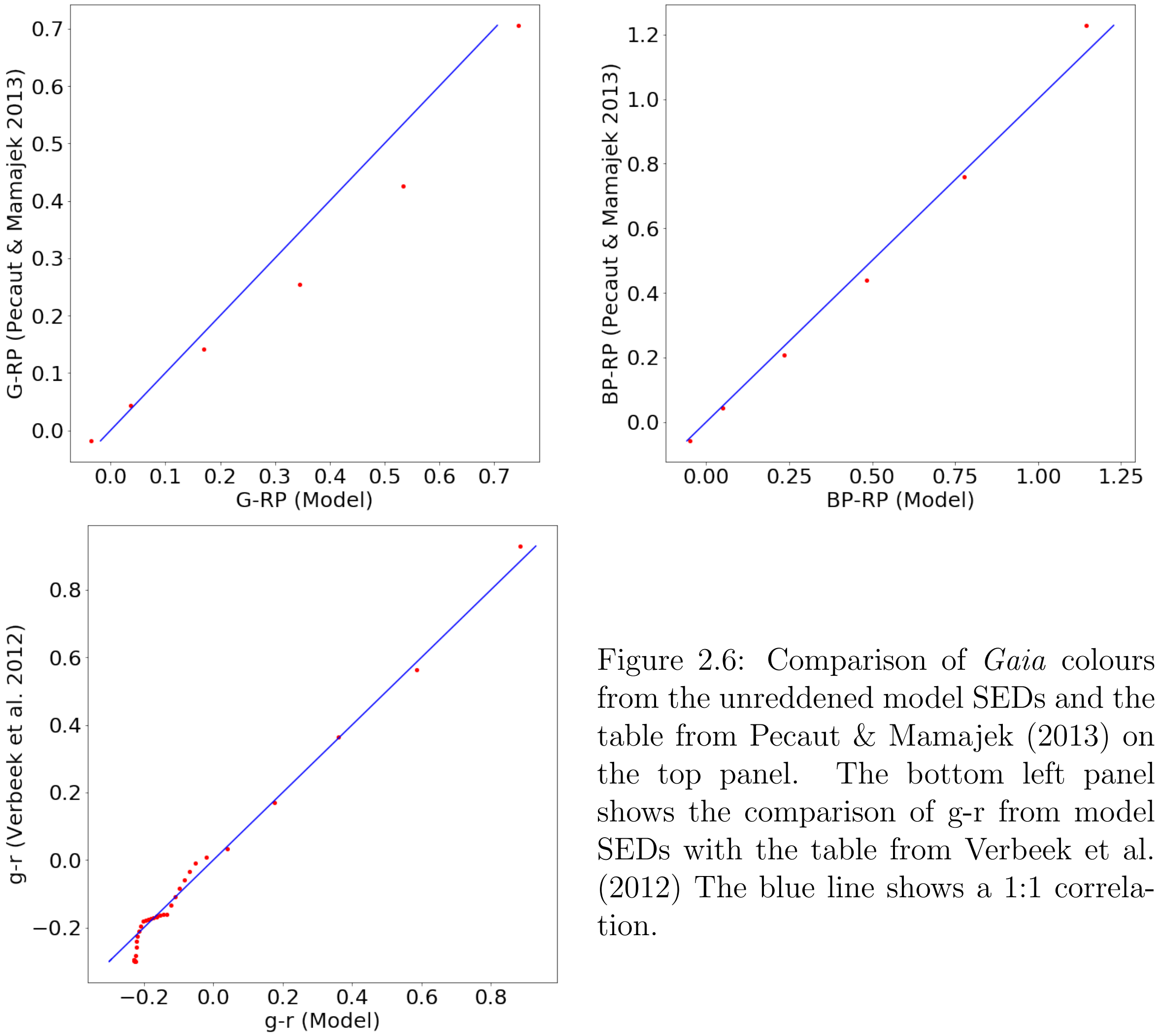


Figure 2.6: Comparison of *Gaia* colours from the unreddened model SEDs and the table from Pecaut & Mamajek (2013) on the top panel. The bottom left panel shows the comparison of g-r from model SEDs with the table from Verbeek et al. (2012) The blue line shows a 1:1 correlation.

is $\log(T_{\text{eff}}/K) = 4.69$. Therefore, this upper mass limit is beyond the allowed maximum temperature of the model.

2. For each value of $\log(M/M_{\odot})$, I find the inferior and superior evolutionary track in the table found in Ekström et al. (2012), and I interpolate the logarithm of the age, effective temperature and luminosity along these tracks.

3. I define a new quantity, the fractional age, labelled Fr(Age), such that:

$$\mathrm{Fr(Age)} = \frac{\mathrm{Age} - \mathrm{Min(Age)}}{\mathrm{Max(Age)} - \mathrm{Min(Age)}} \tag{2.12}$$

   where Min(Age) stands for the minimum age of an interpolated evolutionary track and Max(Age) corresponds to the maximum age. This parameter ranges between 0 and 1. I choose a very small step size of 10^{-5} in order to better capture the late evolutionary stages of a star, when the variation of temperature and luminosity will be the strongest (Ekström et al., 2012).

4. I interpolate again $\log(T_{\mathrm{eff}})$ and $\log(L/L_{\odot})$ along this sequence of fractional age.

By the end of this process, I have obtained a list of unique values of effective temperature and luminosity sampled from a regular grid of $\log(M/M_{\odot})$ and Fr(Age). These interpolated evolutionary tracks are represented in Fig. 2.7. With these interpolated evolutionary tracks, the new model SEDs for each band x can be expressed as:

$$m_x(d) = M(\mathrm{Evo}) + 5\ \log(\mathrm{d}) - 5 \tag{2.13}$$

where $M(\mathrm{Evo})$ is the magnitude taking the evolution into account.

2.3 Extinction

The last ingredient needed to compute the model SED is to estimate the reddening to the source. In an earlier version of this code I used A_V as a free parameter. However, this led to a number of degeneracies between reddened early-type stars and unreddened cool giants when fitting some stars. The strategy to overcome this was to tie the extinction of the modelled star to its model distance. I do this using the 3D extinction map from Green et al. (2019). It was built with the photometry from 2MASS (Cutri et al., 2003) and Pan-Starrs 1 (Chambers et al., 2016) and with the parallaxes from *Gaia* DR2 (Gaia Collaboration et al., 2018). This map covers all declinations higher

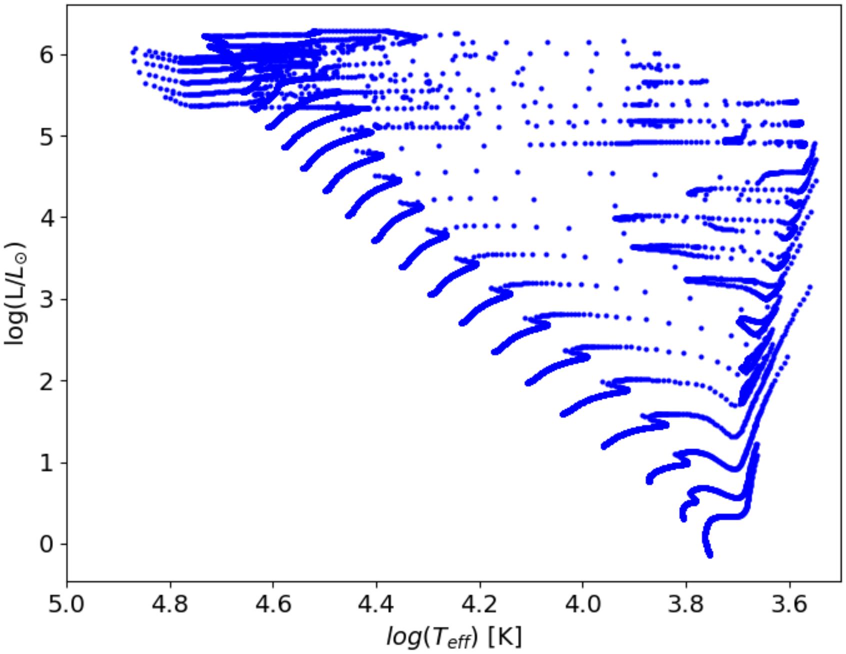


Figure 2.7: Interpolated evolutionary tracks with the grids of masses and ages, based on the evolutionary tracks from Ekström et al. (2012).

than -30° (covering three quarters of the sky) and distances up to several kpc. It thus gives a relation between reddening and distance for given coordinates with good precision.

For this work, I produced extinction curves for each line-of-sight within a given sample. The distance of a source, derived from the parallax (see Section 2.4), was used to estimate the extinction through interpolation. Green et al. (2019) produces values of $E(g-r)$, that I convert into A_V with $A_V = 2.742\,E(g-r)$, where the conversion factor originates from Schlafly & Finkbeiner (2011). A modelled distance for a given source (with known coordinates) can thus provide a model extinction that can be used to model the unreddened SED. This value was then incorporated to build a reddened model SED.

2.4 SED fitting process (version 1.0)

The model SEDs are fitted to the observed SEDs through Bayesian inference and using a maximum-likelihood test, applying a Markov Chain Monte Carlo (MCMC) simulation to explore the parameter space. To that end, I use the `emcee` package in Python from Foreman-Mackey et al. (2013), an MCMC ensemble sampler. This process requires these following ingredients:

- Empirical data, written in the form $d = [d_1, d_2, ..., d_n]$.
- A model for this data, also expressed as $\theta = [\theta_1, \theta_2, ..., \theta_n]$, where n represents the number of model parameters.
- $P(\theta)$ stands for the prior, and includes the interval of possible values for each model parameter, appearing thus as a constraint on the models that can be fitted.
- $P(d|\theta)$, named the likelihood, corresponds to the probability that the data are measured provided the model parameters.

Therefore it is possible to compute the posterior probability, labelled $P(\theta|d)$, which is the probability of a given set of model parameters, given the observations, by applying the theorem of Bayes:

$$P(\theta|d) \propto P(d|\theta)\, P(\theta) \tag{2.14}$$

The list of parameters I am seeking to estimate are:

$$\theta = [\log(M/M_\odot), \mathrm{Fr(Age)}, d, \ln(f)] \tag{2.15}$$

where $\ln(f)$ represents the logarithm of an additional uncertainty added in order to achieve agreement between the observations and the model. Including such a parameter helps χ^2 to converge to 1 for the likelihood calculation (Foreman-Mackey et al., 2013). The new errors will therefore be expressed as (e.g. Casey 2016):

$$s^2 = \sigma^2 + f^2\, \mathrm{Model}^2 \tag{2.16}$$

where Model is the combination of the model SED and the model parallax, the latter related to distance through $\varpi = \frac{1000}{d}$, if ϖ is expressed in mas and d in pc.

The priors for these parameters are written as follows:

$$\ln(P(\theta)) = \begin{cases} \log(\frac{1}{2\,L^3}\,d^2\,\exp\,(\frac{-d}{L})) & \text{if } \begin{cases} -1.0 \leq \log(M/M_{\odot}) \leq 2.0 \\ 0.0 \leq \mathrm{Fr(Age)} \leq 1.0 \\ 0.0 \leq d \leq 5000.0\,\mathrm{pc} \\ -10.0 \leq \ln(f) \leq 1.0 \end{cases} \\ -\infty & \text{otherwise} \end{cases} \tag{2.17}$$

The prior on distance originates from Bailer-Jones (2015), such that the volume density of stars has an exponential drop, with the aim of deducing distances from negatives parallaxes and poor precision measurements. Any of the methods required the introduction of a length scale L. In Astraatmadja & Bailer-Jones (2016), they concluded that $L = 1.35$ kpc gave the most accurate results. More complex methods have been applied in later papers (see e.g. Bailer-Jones et al. 2021), notably the introduction of a scale length dependent upon latitude and longitude, but I stick to a constant value for simplification.

Boundaries of the priors have been chosen as described below:

- Priors on $\log(M/M_{\odot})$ and Fr(Age) have been explained in Section 2.2.6.
- The upper prior on distance, set depending on the region of study, helps the MCMC algorithm to run faster.
- The priors on $\ln(f)$ are quite large in order to give enough freedom to maximize the likelihood. I choose the same limits as the ones used in Foreman-Mackey et al. (2013). Even the use of systematic errors in the observed SED and parallax (see below) does not reduce the necessity of this prior. As a consequence, it will be an indicator of the quality of the fit: the smaller it is, the better the fit.

This posterior distribution constrains the model parameters, which in turn are fitted based on the observations. The latter are written as:

$$\mathrm{Obs} = [g, G_{\mathrm{BP}}, r, G, i, G_{\mathrm{RP}}, J_{\mathrm{2M}}, J_{\mathrm{U}}, H_{\mathrm{U}}, H_{\mathrm{2M}}, K_{\mathrm{2M}}, K_{\mathrm{U}}, \varpi] \tag{2.18}$$

where the photometric bands have been arranged in increasing central wavelength order.

Each of the *Gaia* photometric bands are then corrected for saturation (Riello et al., 2021):

- If $2 < G < 8$: $G = G - 0.09892 + 0.059\,G - 0.009775\,G^2 + 0.0004934\,G^3$
- If $2 < G < 3.94$: $G_{\mathrm{BP}} = G_{\mathrm{BP}} - 0.9921 - 0.02598\,G_{\mathrm{BP}} + 0.1833\,G_{\mathrm{BP}}^2 - 0.02862\,G_{\mathrm{BP}}^3$
- If $2 < G_{\mathrm{RP}} < 3.45$: $G_{\mathrm{RP}} = G_{\mathrm{RP}} - 14.94 + 14.41\,G_{\mathrm{RP}} - 4.657\,G_{\mathrm{RP}} + 0.503\,G_{\mathrm{RP}}^3$

Furthermore, the parallaxes ϖ are corrected for the non-zero parallax zero-point, whose value has been shown to depend on magnitude, ecliptic latitude, and the pseudocolour of the source, using the prescription from Lindegren et al. (2021b).

To these observations are added their associated uncertainties:

$$\sigma = \left[\sigma_g, \sigma_{BP}, \sigma_r, \sigma_G, \sigma_i, \sigma_{RP}, \sigma_{J_{\mathrm{2M}}}, \sigma_{J_{\mathrm{U}}}, \sigma_{H_{\mathrm{U}}}, \sigma_{H_{\mathrm{2M}}}, \sigma_{K_{\mathrm{2M}}}, \sigma_{K_{\mathrm{U}}}, \sigma_{\varpi}\right] \tag{2.19}$$

The photometric uncertainties combine their standard measurement uncertainty and a systematic error. The latter value is taken as 0.03 mag for g, r and i (Barentsen et al., 2014; Drew et al., 2014), 0.01 mag for G, G_{BP}, G_{RP} (Riello et al., 2021), 0.03 mag for J_{2M},0.02 for H_{2M} and K_{2M} (Skrutskie et al., 2006), and 0.03 mag for J_{U}, H_{U}, K_{U} (Hodgkin et al., 2009).

The best fit model minimizes the negative logarithm of the likelihood:

$$\ln(P(Obs|\theta)) = -\frac{1}{2}\sum_{i}^{n}\frac{(\mathrm{Obs(i)} - \mathrm{Model(i)})^2}{s_i^2} + \ln(s_i) \tag{2.20}$$

To perform the SED fitting, I use a Markov Chain Monte Carlo (MCMC) process to explore the posterior distribution and identify the best-fitting parameters. This algorithm draws a certain number of samples from a posterior probability density, given a prior and a likelihood (Foreman-Mackey et al., 2013).

First I choose a number of walkers, corresponding to the size of the ensemble, which are initially distributed randomly across the parameter space. In later iterations the parameter space distribution for a given walker is based on the current positions of the other walkers, thereby translating as some form of mutual dependence. They randomly walk across the parameter space, where at each position, the maximum likelihood is computed. In the end are computed chains containing the values of the parameters for all the steps of the walk. Once enough iterations have passed, the chains will stay around the best-fitting regions of the parameter, thereby showing convergence.

This process should be run long enough that the walkers ‘forget’ their initial positions and end up in a distribution that follows the probability distribution for the model parameters providing a fit to the data. A certain fraction of the number of iterations (the ‘burn-in’) from the beginning of the walker chains are therefore discarded at the end of the simulation, to remove this initial period before which the walkers have forgotten their initial positions. This will suppress the first iterations, and therefore forget the initial estimates for a better accuracy of the results (Foreman-Mackey et al., 2013; Smith, 2016). For an optimal exploration of the parameter space, the combination of 1000 walkers, 100 burnin and 1000 iterations have been selected.

The MCMC process will result in 1,000,000 samples of the posterior distribution (the values of the posterior distributions for each walker at each iteration step, minus the burnin). I use the median of this distribution as the fitted value for $\log(M/M_{\odot})$, Fr(Age), d and $\ln(f)$, and the 16th and 84th percentiles as respective lower and upper bounds. $\log(T_{\rm eff})$ and $\log(L/L_{\odot})$ are then calculated for the entire posterior distribution through an interpolation of the evolutionary models from Ekström et al. (2012).

2.5 Comparison with spectroscopic temperatures

To test the efficiency and reliability of the SED fitting code, I chose the working sample in Cygnus (see Chapter 3) and searched for reliable spectroscopically-observed stars. I crossmatched this sample with the SIMBAD database with a matching radius of 1 arcsec and found 2672 matches with spectral types. From this list, I focused only on those with a clearly identified spectral type, a reference, and with a quality

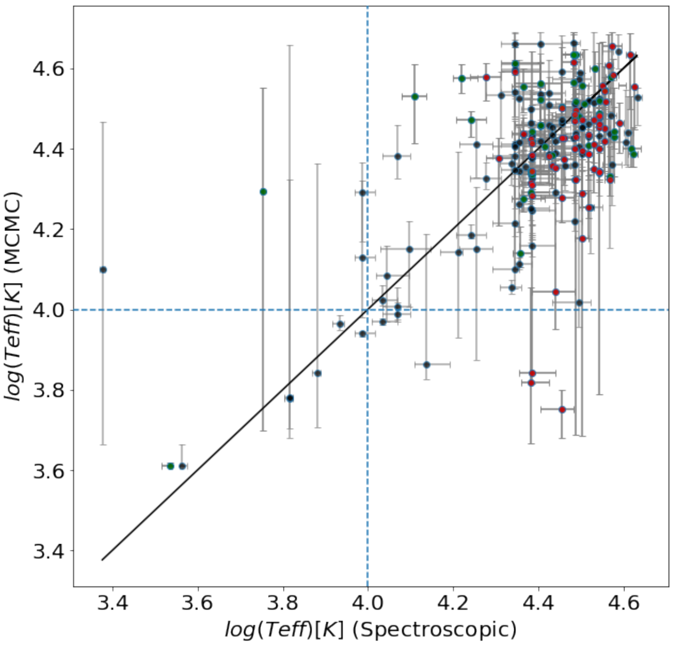


Figure 2.8: Comparison between the literature spectroscopic temperature with that from our SED fits, for the 205 objects with reliable spectral types. Blue dashed lines correspond to the 'limit' at $\log(T_{\text{eff}}/K) = 4$ (dividing OB stars from cooler stars). Green sources are from Blaha & Humphreys (1989) while red ones are from Berlanas et al. (2019).

measurement of 'A', 'B' or 'C' ('D' and 'E' are considered too low quality).

I determined the effective temperatures from the spectral type using the tabulations from Martins et al. (2005) for the O-type stars (observed scale), from Trundle et al. (2007) for early B-type stars, from Humphreys & McElroy (1984) for late B-type stars of luminosity classes 'I' or 'III' and from Pecaut & Mamajek (2013) for the others, interpolating between these tabulations where they overlapped. I required at least a spectral type and a spectral subclass, used the spectral type of the primary star for binaries, assumed a luminosity class of 'V' when unspecified, interpolated for luminosity classes 'II' and 'IV', and chose error bars of one spectral subclass. This process reduced the number of stars with spectroscopic effective temperatures to 205 objects, including 44 from Blaha & Humphreys (1989) and 60 from Berlanas et al. (2019).

Fig. 2.8 shows a comparison between the spectroscopic effective temperatures and those from the SED fits, and reveals that very few cool stars are identified as hot stars by the SED fitting process. I define the recovery rate as $\mathrm{RR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$ where TP stands for the number of true positives and FN corresponds to the number of false negatives. The recovery rate is 97% (96%, 92% and 86%) for stars with SED-fitted $\log(T_{\mathrm{eff}}/K) > 4.0$ (4.1, 4.2, 4.3), and the contamination rate ($\mathrm{CR} = \mathrm{FP}/(\mathrm{TP} + \mathrm{FP})$, where FP is the number of false positives) is between 3 and 14%, depending on the temperature threshold used. It is clear that both these parameters worsens at higher temperature thresholds, however they still remain satisfactory results.

Exploring the fits in more detail, I find that sources with a large extinction ($A_V > 4$ mag) typically have higher $\ln(f)$ values (> -5) and that these corresponded to stars with poorer fitted temperatures (when compared to literature values) and larger uncertainties on the fitted quantities. As an example of this, if I focus the comparison on the 44 objects from Blaha & Humphreys (1989) with literature spectral types (typical extinction of A_V = 1-2 mag, as these are primarily from the less-reddened associations in Cygnus), I find a recovery rate of 92–100%, while if I limit my comparison to the 60 objects in Cyg OB2 (typical extinction of 4-5 mag) listed by Berlanas et al. (2019) the recovery rate drops to 85–95%, slightly inferior. It also has to be noted that this sample is skewed towards hot stars, historically more analysed in the massive OB associations in Cygnus.

Eventually, I explore the impact of using a single value of $\log g$ in the fitted stellar spectral models. This was necessary because of the difficulty of finding models covering the full range of T_{eff} and $\log g$, and with a sufficient wavelength coverage. I estimate the bias with a fixed value of $\log g$ introduced by comparing the effective temperatures and luminosities derived from our SED fits with those derived spectroscopically, for the 199 sources with spectroscopic surface gravities. I divide this sample into stars with low ($\log g < 3.5$) and high ($\log g > 3.5$) surface gravity. I find that stars with a low surface gravity had their SED-fitted effective temperatures over-estimated by $\sim$0.1 dex, while their luminosities were under-estimated by $\sim$0.01 dex. Both of these effects will cause evolved stars with low surface gravities to be fitted closer to the ZAMS than they should be, under-estimating their ages.

2.6 Examples of good SED fits

To illustrate individual SED fits, I show the results for three different sources:

- In Fig. 2.9 is displayed the SED fit for J202147.34+372631.9, a B0.5III star (Massey et al., 2001), quite reddened ($A_V = 4.14^{+0.09}_{-0.24}$ mag) from the extinction map of Green et al. (2019) and with reasonable error bars. Although the SED fitting process placed the star rather close to the main-sequence for its spectroscopic luminosity class, it had a successful recovery of its effective temperature when only using *Gaia* and 2MASS photometry.
- J235714.76+612314.1 is a B5III star (Martin, 1972), whose SED fit is illustrated in Fig. 2.10. Again this star is constrained with reasonable error bars only using *Gaia* and 2MASS photometry, placing the star at an evolved state just like the spectroscopy suggests.
- The SED fit of J003514.47+614626.9 is illustrated in Fig. 2.11. It is a F7V star (Grieves et al., 2018) and shows a good fit to the observed SED and parallax.

2.7 Examples of problematic SED fits

During the SED fitting process I identified certain situations that would produce a poor or problematic SED fit due to one reason or another. I list these situations as follows:

- *Variable stars*: This occurs when a source is characterized by some level of variability. For example, magnitudes from *Gaia* will be different from IGAPS or magnitudes of 2MASS will be different from UKIDSS. A typical example of

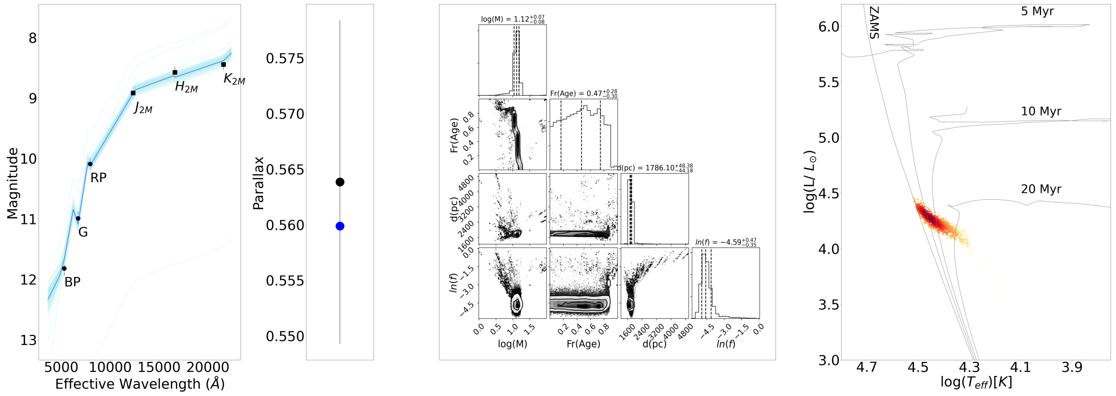


Figure 2.9: SED fit for J202147.34+372631.9 (Cl Berkeley 87 32), fitted with $\log(T_{\text{eff}}/K) = 4.44^{+0.04}_{-0.07}$, compared with a spectroscopic $\log(T_{\text{eff}}/K) = 4.42 \pm 0.04$ (Massey et al., 2001). This includes the SED fit, the triangle plot for the model parameters, and the posterior distributions of $\log(T_{\text{eff}})$ and $\log(L/L_{\odot})$, colour-coded by density in an HR diagram with a zero-age main-sequence and isochrones from Ekström et al. (2012).

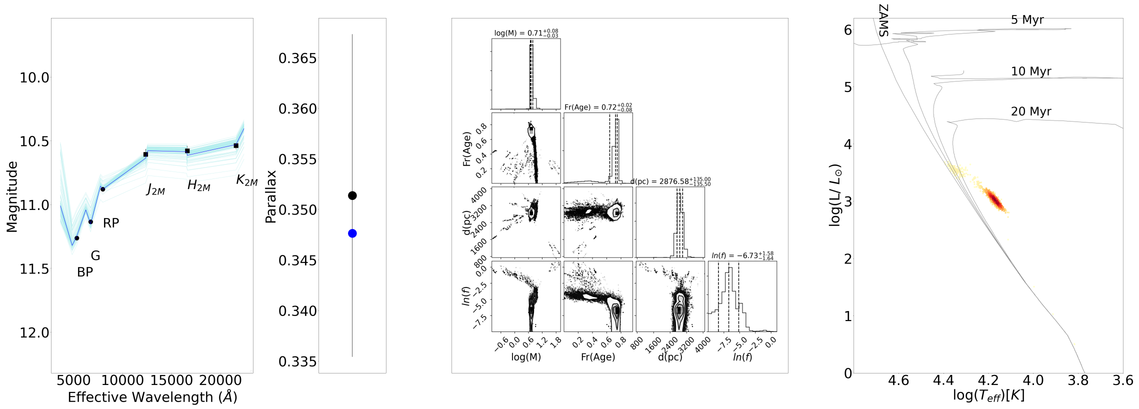


Figure 2.10: SED fit for J235714.76+612314.1 (LS I +61 108), fitted with $\log(T_{\text{eff}}/K) = 4.18^{+0.05}_{-0.03}$, compared with a spectroscopic $\log(T_{\text{eff}}/K) = 4.21^{+0.04}_{-0.01}$ (Martin, 1972).

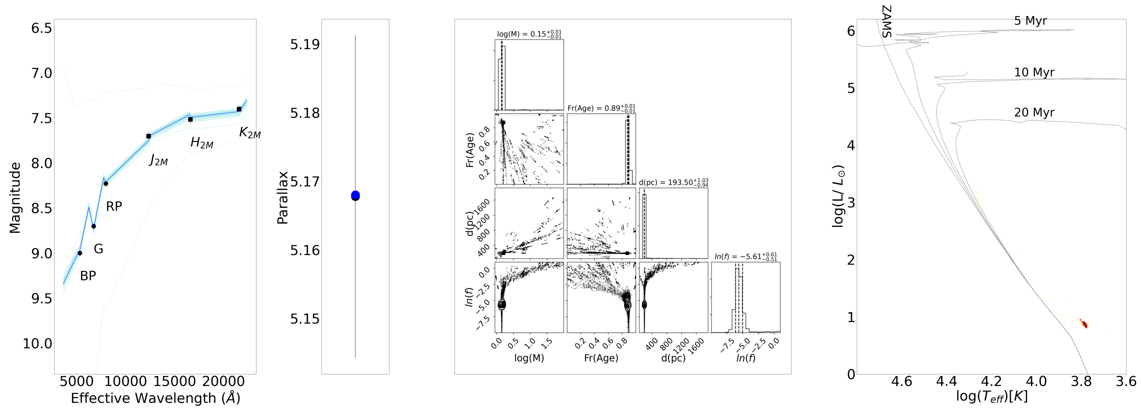


Figure 2.11: SED fit for J003514.47+614626.9 (HD 3130), fitted with $\log(T_{\rm eff}/K) = 3.79^{+0.02}_{-0.01}$, compared with a spectroscopic $\log(T_{\rm eff}/K) = 3.80^{+0.00}_{-0.01}$ (Grieves et al., 2018).

this case is J200223.93+354113.5, as shown in Fig. 2.12. There, the IGAPS photometry was fainter than the *Gaia* one, preventing an accurate fit.

- *Bad extinction*: This happens when there are two possible SED fits for the same observed SED, and the fitting code could not converge in spite of the supplementary iterations. Typically, the code cannot decide whether the observed SED corresponds to an intrinsically cool star or a reddened hot star. The most common cause of this appears to be that the extinction from Green et al. (2019) is incorrect for a specific star. An example is illustrated for J201923.86+364346.6 in Fig. 2.13, where the error bars are very large. The SED fitter wants to fit a higher extinction to correctly fit the very red spectrum, but the extinction map does not allow the star to have such a high reddening.

- *Incorrect parallax*: This occurs when, while the observed SED is well fitted, the observed parallax is not, typically in cases of stars with very high reddening. For instance, Fig. 2.14 shows the SED fit for J203613.07+412807.2, with $A_V = 10.65^{+0.53}_{-2.71}$. The model parallax is much lower than the observed parallax, placing the star at a higher distance, resulting in a double peak in the posterior distribution for the distance, as visible in the triangle plot.

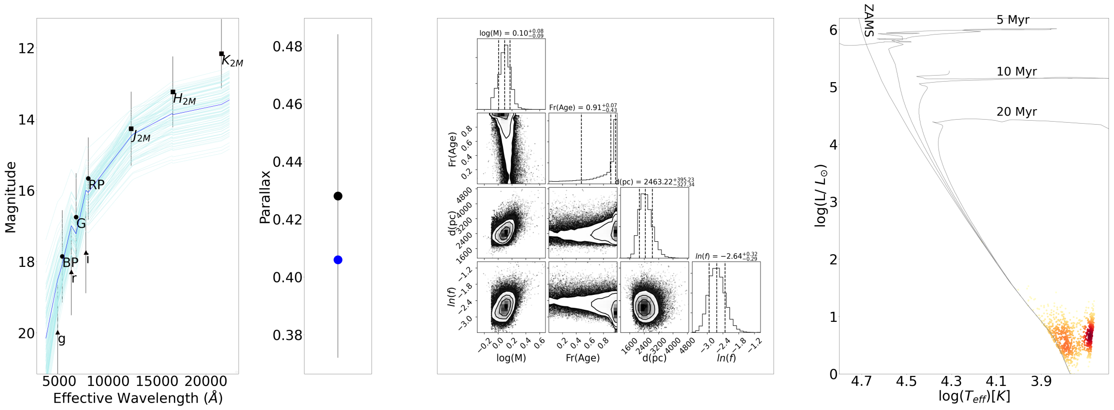


Figure 2.12: SED fit for J200223.93+354113.5, fitted with $\log(T_{\rm eff}/K) = 3.74^{+0.08}_{-0.06}$.

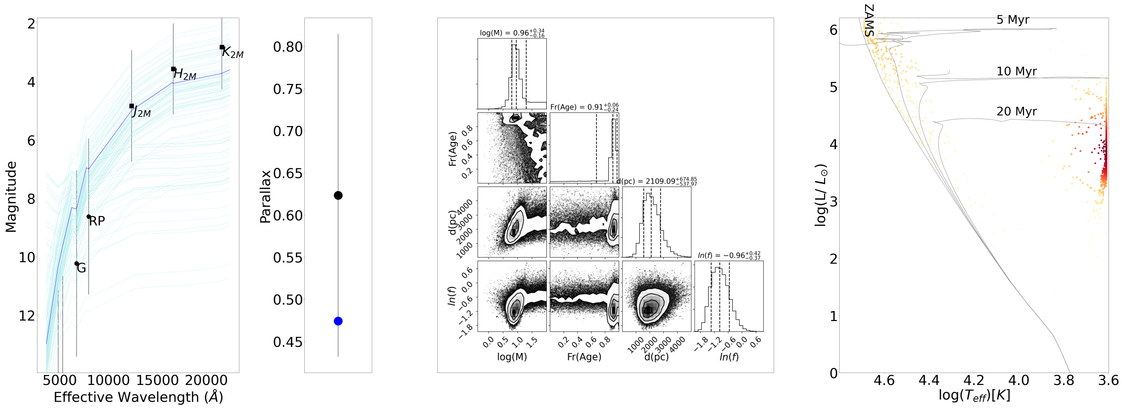


Figure 2.13: SED fit for 201923.86+364346.6, fitted with $\log(T_{\rm eff}/K) = 3.61^{+0.87}_{-0.00}$.

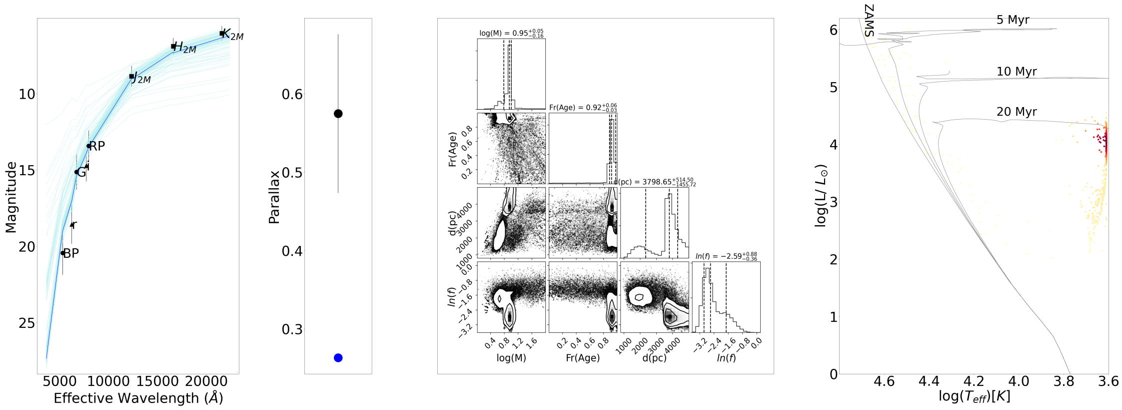


Figure 2.14: SED fit for J203613.07+412807.2, fitted with $\log(T_{\rm eff}/K) = 3.61^{+0.05}_{-0.00}$.

2.8 Improved versions of the SED fitting process

In this section I outline how the SED fitting process was improved throughout the thesis. Exploring different massive star-forming regions revealed the requirement to adapt the SED fitter, along with the availability of the new data to compare and constrain the models.

2.8.1 Supplementary iterations (version 1.1)

Throughout this work I have continuously applied the SED fitting code to several samples (as described in Chapters 3, 5 and 6), and improved it by doing so. Some sources have large uncertainties on temperatures because they can be fitted by both a hot star and a cool luminous star, implying that they require more iterations to fully converge.

To overcome this issue, while keeping the same number of walkers, I have increased the number of burnin iterations to 200 but lowered the number of post-burnin iterations to 200, since 200,000 chains turned out to be enough to explore the posterior distribution. For each source, once these iterations were complete, I checked that both criteria $\log(T_{\text{eff}})P95 - \log(T_{\text{eff}})P5 < 0.5$ and $\ln(f) < -4$ are fulfilled, where I have computed the 5th and 95th percentiles on effective temperature. If either of these conditions fail it is an indication that the resulting fit would have large uncertainties, typically due to a multi-peak posterior distribution. In which case the SED fitter performs an additional 1000 burnin iterations and 200 post-burnin iterations. I resume the SED fitting, and continue to do so until convergence is achieved. I halt this process after 6000 supplementary iterations as it means that convergence is not possible within a reasonable amount of time.

To test this new version, I selected 2000 stars in the final sample of study in the Cygnus region (see Chapter 3). With the version 1.0 of the SED fitting code, 349 had either $\log(T_{\text{eff}})P95 - \log(T_{\text{eff}}P5 < 0.5$ or $\ln(f) > -4$, hence about 17 % of the sample. This number is reduced to 236 (hence about 12 % of the sample) when applying this new version of the SED fitting code.

2.8.2 Scaled extinction (version 1.2)

2.8.2.1 Extinctions from Gaia DR3

The latest data release from *Gaia* was published on June 2022 (Gaia Collaboration et al., 2023a; Babusiaux et al., 2023), during the later stages of my PhD. Amidst this abundance of new data, Gaia DR3 provides astrophysical parameters (Creevey et al., 2023), and notably extinctions, in three different relevant ways:

1. *GSP-Phot* provides the monochromatic extinction A_0 along with the extinctions of the *Gaia* photometric bands. The extinction is constrained by fitting the observed parallax, the G magnitude and the BP-RP spectra are fitted through an MCMC fit. The extinctions are based on the extinction laws from Fitzpatrick (1999) while the model SEDs are integrated from different stellar spectra libraries (Creevey et al., 2023).

2. *ESP-HS* takes advantage of the spectral classification performed by *Gaia* DR3 for the hot stars in the catalogue (effective temperature higher than 7500 K). The monochromatic and *Gaia* extinctions are estimated from the BP-RP spectra together with the radial velocities. These extinctions turn out be generally slightly larger than the *GSP-Phot* extinctions for OB stars (Creevey et al., 2023).

3. *The Total Galactic Extinction map (TGE)* provides an all-sky integrated 2D extinction map for Gaia DR3 sources. To do so, it used the giant stars from Andrae et al. (2022) due to their high brightness, with the effective temperatures and absolute G magnitude from *GSP-Phot*. They required these extinction tracers to be at least 300 pc from the Galactic plane and are associated to a HEALPix, i.e. a pixelisation where a spherical surface is divided such that all pixels cover an identical surface area (Gorski et al., 1999), whose level varies

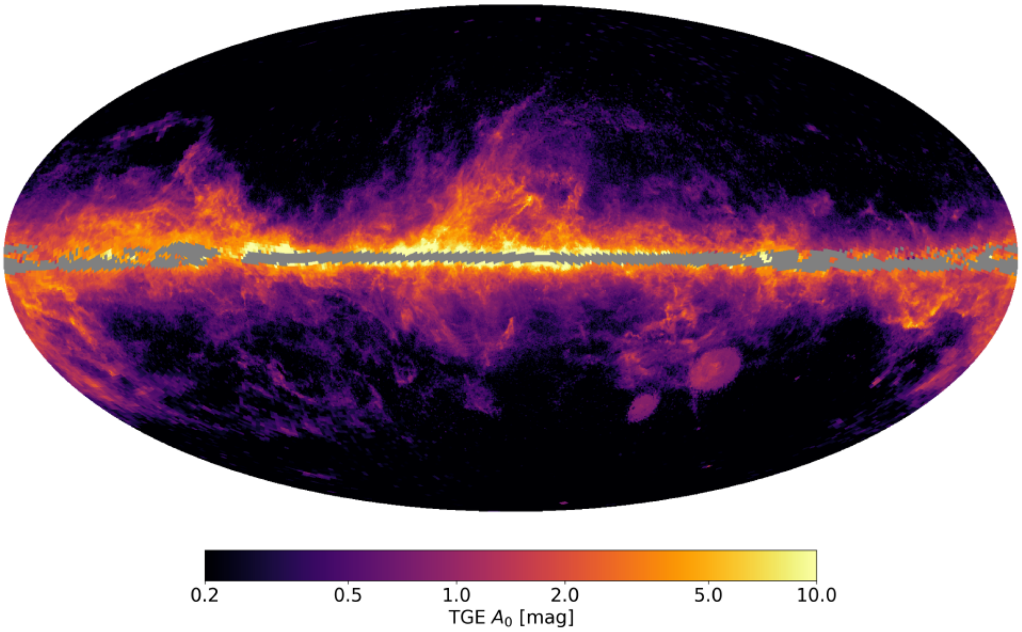


Figure 2.15: All-sky extinction map from Gaia DR3 made with the optimum HEALPixes levels. This figure has been taken from Delchambre et al. (2023).

from 6 (0.839 deg^2) to 9 (0.013 deg^2). A HEALpix will have an assigned extinction only if three tracers or more were available. The optimum HEALPix is then designated as the set of lowest HEALPixes with ten tracers or more (such that the extinction calculation is trustworthy). If the level 6 HEALPix has less than ten tracers, then it is automatically considered as optimum. This extinction map is shown in Fig. 2.15 and reveals an incomplete coverage of the Galactic plane ($|b| < 5°$). Indeed, while the Galactic plane coverage is 97.2 % for the level 6, it is only 33.7 % for the level 9, because the number of tracers is not high enough (Delchambre et al., 2023).

Delchambre et al. (2023) also compared the TGE extinctions (A_0) with the *Bayestar* extinctions (A_V) from Green et al. (2019), and estimated a linear fit of 1.20 ± 0.04. As pointed out in Andrae et al. (2022), Green et al. (2019) tend to underestimate extinction compared with other extinction maps. This encouraged me to modify the SED fitting process.

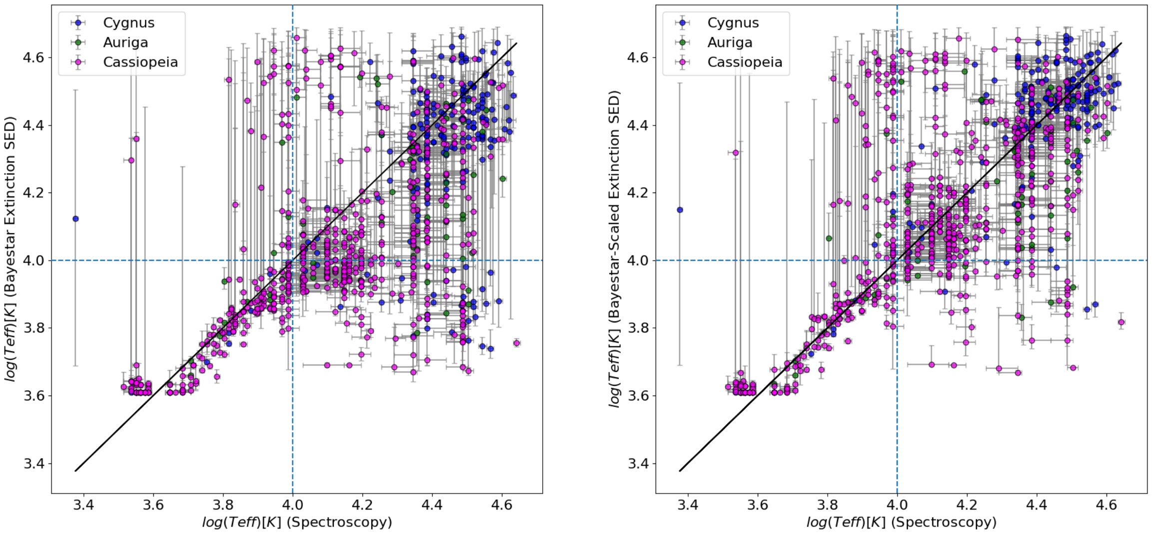
2.8.2.2 New extinctions

To test which extinctions should be used, I created three different sets of extinctions. The first one includes the original *Bayestar* extinctions from Green et al. (2019). The second one contains the same *Bayestar* extinctions, but increased by 22%. This factor arises from Delchambre et al. (2023), with A_0 converted to A_V from $\frac{A_V}{A_0} = 0.98$ (Sale et al., 2014; Green et al., 2018). The last one uses directly the extinctions from *Gaia* DR3 (Creevey et al., 2023), either *ESP-HS* when available (more suited for hot stars), or *GSP-Phot* otherwise. I do not utilise directly the *Gaia* DR3 extinction map due to its incomplete coverage of the Galactic plane (see Fig. 2.15, Delchambre et al. 2023).

Each of these samples was applied to a spectroscopic catalogue to ensure a temperature comparison similar to that in Section 2.5. This catalogue includes the original Cygnus sample from Section 2.5, along with spectroscopically-reliable sources from Auriga (see Chapter 5) and Cassiopeia (see Chapter 6), the two other regions of study of this thesis. The resulting sample contains 879 stars, but only 532 of them have a *Gaia* DR3 extinction.

Fig. 2.16 shows a comparison between the SED effective temperatures derived from these three sets of extinctions with the spectroscopic effective temperatures extracted from *Simbad*. I use the threshold of $\log(T_{\text{eff}}/K) = 4, 4.1, 4.2, 4.3$, just like in Section 2.5, and the same quality metrics that are the recovery and contamination rates (RR and CR). The comparison is shown in Table 2.1.

These statistics turn out to be worse than the sole Cygnus sample from Section 2.5. One of the reasons for this is the Cygnus sample was dominated by early B-type and O-type stars, far from the lower temperature thresholds, while the Auriga and Cassiopeia samples have more late B-type stars that are closer to the temperature thresholds used. These results also show a higher contamination rate, but this is mostly due to the supergiants and bright giants that have been observed in Cassiopeia, therefore skewing the total spectroscopic sample. Removing these stars from the sample



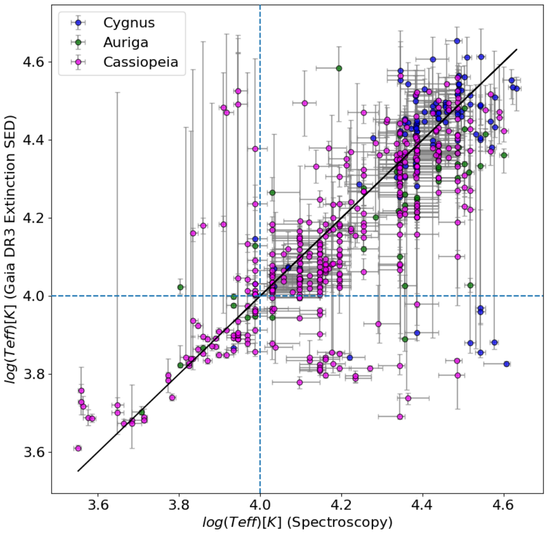


Figure 2.16: Comparison between the spectroscopic temperature from *Simbad* in abscissa and the SED temperatures panels in ordinate. Top left, top right and bottom left panels show respectively the results for *Bayestar*, scaled *Bayestar* and *Gaia* DR3 extinctions. The dots have been colour-coded following the regions they are from.

indeed lowers considerably the contamination rates (they become lower than 10 % for every temperature threshold), but the reason why their temperatures are overestimated by the SED fitting process still needs to be investigated.

From these results, it is also clear that scaling the *Bayestar* extinctions increases

Table 2.1: Recovery and contamination rates for the different temperature thresholds. The first set corresponds to the results using the *Bayestar* extinctions, the second using the scaled *Bayestar* extinctions and the third using the *Gaia* DR3 extinctions.

	$\log(T_{\text{eff}}/K) = 4$	$\log(T_{\text{eff}}/K = 4.1$	$\log(T_{\text{eff}}/K) = 4.2$	$\log(T_{\text{eff}}/K) = 4.3$
RR1 (%)	68.71	58.82	61.14	53.15
CR1 (%)	7.41	11.84	15.69	21.40
RR2 (%)	89.78	78.86	79.36	73.15
CR2 (%)	11.35	17.97	21.36	23.83
RR3 (%)	89.44	80.05	86.72	77.19
CR3 (%)	5.91	8.23	5.55	8.56

the recovery rate by $\sim$20 %. The SED temperatures from *Gaia* DR3 give slightly better results than the scaled *Bayestar* extinctions, but the former catalogue contains fewer of the problematic objects (see Fig. 2.16). However, the *Gaia* DR3 extinctions have a partial coverage, therefore the *Bayestar* extinctions are better suited for large-scale SED fitting, especially in the Galactic plane. Utilising *Gaia* DR3 extinctions would require us to extract them from different sets of measurements (i.e. *GSP-Phot* and *ESP-HS*) while the *Bayestar* extinctions come from only one set and are consequently more homogeneous.

The version 1.2 of the SED fitting process now contains the *Bayestar* extinctions increased by 22 %.

2.9 Comparison with APOGEE temperatures

Finally, to fully test the reliability and quality of the improved SED fitting code, I decided to compare the results with APOGEE. As a large-scale spectroscopic survey with data collected in the near-infrared, its sample of high-resolution spectra of stars across the Milky Way makes it ideal for comparison (Majewski et al., 2017).

Table 2.2: Recovery and contamination rates for the different temperature thresholds between the SED-fitted and the APOGEE temperatures of the comparison sample. The first statistics correspond to the results for the whole sample, while the following ones specifically show them for the Cygnus (1), Auriga (2) and Cassiopeia (3) subsamples.

	$\log(T_{\text{eff}}/K) = 4$	$\log(T_{\text{eff}}/K) = 4.1$	$\log(T_{\text{eff}}/K) = 4.2$
RR (%)	84.79	73.98	64.24
CR (%)	23.31	33.89	38.99
RR1 (%)	72.95	60.50	63.01
CR1 (%)	32.16	28.00	28.12
RR2 (%)	87.96	77.92	40.90
CR2 (%)	19.23	34.78	62.50
RR3 (%)	89.80	84.55	75.00
CR3 (%)	21.03	36.96	40.84

For this work I specifically exploited data release DR17 (García Pérez et al., 2016; Abdurro'uf et al., 2022). I have selected the sources in APOGEE with a measured T_{eff} from the pipeline, but have discarded those whose T_{eff} measurement indicated a warning. I do not consider these results reliable because of their proximity with the upper limit of APOGEE measurements (20,000 K). This means that the highest temperature threshold for comparison is $\log(T_{\text{eff}}/K) = 4.2$, and that my SED fitting code can study hotter stars than some spectroscopic surveys.

Crossmatching my sample within 1" with APOGEE DR17 and applying the warning criterion, I obtained a sample of 1143 stars, including 335 in Cygnus, 331 in Auriga and 477 in Cassiopeia. The comparison between the scaled *Bayestar* SED-fitted temperatures and the APOGEE temperatures is displayed in Fig. 2.17. Again I have used the the recovery and contamination rates (RR and CR) to assess the quality of the calibration, as shown in Table 2.2.

Table 2.2 shows higher contamination rates when comparing the SED-fitted temperatures with APOGEE than with SIMBAD. It also shows a worse recovery rate at

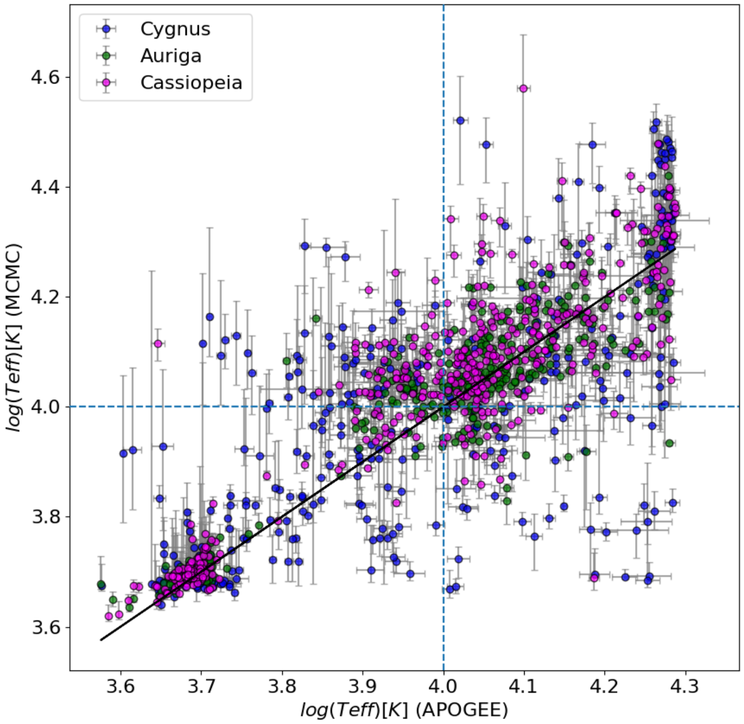


Figure 2.17: Comparison between the APOGEE and the new MCMC temperatures for the comparison sample of 1143 stars.

lower temperature thresholds for the Cygnus sample than for the Auriga and Cassiopeia samples, which is surprising given the previous results. The Auriga sample has a worse recovery rate at the highest temperature threshold, which could be due to the sparsity of very hot stars in this region compared with Cygnus and Cassiopeia.

3 Revisiting the Cygnus OB associations

This third chapter presents the first results and application of the SED fitting process. Specifically, it is dedicated to a new exploration of the OB associations in Cygnus and is based on the work published in Quintana & Wright (2021). I outline how I selected and filtered out the working sample in the Cygnus region, illustrate the results of the SED fitting and detail the method used to identify new kinematically-coherent OB associations, before analysing them.

3.1 Historical Cygnus OB associations

Members of the historical Cygnus OB associations, as in other regions, were mainly selected through the photometry of their brightest members, with the observations and techniques available several decades ago (see e.g. Humphreys 1978). Since then, while new telescopes offered the possibility of studying the kinematics of OB associations (de Zeeuw et al., 1999), the membership of some of the known OB associations remained similar and continues to be used (Melnik & Dambis, 2020). While the membership of Cyg OB2 has been revised many times (Knödlseder, 2000; Wright et al., 2015; Berlanas et al., 2018b; Orellana et al., 2021), it is not the case for the other Cygnus OB associations.

Proper motions and parallaxes from *Gaia* (Gaia Collaboration et al., 2016) allow a new exploration of these OB associations to verify their validity. To that end, I have used the data from *Gaia* EDR3 (Gaia Collaboration et al., 2021), giving the most up-to-date astrometry.

I specifically focus on a 60 deg^2 area inside the Cygnus constellation, spanned by l = [71°,81°] and b = [-1°, 5°]. In this region shine the OB associations Cyg OB1, OB2, OB3, OB8 and OB9, the five most well-analysed associations in Cygnus, particularly Cyg OB2. 178 OB stars were identified by Blaha & Humphreys (1989) and divided between these five associations based only on their positions on the sky and photometry, and the same members were used by Melnik & Dambis (2020). Cyg OB4, OB5, OB6 and OB7, the other Cygnus OB associations, are not considered here because they are either much more extended or located at higher longitudes (see Chapter 1). Consequently, this choice of coordinates allows me to analyse an area of reasonable size. For this work I use the list of 164 OB stars for Cyg OB1, OB3, OB8 and OB9 from Blaha & Humphreys (1989). For Cyg OB2 I use the updated census of OB stars from Berlanas et al. (2019). Combining them yielded a total of 341 OB stars.

I crossmatched this catalogue of Cygnus OB association members with *Gaia* EDR3 data, using a matching radius of 1 arcsec. 329 sources had a successful match for further analysis. To filter this data, I have applied the recommended astrometric criterion from Lindegren et al. (2018): the renormalised unit weight error (RUWE) of a source must be smaller than 1.4. 269 sources fulfilled this condition. Distances of sources with good astrometry were obtained from the geometric distances of Bailer-Jones et al. (2021).

Fig. 3.1 shows the distribution of these stars in Galactic coordinates (where all the 329 stars are displayed), while Fig. 3.2 shows them in distance plotted against l and Fig. 3.3 in Galactic proper motions (where only the 269 sources that passed the astrometric quality tests are shown as the others are considered unreliable). Most members are found at distances between 1.5 and 2.0 kpc with a slight trend of increasing distance with decreasing Galactic longitude. As a result, Fig. 3.1 already illustrates that some stars initially included within the associations turn out to be located either closer or further away than the other association members, and as such can be considered as contaminants. It shows as well that, in this region, higher extinctions are encountered at higher longitudes and lower latitudes, reaching their maximum around Cyg OB2. From Fig. 3.3, Cyg OB2 is the most distinct and highly concentrated of the associations, contrary to the other ones (although Cyg OB3 exhibits some level of kinematic coherence).

I attempted to refine the association membership, as originally defined by Blaha & Humphreys (1989), by removing outliers in distance and proper motion, and also applied a number of modern clustering algorithms (notably DBSCAN, see Pedregosa

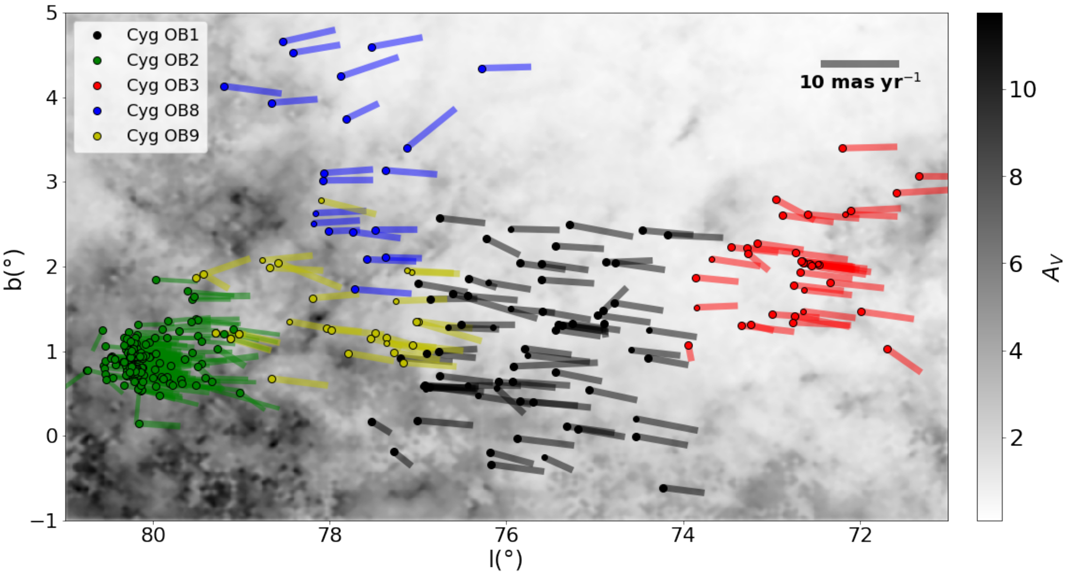


Figure 3.1: Top left: spatial distribution of Cygnus OB association members from Blaha & Humphreys (1989) and Berlanas et al. (2019), with their PMs as vectors, and a background extinction map from Green et al. (2019), representing the integrated reddening up to a distance of 1.8 kpc. Objects with $RUWE > 1.4$ have been displayed as small circles.

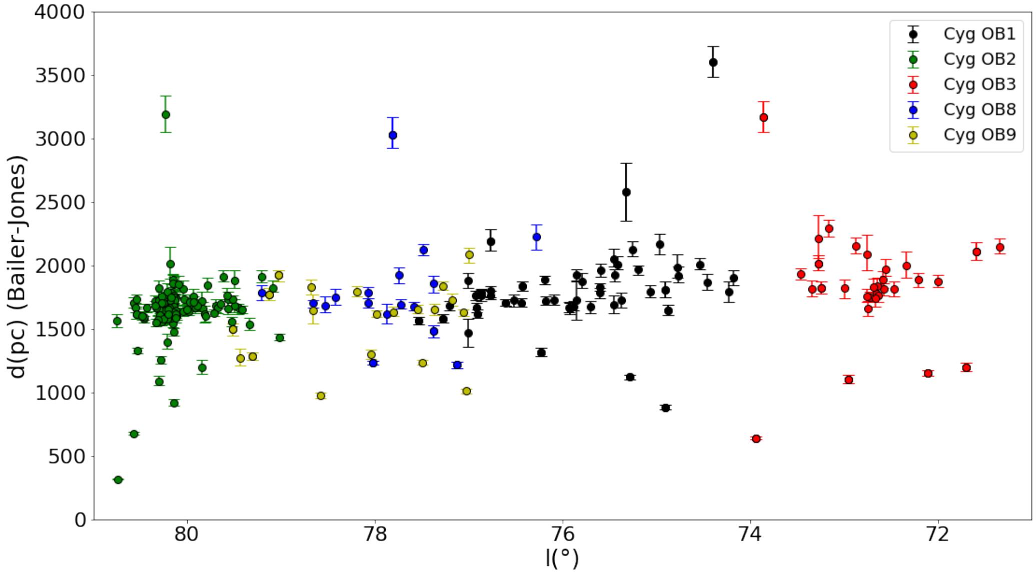


Figure 3.2: Distances from Bailer-Jones et al. (2021) for Cygnus OB association members with $RUWE < 1.4$ as a function of Galactic longitude.

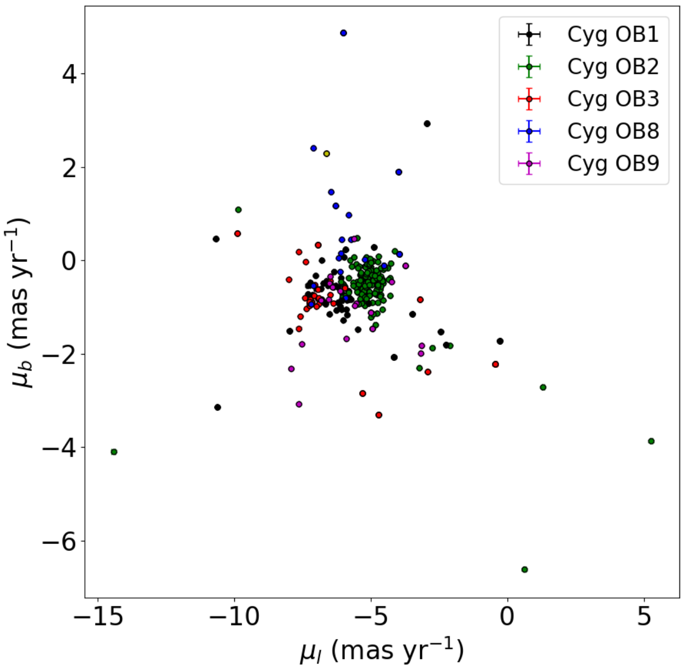


Figure 3.3: Galactic PM distribution for the Cygnus OB association members for sources with $RUWE < 1.4$.

et al. 2011) to this list of OB stars to "re-discover" these associations. However, only Cyg OB2 and OB3 are characterized by some level of kinematic coherence or could be recovered in some way using DBSCAN that would suggest they are true associations. It implies that Cyg OB1, OB8 and OB9 are probably not genuine OB associations and should consequently be revisited.

There have been numerous studies of Cyg OB2 (see e.g. Knödlseder 2000; Wright et al. 2015; Berlanas et al. 2019; Orellana et al. 2021), therefore it is unsurprising that it turns out to be the most high-confidence OB association in this region. In contrast, individual studies about the other Cygnus OB associations are lacking (see Chapter 1). This means that the best option to analyse the Cygnus OB associations is to identify new associations, composed of OB stars identified with my own method to avoid bias. This will be the focus of this chapter.

3.2 Identifying and characterising OB stars

This section details the selection of the sample of study in the Cygnus region and presents the general results of the SED fitting process on this working sample.

3.2.1 Data and selection process

This study uses the twelve photometric bands from *Gaia* EDR3, IGAPS, 2MASS and UKIDSS introduced in Chapter 2, along with the observed parallax from *Gaia* EDR3 corrected for the zero-point from Lindegren et al. (2021b)

There are a total of 8,474,415 sources in *Gaia* EDR3 in the Cygnus region defined in Section 3.1, 7,434,799 of which have a valid five or six-parameter astrometric solution. Two-parameter astrometric solutions only contain information about equatorial coordinates, without any observed parallaxes and proper motions, and are therefore not usable as I aim at conducting a kinematic analysis.

This sample was crossmatched with 2MASS, UKIDSS and IGAPS with a matching radius of 1". While both 2MASS and IGAPS have J2000 equatorial coordinates that allow a direct correspondence, I projected the coordinates from *Gaia* EDR3 into averaged UKIDSS epochs with *Gaia* proper motions.

I found near-IR matches for 6,347,730 sources and IGAPS matches for 6,158,261 sources. I removed the 1,087,069 *Gaia* sources that lacked a near-IR counterpart (the combination of visible and near-IR SED provides a better $T_{\rm eff}$ accuracy, especially for reddened objects). These sources generally had $G > 18$ mag, too faint to be detected by 2MASS and lying in one of the areas of incomplete coverage for UKIDSS. They were unlikely to correspond to OB stars in the region of interest.

I discarded bad photometry according to the following criteria. For IGAPS, photometry was considered reliable if saturation was avoided and if its associated class indicated a star or probable star (Monguió et al., 2020). For *Gaia*, the condition is $|C^*| < 3\,\sigma_{C*}$ where C^* stands for the corrected excess flux factor in the $G_{\rm RP}$ and $G_{\rm BP}$ bands and σ_{C*} is a power-law on the G band with a chosen 3σ level (Riello et al., 2021).

For 2MASS, a good quality flag (A, B, C or D) on each photometric band was required, the other quality flags indicating a bad detection (X), lower-quality photometry (E), no reliable estimate of photometric uncertainty (F) or upper limit on magnitude reached (U) (Cutri et al., 2003). For UKIDSS, it was imposed that $ErrBits < 256$, otherwise the number of zero-confidence pixels was too high, and I excluded as well photometry with either $J < 13.25$, $H < 12.75$ and $K < 12$ mag, below which the photometry risks saturation (Lucas et al., 2008). These bad photometric bands had a magnitude set to -999 and their associated uncertainty to 1e8, meaning that they will not influence the SED fit of the object.

At least one good photometric measurement in a blue band (g, G_{BP} or G) and one in a near-infrared band were required, to ensure reliable SED fits. This is a lower limit but it has to be noted that, unsurprisingly, a larger number of good photometric bands will result in a more accurate fit to the observed photometry (except for the rare case of a variable star, see Chapter 2). This step reduced the size of the sample to 5,399,862 sources.

To ensure the quality of the astrometry, I also imposed $RUWE < 1.4$, leaving 5,211,036 sources. Furthermore, too large uncertainties on parallaxes will translate into poorer constraints on distance and luminosity, this is why I added another constraint, that is $|\frac{\varpi}{q*\sigma_{\varpi}}| > 2$. Here σ_{ϖ} is the standard parallax uncertainty, and q is a multiplicative factor that inflates the published standard uncertainty on parallax as it was found to be underestimated, notably for bright sources (El-Badry et al., 2021). This left 2,145,918 sources. [1]

In Section 3.1 it was shown that most of the stars in the Cygnus OB associations are located between 1.0 and 2.5 kpc. I thus removed all sources that fell outside this range of distances using the geometric values calculated by Bailer-Jones et al. (2021), trimming the sample to 896,946 sources. This parameter was chosen as a simplification

---

[1] Maíz Apellániz et al. (2021) offer an alternative method for filtering *Gaia* data and correcting the astrometry that involves relaxing the constraint on RUWE and inflating the parallax uncertainties. This approach typically increased the parallax errors by 10-20%, but would have made the final sample 5% larger (as the RUWE constraint is slightly relaxed). This change would therefore not have had a significant influence on the results.

as higher extinctions may make the photogeometric distance from Bailer-Jones et al. (2021) more reliable, but in any case, the SED fitting process will provide my own estimation of distance.

The primary target of this study are OB stars, encouraging me to limit the sample to sources with absolute magnitudes, $M_K < 1.07$, where $M_K = K - 5\ \log(d) + 5$ with d in pc standing for the distance from Bailer-Jones et al. (2021). This value of M_K corresponds to spectral types earlier than A0V (Pecaut & Mamajek, 2013). For sources that lacked K-band photometry, I used instead $M_H < 1.10$ or $M_J < 1.07$, which offer equivalent criteria in the J and H bands. Either 2MASS or UKIDSS photometry was used for this filtering depending on the availability and quality of the photometric bands. To that end I prioritized the photometric band with the lowest observed uncertainty, which was typically UKIDSS when photometry from both surveys was available. NIR bands were chosen for the cut in absolute magnitude since they are less affected by extinction (Fitzpatrick et al., 2019). With these conditions, the sample was reduced to 47,498 sources.

Finally I exploited the near-IR colour-colour diagram to remove giants by retaining only objects that had larger H-K colours than the dashed line in Fig. 3.4 (projecting UKIDSS photometry onto the 2MASS system with the equations from Hodgkin et al. 2009). This line separates early-type stars of every extinction level from the majority of red giants. Sources without sufficient photometry to be plotted in this diagram were kept in the working sample. Albeit this method clearly discarded many cool giants from the sample, a lot of them will still be included in the final sample, hence the importance of proper temperature and luminosity constraints through the SED fitting process.

This selection process yielded a final sample in the Cygnus region of 20,498 objects. This is the total number of objects that will undergo the SED fitting process. Out of those, from the 152 sources listed by Blaha & Humphreys (1989), 87 passed all these tests and are found in the working sample.

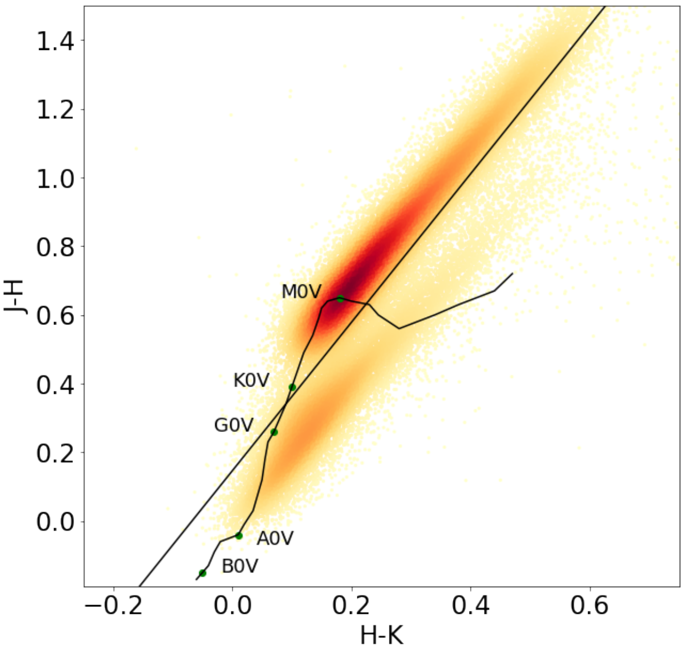


Figure 3.4: Near-IR colour-colour diagram displaying the source density from the photometric and astrometric sample. The black curve corresponds to the unreddened main-sequence from Straižys & Lazauskaitė (2009). The black line, which is parallel to the near-IR reddening vector, is used to separate early-type stars of any reddening (below the line) from red giants (above the line).

3.2.2 General results

Version 1.0 of the SED fitting code described in Chapter 2 was applied to the 20,498 objects in Cygnus and the distribution of their estimated physical parameters is displayed in Fig. 3.5.

The distribution of $T_{\rm eff}$ shows that the sample is clearly dominated by cool red giants, in spite of the previous cuts. 4680 stars have $\log(T_{\rm eff}/K) > 4$ (OB stars, 22.8%), and 818 have $\log(T_{\rm eff}/K) > 4.3$ (O stars, 4.0%), confirming the presence of a large population of hot stars within the region. The distributions of $\log(L/L_\odot)$ and $\log(M/M_\odot)$ reveal, respectively, peaks between 10 and 100 $L_\odot$ and between 1 and 3 $M_\odot$, another strong indicator of the dominance of cool giants.

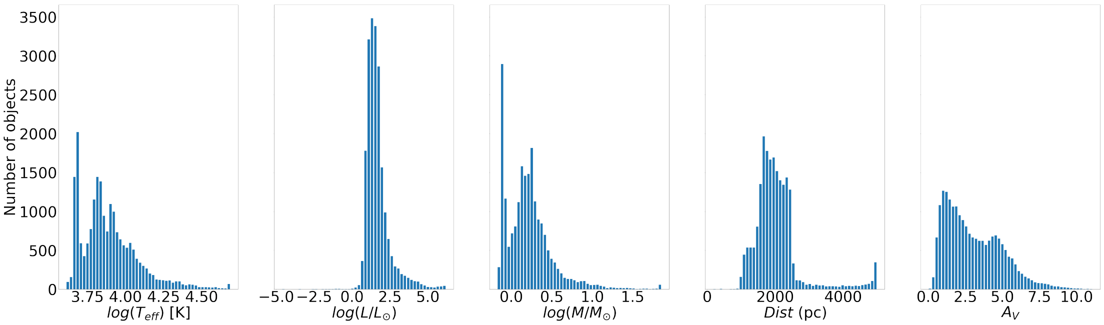


Figure 3.5: Distribution of the median physical parameters derived from the SED fitting process for the 20,498 sources.

As expected from the distance cuts, most of the fitted distances fall between 1.0 and 2.5 kpc, with a peak between 1.5 and 2 kpc, which is the range of distances spanned by the Cygnus OB associations in this area (see Section 3.1). Sources fitted close to the upper prior on distance (5 kpc) are often indicators of poor fits. A few stars with $d > 2.5$ kpc reveal the slight discrepancy between the SED-fitted distances and the geometric distances from Bailer-Jones et al. (2021).

The reddening of the fitted objects, as derived from the extinction map of Green et al. (2019), shows a broad spread from $A_V = 0$ to 6 mag. There are in particular two peaks around $A_V = 1 - 2$ mag and around $A_V = 4 - 5$ mag, early illustrations of the region's extinction structure (the second peak being the level of extinction in the Cyg OB2 area).

3.3 Identification of new OB associations

This section is dedicated to the search for OB associations in the Cygnus region. For this purpose, I utilised a new flexible clustering method, whose results will be contrasted with existing clustering algorithms. This allows me to identify groups without using the historical Cygnus OB associations.

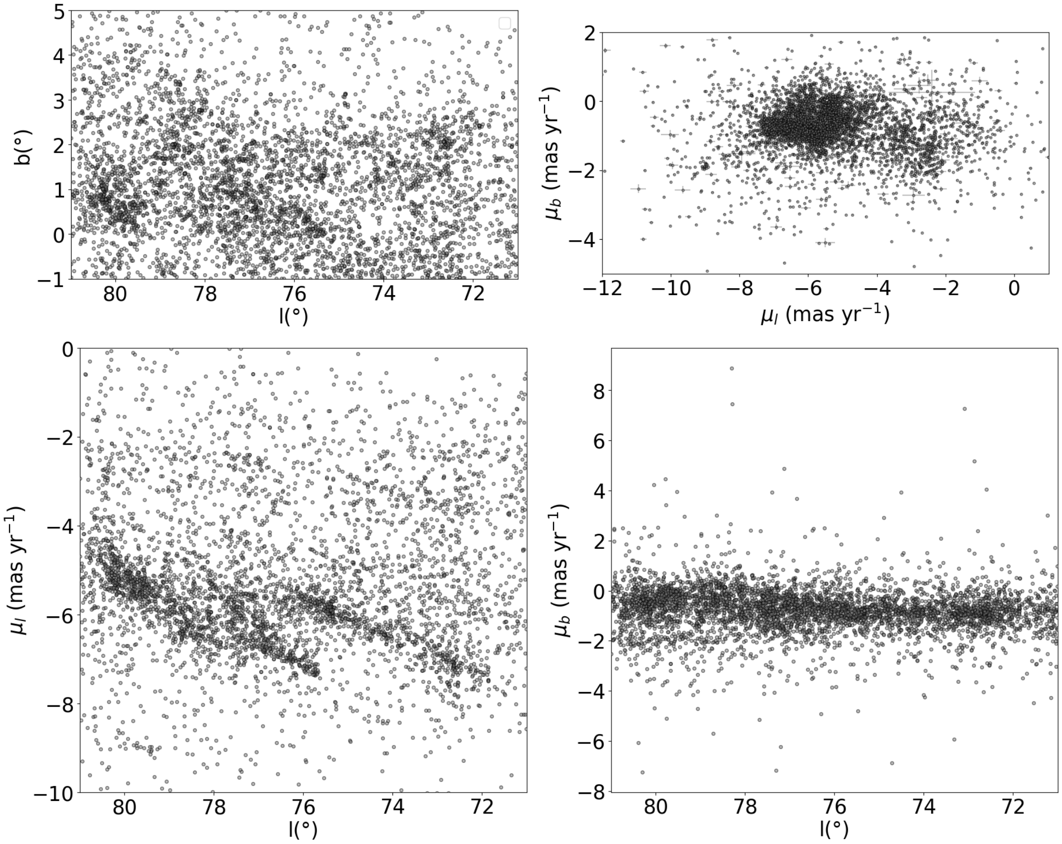


Figure 3.6: Spatial and proper motion distributions of the 4680 objects fitted with $\log(T_{\text{eff}}) > 4$.

3.3.1 Broad kinematics distribution of OB stars

In Fig. 3.6 is displayed the broad spatial and kinematic distribution of the SED-fitted OB stars ($\log(T_{\text{eff}}/K) > 4$). The spatial distribution (upper left panel) shows a broad density gradient as a function of Galactic latitude, as expected, with multiple overdensities that may correspond to OB associations (including one in the position of Cyg OB2 at $l \sim 80°$, $b \sim 1°$). The proper motion distribution (upper right panel) shows that the bulk of the stars have proper motions between μ_l = -8 and -2 mas yr^{-1} and around $\mu_b \sim$ -1 mas yr^{-1}, divided between a main group at the centre and a smaller one at the right. The graph displaying l vs μ_l (lower left panel) reveals two diagonal structures, above and below $l = 76°$, an interesting saw-tooth pattern indicating a correlation between position and velocity that is usually indicative of some form of expansion (I will return to this feature in Section 3.4). No such feature was observed in μ_b (see lower right panel of Fig. 3.6).

The expected motion of stars in this sightline can be estimated through a simple Galactic rotation model. Eilers et al. (2019) determined a Galactic rotation curve of $v = 229\,\mathrm{km\,s^{-1}} - 1.7\,\mathrm{km\,s^{-1}\,kpc^{-1}}(\mathrm{R} - \mathrm{R_{\odot}})$. Along the line of sight of the Cygnus region, these stars have a Galactocentric radius of $R \sim 7.7$ kpc, assuming d = 1.8 kpc and $l = 72°$, at an angle intercepting the line of sight of 4.5 degrees the rotation speed is equal to 229.6 km s^{-1}. This translates into a motion of about -18 km s^{-1} in the l direction. With the local Solar motion of U = 11.1 km s^{-1} and V = 12.24 km s^{-1} (Schönrich et al., 2010), the Sun requires a correction of about -6.73 km s^{-1} in that direction. This therefore results in an apparent motion of -24.73 km s^{-1}, or a proper motion of about -2.74 mas yr^{-1} assuming a distance of d = 1.8 kpc. The majority of stars in this direction have a proper motion in l more negative than this, suggesting that the Cygnus region is moving towards smaller l values (towards the inner Galaxy) relative to other stars orbiting within the Milky Way.

Stars in this sightline carry typical proper motions in Galactic latitude of -1 mas yr^{-1}. If the motion of the Sun is 7 ± 0.5 km s^{-1} (Bland-Hawthorn & Gerhard, 2016) in the positive b direction, it means that stars in the Galactic disk will have negative proper motions of a similar magnitude. This would equate to $\sim$ -1 mas yr^{-1} at a distance of 1.8 kpc, consistent with the observations along this sightline.

3.3.2 Identifying kinematic groups

For the purpose of identifying OB associations, I limited the sample of stars to those with $\log(T_{\mathrm{eff}}/K) > 4.2$ (approximately equivalent to spectral type B5 or earlier, Pecaut & Mamajek 2013). This reduces the sample down to 1349 stars. These hot and massive stars are almost definitely young, and would therefore have similar kinematics to their birth environment. With these subsamples I will apply a 4D clustering - two dimensions for position and two for velocity, hence a plane of the sky approach, an approach that is validated by the limited range of distances over which the Cygnus OB associations lie (see Section 3.1).

To identify OB associations or groups of kinematically-coherent stars within the wider distribution of OB stars, I sought to find groups of stars that are both relatively close together on the plane of the sky, but also with more similar kinematics between them than the general distribution of OB stars within the region. A good starting point consisted of visualising how the relative kinematics of stars changes over the area of study. I thus defined a grid with a cell size of 0.1° and at each point in this grid I selected the ten nearest stars from the sample of 1349 stars (hence nearest neighbours). I then performed a Kolmogorov-Smirnov (KS) test comparing the proper motion distribution of these ten stars with the proper motion distribution of all 1349 stars. Such a process was calculated for both the proper motion in l and in b, obtaining a P-value for each. I multiplied these two quantities together and this resulted in the probability that the kinematics of stars in that area are consistent with that of the wider population.

The distribution of this probability is illustrated in Fig. 3.7, over the area of study, effectively highlighting regions that are kinematically distinct compared to the wider area. Contours are shown at 2, 3, and 4σ significance. Immediately evident are several areas with significantly distinct kinematics ($\log(P) < -6$), including one located in the direction of Cyg OB2 ($l = 80°$, $b = +1°$). This provides an early verification of this new method as Cyg OB2 is the most kinematically-coherent of the historical Cygnus OB associations (see Section 3.1).

To test the statistical significance of this method I repeated the analysis 100 times, randomising the proper motions of the 1349 stars in each iteration and computing the log(P) value at each position. The goal of this process was to check whether the probabilities derived from the actual proper motions are consistent with a random distribution of kinematics. Fig. 3.8 compares the observed distribution of probabilities with that derived from the randomised velocities and shows the former reaches considerably smaller values of log(P). This suggests that the majority of the kinematic structures observed in Fig. 3.7 are likely to be real.

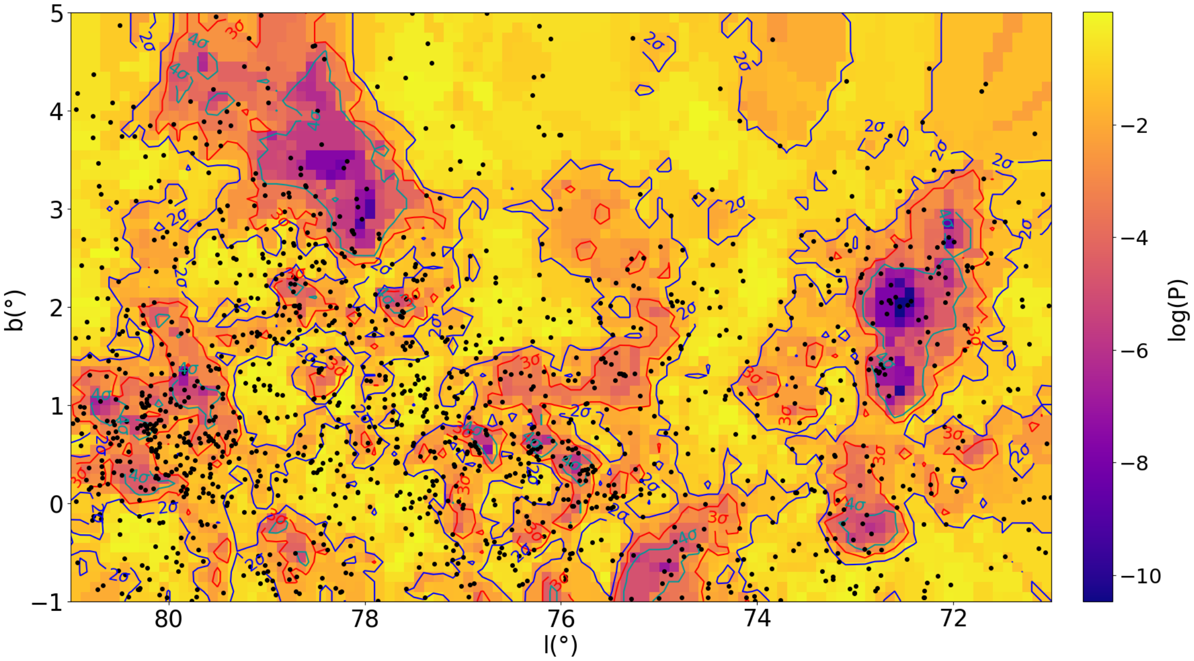


Figure 3.7: The Cygnus region, colour-coded according to probability, log(P), that the kinematics of stars in that vicinity are consistent with that of the wider population. The 2σ, 3σ, 4σ contours represent respectively levels at log(P) = -1.3,-2.6 and -4.2.

To identify stars in a given group, I chose the 3σ contour level and defined all stars falling within such a contour as being members of that group. This value was chosen as a balance between the 4σ level excluding too many stars and the 2σ level where the kinematic coherence was not strong enough. Given the number of contours, I required each group to contain a minimum of 10 stars, which led to the identification of four distinct kinematically-coherent groups of stars.

This method could include contaminants projected against real groups, so to exclude them I calculated the median, 16th and 84th percentile values of μ_l, μ_b and ϖ, and excluded sources that were more than 3σ from the median value in each dimension, where σ is the intrinsic dispersion of a parameter computed through the outlier-resistant method from Ivezić et al. (2014). This process reduced the membership of each group by $\sim$ 20–30%.

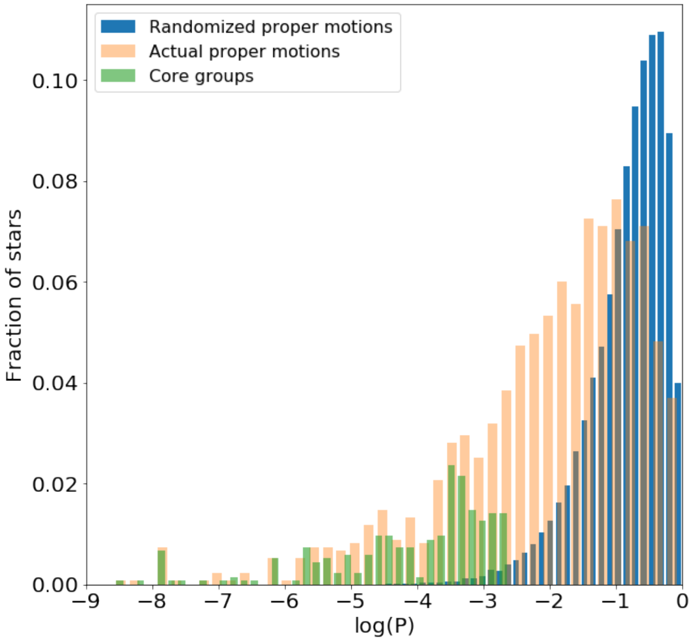


Figure 3.8: log(P) values across our survey area deduced from the randomized proper motions, from the actual proper motions, and for the core groups (before the removal of outliers).

Five associations (see Fig. 3.9) were identified and defined as:

1. Association A is located in the upper-right part of Fig. 3.7, with $l = 71.5 - 73°$ and $b = 0.5 - 3.0°$.
2. Association B spans the upper-left part of Fig. 3.7, in the region $l > 77 - 80°$ and $b = 2.5 - 5.0°$.
3. I initially defined a third association in the region of l from 75 to 77.5° and b from -0.5 to 2°, but further investigation suggested this constituted two associations with distinct μ_l distributions and so this was separated into two associations, C and D, with μ_l less than and greater than $\mu_l = -6.2$ mas yr^{-1}, respectively.
4. The association in the lower left part of Fig. 3.7 corresponds to Cyg OB2 and

its surroundings. It forms association E and is the largest structure identified within the region.

The initial selection of association members is based on stars with $\log(T_{\rm eff}/K) > 4.2$. This threshold was chosen to provide a balance between a large enough sample to identify these groups and using only stars that are sufficiently hot to be very massive and young. Once the 'core' groups were outlined, I expanded their membership by including cooler stars, which fulfill either $\log(T_{\rm eff}/K) > 4$ or $\log(L/L_{\odot}) > 2.5$. The goal was to include all possible OB stars and notably massive post-main sequence stars that have evolved to cooler temperatures. Besides adding to each group any star that fulfils these criteria, I required that they fell within the same 3σ contours, and were located within 2 standard deviations of the median value of parallax and proper motion (with the median and standard deviations defined from the current association membership). This increased the membership of the existing associations to a total of 109, 100, 75, 65 and 168 stars, respectively.

Finally, whilst analysing further the kinematics of these stars, I became aware of a 'gap' between associations A and D in their distributions in l, b, and μ_l. I identified the stars that fell between associations A and D in both l, but also in the l vs μ_l diagram (see Fig. 3.10), trimmed the sample and then added in slightly cooler stars, all as described above (this time also discarding the outliers in l and b). This resulted in an association of 147 stars called association F. It should be noted that the association has been selected manually, contrary to the other associations. Fig. 3.9 shows the spatial distribution of all the final associations, while Fig. 3.10 shows the proper motion distribution (left panel) and l vs μ_l (right panel) distribution of the associations.

3.3.3 Comparison with historical associations

I compared the membership of these new associations with that of the historical associations introduced in Section 3.1 and the results confirmed the findings in Section 3.1. While Cyg OB2 stands out as a real association, corresponding roughly to association

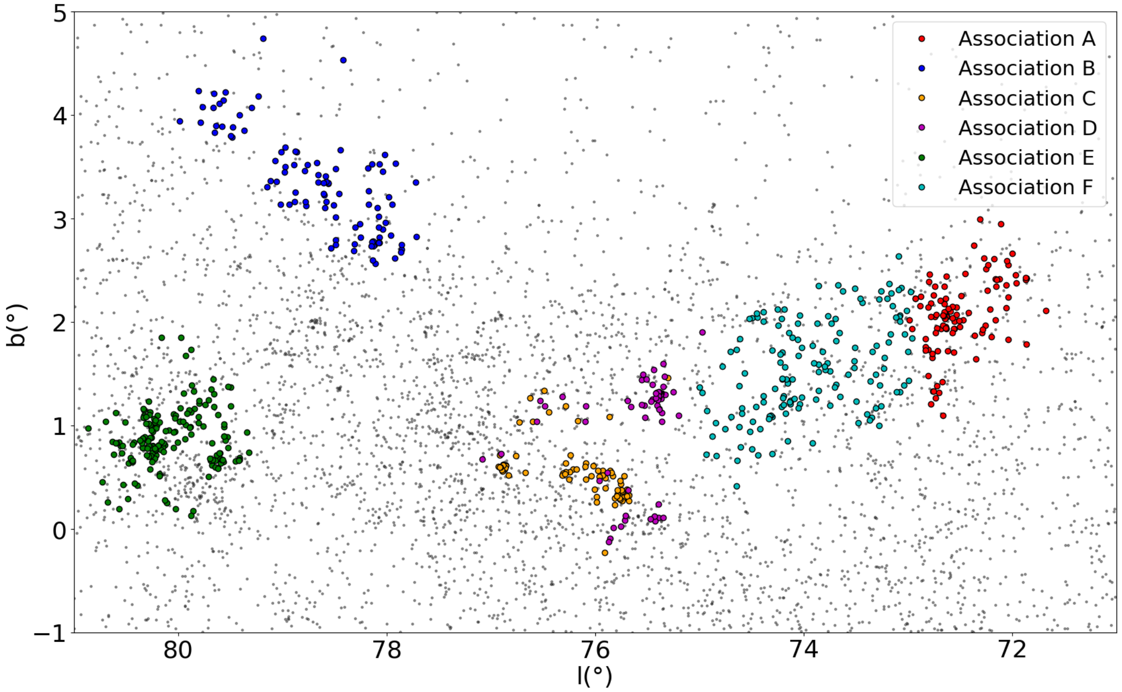


Figure 3.9: Spatial distribution of the six newly identified kinematically-coherent OB associations (coloured symbols), plotted against the field OB stars (faint black dots).

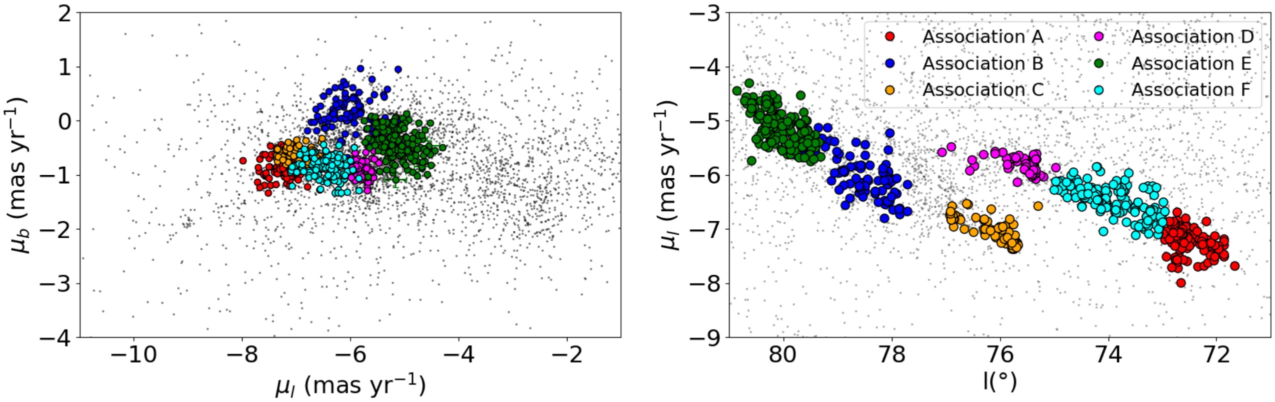


Figure 3.10: Left: proper motion distribution of all 6 new OB associations (coloured symbols), plotted against the field OB stars (black dots). Right: proper motion in l plotted against l for the same objects. The uncertainties in μ_l are comparable to, or smaller than, the symbol size, and thus are not shown.

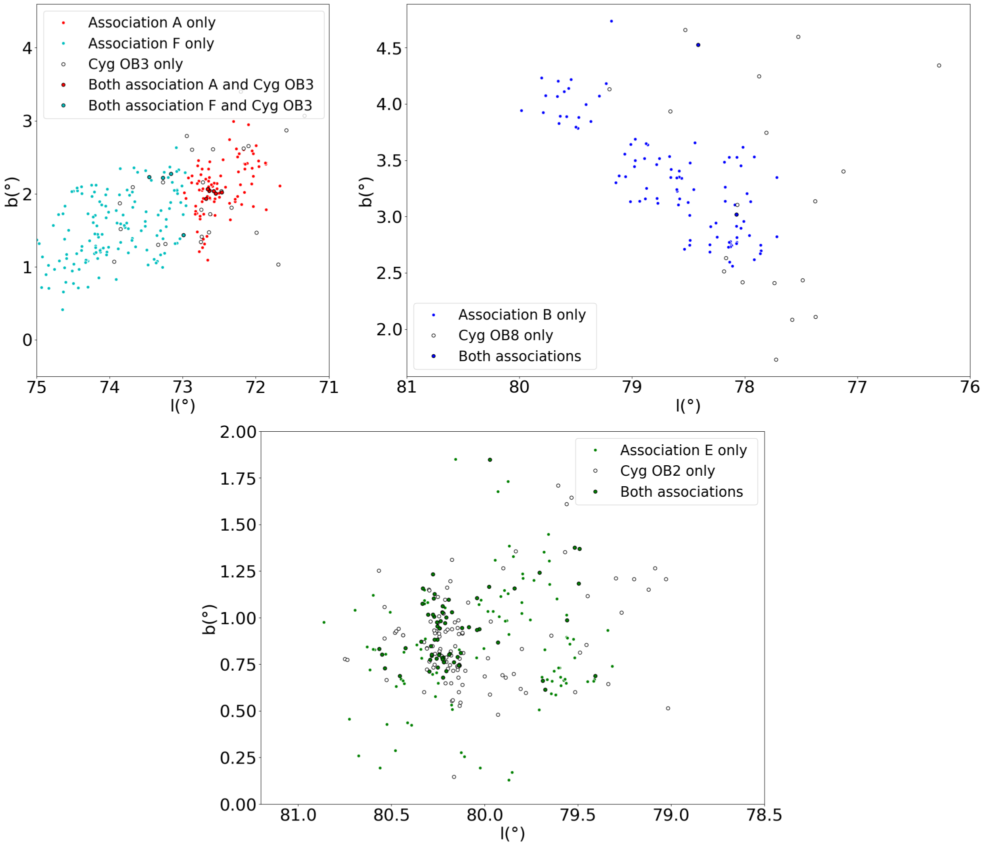


Figure 3.11: Comparison between the new associations and the historical associations Cyg OB2, OB3 and OB8.

E with 64 stars in common with the list from Berlanas et al. (2019), and Cyg OB3 has some overlap with association A, the same is not true for the other associations. They indeed show very little overlap with the historical OB associations (typically 5-20% of the historical association members are found in the newly identified associations). Comparison plots between the historical associations and the new associations are shown in Fig. 3.11 for the relevant ones.

Each of the historical associations contains fewer stars than the newly identified associations, with the exception of Cyg OB2. This can be explained by the availability of larger surveys and modern techniques to identify OB stars, compared with what was available to Humphreys (1978) and Blaha & Humphreys (1989). Also, these new OB associations contain late B-type stars, compared to the historical associations which were composed mostly of O and early B-type stars.

3.3.4 Comparison with DBSCAN

To test my group-finding method and compare the results to those from modern clustering algorithms, I again selected the 1349 stars of the sample with $\log(T_{\text{eff}}/K) > 4.2$ and used the clustering algorithm DBSCAN (Pedregosa et al., 2011). I chose the number of nearby stars to be ten for similarity with the method and a maximum normalised distance between stars (in both spatial and proper motion coordinates) to be considered part of the same group of 0.05. They were the most relevant parameters of DBSCAN to perform the clustering.

DBSCAN was able to identify six associations, respectively containing 154, 169, 23, 48, 13 and 33 stars. Fig. 3.12 shows how these associations compare with the associations identified with the P-value method, some of which match up well, while others do not. For example, association A agrees well with DBSCAN association 6 and association E matches with DBSCAN association 2, though in both cases my associations are more extended than those identified by DBSCAN. In other cases the agreement is less good: associations C and D overlap with DBSCAN associations 1 and 3 in a complex manner. The most obvious distinction lies between association 5 from DBSCAN and our association B. DBSCAN was unable to pick out the bulk of association B. This is because, as outlined above, one of the main parameters of DBSCAN is the relative distance between the objects used to form clusters. Consequently, since the same value was used throughout the entire region, it was impossible to detect this lower-density association, while still detecting the higher-density associations found elsewhere. My method, on the other hand, only relies on the number of neighbour stars, and therefore is not biased by density in this regard.

Sometimes the opposite trend is observed, i.e. some stars have been identified as part of the DBSCAN associations but not with my method. This stems from my

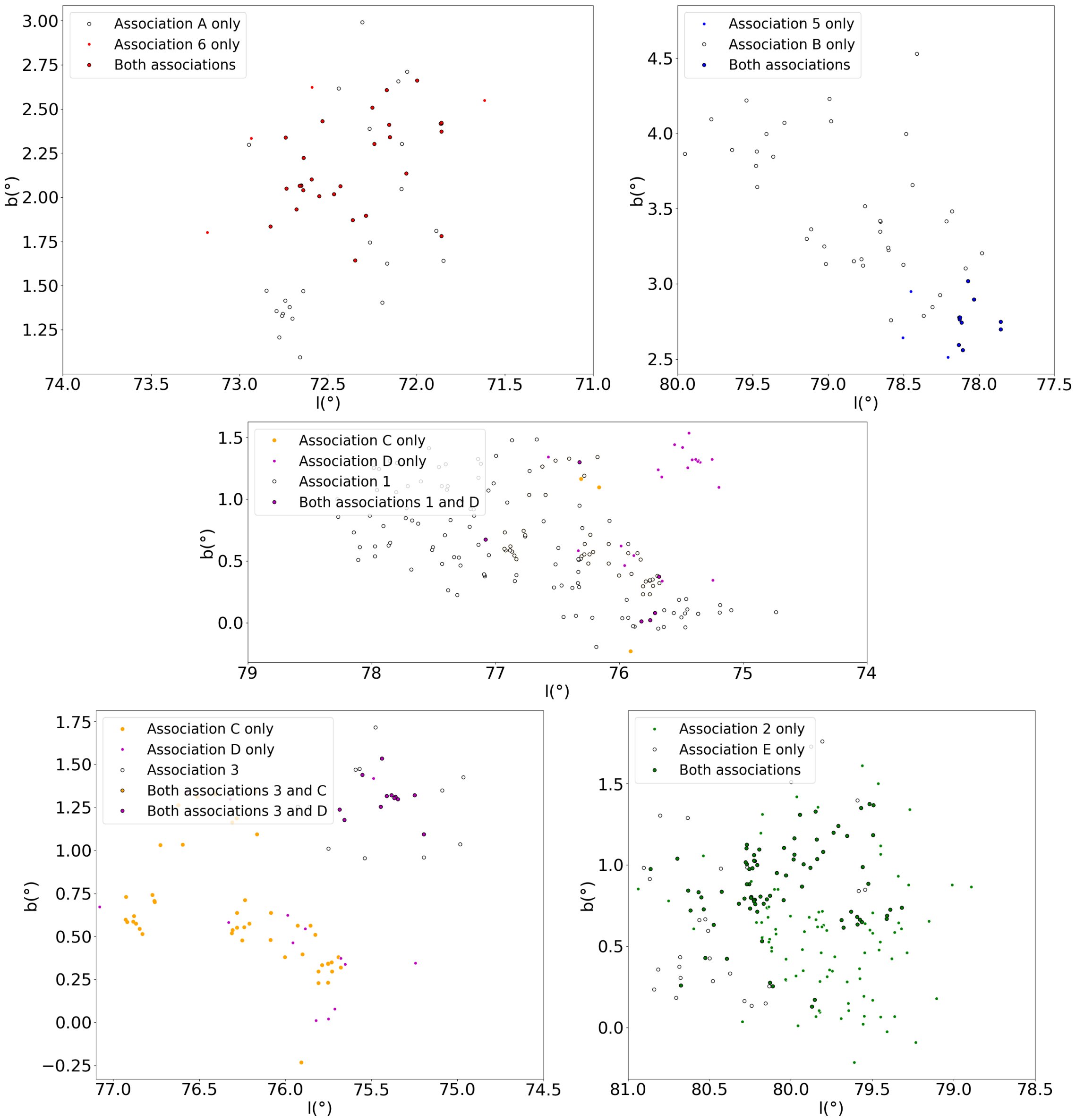


Figure 3.12: Comparison between our new associations and the associations identified by DBSCAN.

selection: I chose sources within 3σ contours, rejecting sources that fell outside of these contours and which may represent members in the periphery of a association. Choosing 2σ contours may have solved that issue, however, this would have lowered the kinematic coherence of these associations.

3.4 Analysis of the new OB associations

In this section I analyse the physical and kinematic properties of the new associations. Relevant physical parameters are their total stellar mass, number of O- and B-type stars, and to accompany these results I produce HR diagrams. For the kinematic properties, I estimate both their velocity dispersions and velocity gradients with the aim of measuring their expansion rate. Tables 3.1 and 3.2 summarise some of the physical and kinematic properties of each association, which were calculated as follows.

3.4.1 Physical properties of the individual associations

To estimate the observed number of O and B stars in each association, I defined B-type stars as those with $\log(T_{\mathrm{eff}}/K) > 4$ and $\log(T_{\mathrm{eff}}/K) < 4.47$ and O-type stars as those with $\log(T_{\mathrm{eff}}/K) > 4.47$. Where a spectral type is available from the literature I used the spectroscopic T_{eff}, otherwise I used the effective temperature derived from the SED fits. The uncertainties on the number of O and B stars was estimated using a Monte Carlo experiment, varying the temperature of each star according to its uncertainties (which are typically lower for the spectroscopic temperatures). It has to be noted that some O-type stars may have already undergone supernovae, which may result in an under-representation of O-type stars in older associations.

I then attempted to calculate the total stellar mass of each association. To that end, the SED-fitted masses were used for consistency. Subsequently, to determine the complete mass range, I inspected the mass function, chose the turn-over point of the

Table 3.1: Properties of the three first new OB associations. The first column indicates the parameters, where the subscript 'm' indicates the median value and 'σ' the dispersion.

Parameters	Units	Assoc. A	Assoc. B	Assoc. C
$\mathrm{RA(ICRS)_m}$	deg	301.45	304.38	305.45
$\mathrm{DE(ICRS)_m}$	deg	35.69	41.41	37.77
l_m	deg	72.62	78.56	76.02
b_m	deg	2.06	3.32	0.54
d_m	pc	1894.5	1729.6	1709.6
μ_{l_m}	mas yr^{-1}	-7.20	-5.96	-7.01
σ_{μ_l}	mas yr^{-1}	0.20	0.45	0.27
μ_{b_m}	mas yr^{-1}	-0.88	0.10	-0.61
σ_{μ_b}	mas yr^{-1}	0.17	0.28	0.11
σ_{v_l}	km s^{-1}	$1.84^{+0.48}_{-0.41}$	$3.66^{+0.76}_{-0.64}$	$2.23^{+0.49}_{-0.35}$
σ_{v_b}	km s^{-1}	$1.57^{+0.35}_{-0.28}$	$2.28^{+0.43}_{-0.34}$	$0.92^{+0.21}_{-0.17}$
Observed number of B stars		100^{+2}_{-3}	88^{+3}_{-3}	61^{+3}_{-2}
Observed number of O stars		3^{+1}_{-1}	5^{+1}_{-1}	8^{+2}_{-1}
Estimated total stellar mass	$M_\odot$	1883^{+243}_{-211}	1828^{+229}_{-196}	1534^{+213}_{-189}
Velocity gradient (l)	mas yr^{-1} deg^{-1}	0.22 ± 0.07	0.43 ± 0.05	0.36 ± 0.04
Velocity gradient (l)	km s^{-1} pc^{-1}	0.06 ± 0.02	0.12 ± 0.01	
Velocity gradient (b)	mas yr^{-1} deg^{-1}	0.40 ± 0.04	-0.03 ± 0.04	
Velocity gradient (b)	km s^{-1} pc^{-1}	0.11 ± 0.01	-0.01 ± 0.01	0.03 ± 0.01
Expansion age (l)	Myr	16.67 ± 5.55	8.33 ± 0.69	10.00 ± 1.00
Expansion age (b)	Myr	9.09 ± 0.83	-	-

mass function as the lower limit and the post-main sequence turn-off mass at 15 Myr for the upper limit (Ekström et al., 2012). The samples, based on this, are relatively complete in the mass range of 3.2 – 13.8 $M_\odot$, thus covering the range of B-type stars.

Table 3.2: Same than Table 3.1 but for the three other new associations.

Parameters	Units	Assoc. D	Assoc. E	Assoc. F
$\mathrm{RA(ICRS)_m}$	deg	304.28	308.08	302.86
$\mathrm{DE(ICRS)_m}$	deg	37.65	41.30	36.58
l_m	deg	75.42	80.16	74.03
b_m	deg	1.24	0.84	1.52
d_m	pc	1998.7	1665.4	1985.3
μ_{l_m}	mas yr^{-1}	-5.83	-5.13	-6.44
σ_{μ_l}	mas yr^{-1}	0.15	0.30	0.29
μ_{b_m}	mas yr^{-1}	-0.87	-0.42	-0.85
σ_{μ_b}	mas yr^{-1}	0.17	0.32	0.16
σ_{v_l}	km s^{-1}	$1.45^{+0.42}_{-0.38}$	$2.38^{+0.75}_{-0.55}$	$2.71^{+0.55}_{-0.47}$
σ_{v_b}	km s^{-1}	$1.36^{+0.41}_{-0.37}$	$2.52^{+0.66}_{-0.37}$	$1.55^{+0.31}_{-0.26}$
Observed number of B stars		58^{+2}_{-2}	116^{+4}_{-3}	134^{+3}_{-2}
Observed number of O stars		2^{+1}_{-1}	34^{+3}_{-2}	5^{+2}_{-1}
Estimated total stellar mass	$M_\odot$	1019^{+183}_{-152}	3502^{+317}_{-303}	1972^{+245}_{-219}
Velocity gradient (l)	mas yr^{-1} deg^{-1}	0.16 ± 0.04	0.46 ± 0.06	0.31 ± 0.03
Velocity gradient (l)	km s^{-1} pc^{-1}	0.04 ± 0.01	0.13 ± 0.02	0.08 ± 0.01
Velocity gradient (b)	mas yr^{-1} deg^{-1}	0.11 ± 0.03	0.12 ± 0.04	0.06 ± 0.03
Velocity gradient (b)	km s^{-1} pc^{-1}	0.02 ± 0.01	0.03 ± 0.01	0.02 ± 0.01
Expansion age (l)	Myr	22.98 ± 5.74	7.69 ± 1.18	11.86 ± 1.48
Expansion age (b)	Myr	33.42 ± 11.14	30.63 ± 10.21	-

The filtering of the data (Section 3.2.1) means that my samples of members will be slightly incomplete. The incompleteness was estimated by calculating the fraction of stars, as a function of magnitude, that were discarded at each step in Section 3.2.1. The completeness level varies for each step, and as a function of magnitude (shown in Fig. 3.13), with a total completeness of 70–80% in the magnitude range of interest.

Fig. 3.13 shows these incompleteness curves and an example of how the membership of one of my associations increases once this incompleteness is accounted for.

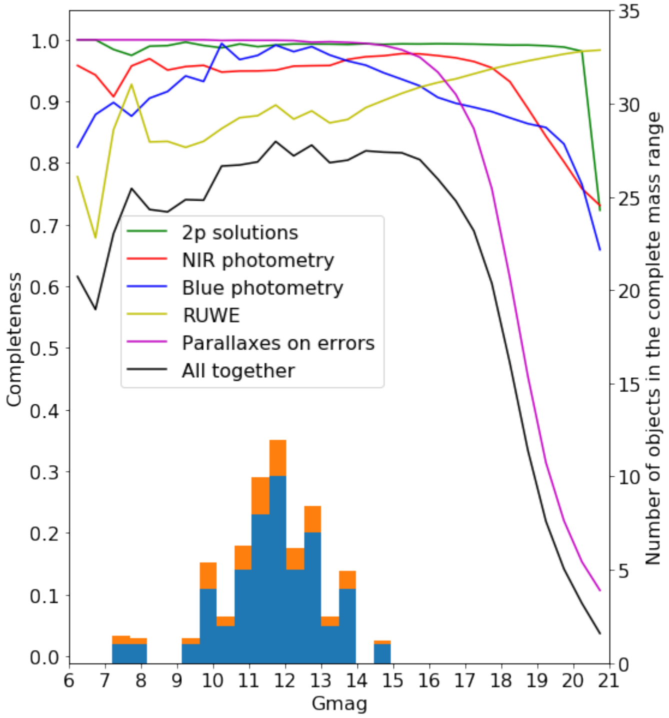


Figure 3.13: Completeness as a function of G magnitude for the various data filtering steps in Section 3.2.1. The black curve corresponds to the combination of all the filtering effects. The histogram show the number of stars in association A within the complete mass range both before (blue) and after (orange) applying the incompleteness corrections.

The last source of incompleteness stems from the false positive and negative rates from the SED fitting process. In Chapter 2, for this same sample, I quantified the recovery rates at different temperature thresholds by comparing the SED-fitted temperatures with literature spectroscopic temperatures. I repeated this here using as a threshold the minimum mass of the complete mass range ($\log(T_{\rm eff}/K) = 4.11$). For sources with $A_V < 4$ (appropriate for the low-extinction associations) the false negative rate is 5.0% and the false positive (contamination) rate is 5.9%, while for sources with

$A_V > 4$ (appropriate for Cyg OB2), these quantities are respectively equal to 3.5% and 0.0%. The 'true' number of stars, labelled as T is written as:

$$T = \frac{N}{1-F} - N\,P \tag{3.1}$$

where N is the corrected number of stars within the complete mass range, F is the false negative rate and P the false positive rate.

I applied both of these corrections in order to estimate the total stellar mass of each association. Subsequently, I performed a Monte Carlo simulation applying the mass functions from Maschberger (2013) to sample stellar masses, counting the number of stars in the representative mass range as well as the total number (and mass) of stars. When the simulation reached the total number of observed stars in the representative mass range, I halted it and took the total mass of the association of stars. I repeated this 10,000 times, every time varying the number of stars observed in the representative mass range (according to their uncertainties), in order to extract the errors of the total stellar mass.

3.4.2 HR diagrams and ages of association members

Placing stars in an HR diagram requires their luminosity and effective temperature. While the SED fitting process provides these values, I preferentially use the spectroscopic parameters when available for better accuracy. For this subsample, I derived stellar luminosities through the following steps. First I defined their extinction in the G-band:

$$A_G = 1.0327\,[(G-K)-(G-K)_0] \tag{3.2}$$

where G and K are the magnitudes from *Gaia* and 2MASS respectively and $(G-K)_0$ is computed from the unreddened model SEDs.

The extinction was then used to calculate the absolute magnitude in the G band:

$$M_G = G - 5\,\log(d) + 5 - 1.088\,A_G \tag{3.3}$$

where A_G was converted to A_V using the conversion from Jordi et al. (2010) (although it may be outdated, see end of Chapter 2).

Finally Eq. (3.3) was used to compute the stellar luminosity of the spectroscopic sample:

$$\log(L/L_\odot) = \frac{4.75 - M_G - BC_G}{2.5} \tag{3.4}$$

where BC_G is the bolometric correction for the G band from Jordi et al. (2010).

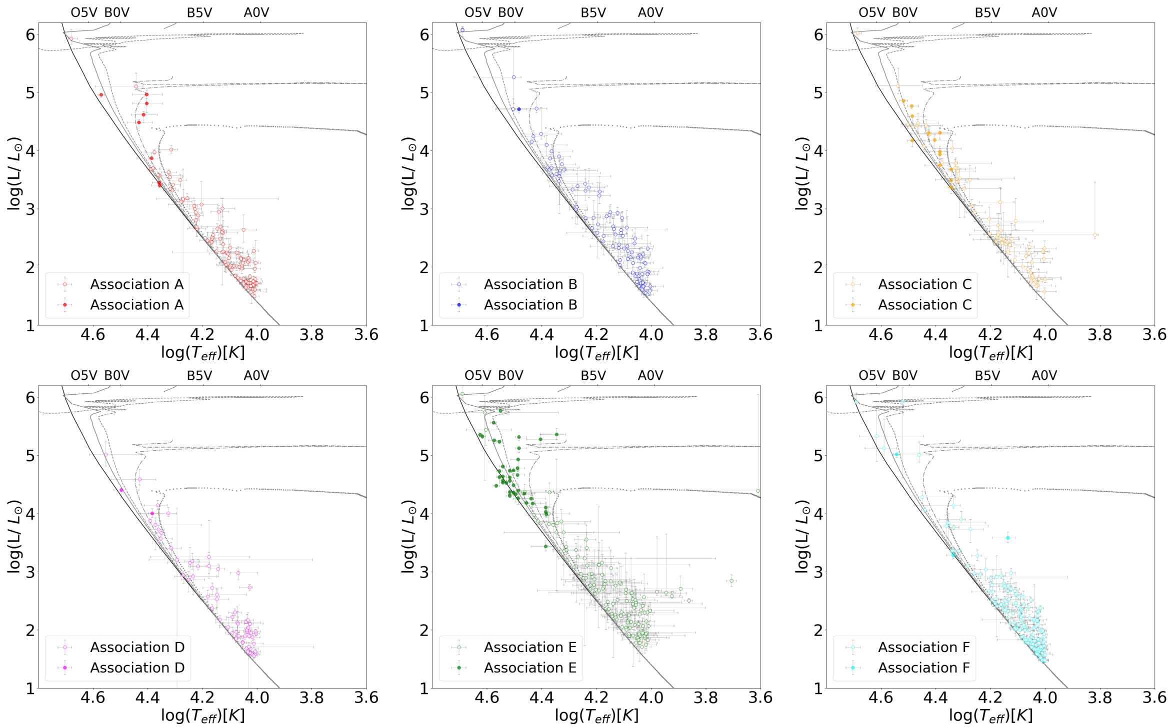


Figure 3.14: HR diagrams for all the OB associations. Full coloured symbols correspond to those with spectroscopy while the empty symbols are those without. Isochrones from the rotating models in Ekström et al. (2012) are also shown. From top to bottom these are the 3.16, 5, 10 and 20 Myr isochrones. Positions of some spectral types are indicated on the top horizontal axis.

The HR diagrams for each association member are shown in Fig. 3.14. From these plots, it is clear that estimating the ages of these associations is difficult. Many of the associations include members close to the ZAMS at high luminosities, implying the existence of particularly massive (and therefore young) stars. These would possibly need to be as young as 3–5 Myr, depending on initial rotation rate. There are notable examples of this in associations B, C and E (Cyg OB2). Some of the associations also include members moving away from the main sequence, particularly around the 10 Myr isochrone, implying the presence of stars around this age. Examples of this can be found in associations A, D, E (to a lesser extent) and F.

With these indicators of age, I can place the associations in an approximate age order, with associations B, C and E as the youngest (due to the presence of very luminous stars within them), and associations A, D and F as the oldest (due to the presence of stars at or beyond the 10 Myr isochrone). However, I note that many OB associations, including these, exhibit large age spreads (see Wright 2020 for examples) and therefore assigning a single age to any association is difficult. Furthermore, there are other potential source of age spreads. One of them is binarity, as my SED fitter currently models single stars only, while in Section 1.2.1 I mentioned that massive stars are more likely to belong to binary or multiple systems. Another one is stellar rotation, as the models from Ekström et al. (2012) use a single rotation rate, while the rotation rate of a star is expected to slow down with age due to the effect of magnetic field (Epstein & Pinsonneault, 2014; Vidotto et al., 2014).

3.4.3 Kinematic properties of the individual associations

I calculated the median positions and proper motions of each association along with their distance. Each association is typically located between 1.5 and 2.0 kpc, just like the historical Cygnus OB associations, associations B, C and E being the closest and associations A, D and F being the most distant.

The proper motion dispersions were computed using the method from Ivezić et al. (2014), with random gaussian sampling to estimate uncertainties. Velocity dispersions range between 1 and 4 km s^{-1} as indicated in Tables 3.1 and 3.2, consistent with other OB associations (Wright, 2020).

Determining whether the associations are expanding required me to search for evidence of velocity gradients in the kinematics of the stars. I fitted a linear relationship using an MCMC simulation and the *emcee* package, and obtained velocity gradients for each association (see Tables 3.1 and 3.2). An example fit is shown in Fig. 3.15 for association E which exhibits the strongest expansion signature. I note that the velocity gradients are generally anisotropic and larger in the l direction compared to the b direction, similar to the pattern of anisotropic expansion observed in other OB associations such as Sco-Cen, which has been attributed to the existence of kinematic substructure (Wright & Mamajek, 2018). Alternatively, the gravitational potential from the Galactic disk could engender an anisotropic expansion.

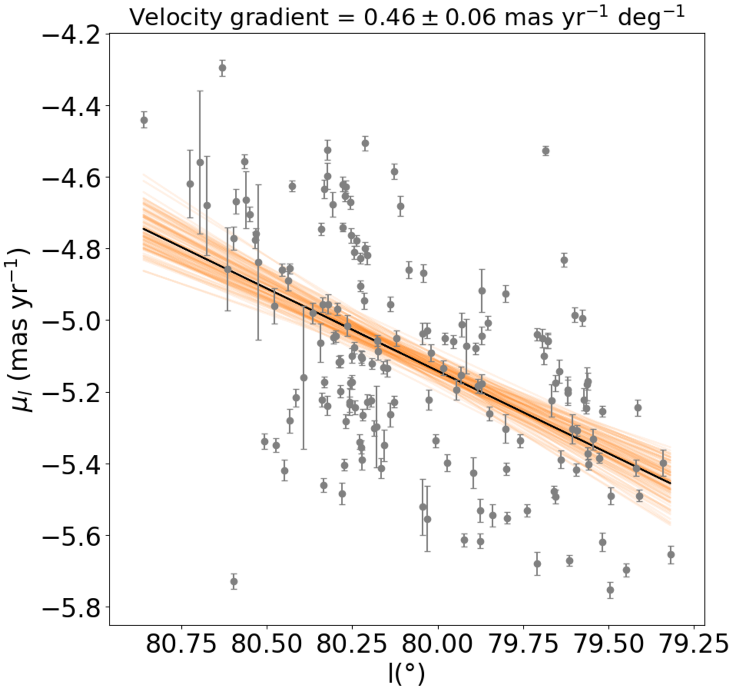


Figure 3.15: Results of the MCMC fit between position and proper motion in the l direction for association E. 100 random samples have been drawn from the posterior distribution and shown in orange. The black line corresponds to the best-fit value for the velocity gradient which has been indicated on the top of the plot.

The next step consisted of converting velocity gradients into expansion ages. This was achieved by inverting the expansion gradient, and the resulting expansion ages are listed in Tables 3.1 and 3.2. Expansion ages are based on the assumption that the entire association expanded from a compact region of space, and therefore represent upper limits rather than precise ages. I have not calculated these quantities in Galactic latitude for associations B, C and F as their velocity gradients were either too small or negative, as shown in Tables 3.1 and 3.2 (both of which result in unphysical expansion ages). On the other hand, results for the l direction tend to be closer to the likely ages for these systems. The expansion ages in the l and b directions do not agree for any of the associations except association D, as has been observed in many other OB associations (see e.g. Wright 2020).

While these expansion ages should be considered as upper limits, it is reassuring to see that the three associations identified as being the youngest (B, C and E, see Section 3.1) have the smallest expansion ages, and the three older associations have the largest expansion ages.

3.4.4 Large-scale dynamics of the Cygnus associations

In the right-hand panel of Fig. 3.10 the proper motion in Galactic longitude is plotted against Galactic longitude. A very clear trend of increasing proper motion is shown as a function of Galactic longitude for the stars in two combinations of associations: associations A, D and F at low longitudes and associations B, C and E at high longitudes. While this pattern was already identified in Fig. 3.6, it is significantly clearer in Fig. 3.10 where just the association members are shown. Notably, these two sets correspond to the three young associations and the three old associations (see Section 3.4.2), hinting at a connection in both dynamics and age. This pattern repeats itself on a scale of five degrees in longitude (from l = 81° to 76° and from l = 76° to 71°). and approximately 3 mas yr^{-1} in proper motion. At a rough distance of 1.8 kpc these equate to ~150 pc and ~25 km s^{-1}, respectively. No such pattern is seen for μ_b as a function of either l or b, again emphasising the preferential expansion along the l direction.

3.5 Discussion

In this section I discuss the results, how they compare to previous studies and what they imply for our general understanding of OB associations. The main results are:

- Most of the historical OB associations in the Cygnus region lack kinematic coherence and therefore do not appear to be genuine OB associations. I identified six new systems that I compared to the previous divisions. There is a strong overlap between my association E and Cyg OB2, a partial overlap between association A and Cyg OB3, but no significant overlap between my other associations and the historical associations.

- I calculated the broad properties of each new association, including estimates of their total stellar mass, quantifying their incompleteness. HR diagrams were produced for every association. Association E (Cyg OB2) contains by far the largest number of O-type stars and is the most massive.

- I computed the velocity dispersions of each association and searched for evidence that they are expanding. I found strong evidence of expansion for all of the associations in the l direction and in associations A, D and E in the b direction, with asymmetric expansion patterns consistent with what has been observed for other OB associations.

- I discovered a correlation between l and μ_l on large scales across the entire Cygnus region, specifically connecting the three oldest and the three youngest associations, thus showing a connection between age and dynamics.

3.5.1 The new Cygnus OB associations

I have identified 6 new OB associations inside the Cygnus region, with properties consistent with being OB associations (Wright, 2020), notably with their total stellar mass between 1000 and 3500 $M_{\odot}$. These associations are remarkably kinematically-coherent (see Fig. 3.9 and 3.10), especially when contrasted with the historical OB associations (Section 3.1). Comparing these associations to the historical associations, the best match is between association E and Cyg OB2 with 64 stars in common, yet association E has a better kinematic coherency due to its selection method.

Zari et al. (2021) provided a map of hot and luminous OBA stars in the local part of the Milky Way and noticed that massive star-forming regions and OB associations appeared as overdensities. Yet only Cyg OB2 is clearly prominent in their map, meaning that, in spite of their completeness, they were not able to identify some of the associations that I have identified. Cyg OB2 is the densest and most massive association of the region, while the other associations have a lower density, by approximately one order of magnitude, explaining why Zari et al. (2021) did not detect such groups. Therefore my method is shown to be capable of detecting less dense features.

Berlanas et al. (2019) divided Cyg OB2 in two substructures along the line of sight, with a main group situated around 1760 pc and a smaller, foreground group at 1350 pc, with association E matching the former. On the other hand, Orellana et al. (2021) identified 2767 proper motion members of Cyg OB2 at 1683±5 pc. This includes 300 stars in common with previously known spectroscopic and photometric member of this association, having a mean distance of 1669 ± 5 pc. This is very close to the median distance of association E in Table 3.2). Orellana et al. (2021) also suggested that Cyg OB2 can be divided into several substructures.

The existence of two separates structures composing Cyg OB2 could explain why association E does not include some of the classical members of Cyg OB2 and why my estimate of the total mass of association E is smaller than the mass of Cyg OB2 calculated by Wright et al. (2015). Nonetheless, given their close characteristics, such as central coordinates, distance, total stellar mass, age and high number of O-type stars (Wright et al., 2015, 2016; Berlanas et al., 2019; Orellana et al., 2021), the newly identified stars in association E are likely to be part of Cyg OB2.

3.5.2 Expansion

The expansion of OB associations has been assumed since their very first discovery as they were argued to be the expanded remnants of dense star clusters (Ambartsumian, 1947). Wright (2020) emphasized the two 'extreme' scenarios for the origin of OB associations being hierarchical and clustered star formation (see Chapter 1). In the latter case, OB associations are formed when compact star clusters are disrupted by processes such as residual gas expulsion (see e.g. Lada & Lada 2003), while in the former case multiple structures may form hierarchically and drift apart without a clearly coherent expansion pattern. Which of these two theories can best explain the highly anisotropic expansion patterns in many OB associations remains to be seen (though Kruijssen 2012 outlined a possible cause of asymmetric expansion).

I measured a clear expansion trend in all of the identified associations in the l direction along with a similar trend in the b direction for associations A, D and E. By comparison, Wright et al. (2016) found no clear evidence of expansion in Cyg OB2. Wright & Mamajek (2018) also failed to detect expansions in both directions in Sco-Cen, neither did Ward & Kruijssen (2018) for their 18 studied OB associations. Using *Gaia* DR2 data, Melnik & Dambis (2020) only identified expansion in 6 out of their 28 selected OB associations, and not in any of the historical Cygnus OB associations. On the other hand, Cantat-Gaudin et al. (2019) and Armstrong et al. (2020) measured expansion in sub-groups inside the Vela-Puppis region, albeit anisotropic expansion, while Kounkel et al. (2018) made a similar discovery in the substructures of the Orion complex.

Of all these studies, my results fall in the second category and share the feature with those studies that the OB associations have been identified kinematically. The fact that association E is expanding while Wright et al. (2016) did not detect any expansion in the historical Cyg OB2 illustrates this trend: Wright et al. (2016) included some stars in Cyg OB2 with different kinematics and potentially at a different distance (Berlanas et al., 2019). This is shown in Fig. 3.16, where the stars in Cyg OB2 that are not in my association E have a broader distribution in μ_l and therefore do not show the signature of expansion that I have detected here.

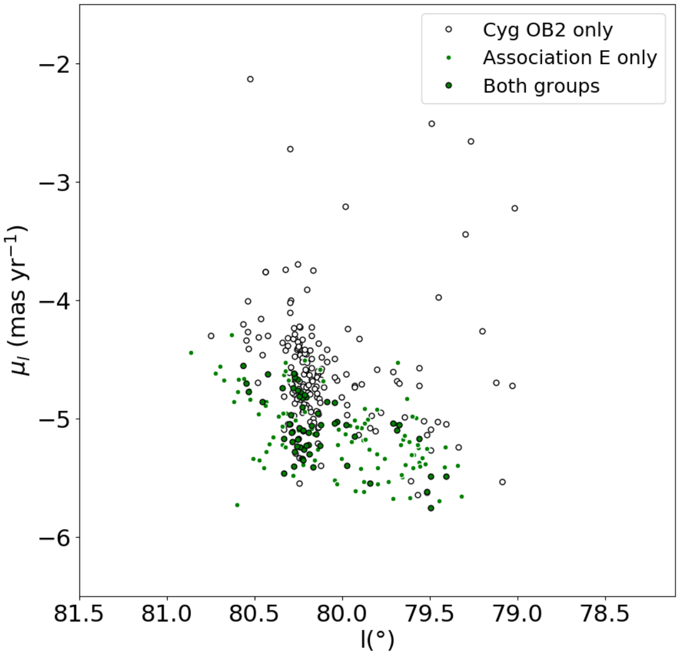


Figure 3.16: Proper motion in Galactic longitude as a function of Galactic longitude for association E members and the Cyg OB2 members from Berlanas et al. (2019), highlighting the 64 stars in common.

It is clear that studies that use the historical membership of an OB association have typically failed to find evidence of expansion, while those studies that have redefined the membership using kinematic information have more consistently detected expansion trends. These results encourage a broader analysis of OB associations to confirm whether expansion trends are common.

3.5.3 Large-scale kinematics and expansion

The large-scale kinematics seen in the right panel of Fig. 3.10 shows two patterns with a length of about five degrees in Galactic longitude and 3 mas yr^{-1} in proper motion, equating to a length of about 150 pc and a velocity of 25 km s^{-1} (at a distance of 1.8

kpc). The direction of Galactic longitude towards the Cygnus region is approximately transverse to the direction of the Cygnus spur (or spiral arm) within the plane of the Galactic disk. A similar kinematic pattern has been observed amongst the OB stars in the Carina region by Drew et al. (2021).

The nature of this large-scale modulating pattern could have various explanations. I explored whether such a kinematic pattern could emerge due to the projection of Galactic rotation using the rotation law of Eilers et al. (2019). The resulting kinematic signature does show a dependence between l and μ_l, but it is smaller in μ_l by almost an order of magnitude than that observed, and does not exhibit the saw-tooth pattern seen in Fig. 3.10. Additionally, the fit of the Galactic rotation curve calculated by Gaia Collaboration et al. (2023b) for their all-sky sample of B-type stars predicts a flat distribution in the Cygnus region, as shown in Fig. 3.17. This implies that the observed kinematic pattern originates not from Galactic rotation but from the local dynamics in Cygnus.

A correlation between distance and velocity in the same direction is often an indication of expansion (see e.g Wright et al. 2019; Armstrong et al. 2020), whose main invoked source could be feedback processes. Chevance et al. (2020) have argued that feedback within HII regions promotes velocities of order ∼15 km s^{-1}. Feedback would preferentially happen in opposite directions and consequently engenders a kinematic pattern of magnitude $\sim$ 30 km s^{-1}, very similar to the 25 km s^{-1} observed pattern. These phenomena were noticed by Chevance et al. (2020) in other galaxies on scales of 100–300 pc, again consistent with the scale of 160 pc observed in this work. This would suggest that the kinematic pattern in Cygnus was introduced by a previous generation of stars that exerted feedback on the surrounding gas clouds. Not only would they have formed the currently observed stars, but these new-born stars would have inherited their kinematics. This scenario is explored further in Chapter 4.

I also note that the 6 new OB associations are arranged in this large-scale kinematic pattern in groups of 3 OB associations. From analysis of the HR diagrams of these groups I estimated that the three youngest associations are part of one of these large-scale patterns and the three oldest associations are part of the second pattern.

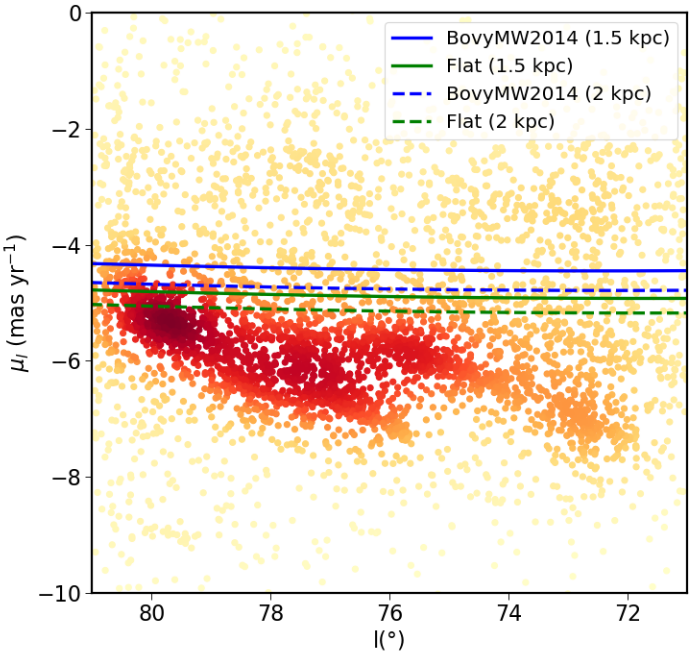


Figure 3.17: Proper motion in Galactic longitude as a function of Galactic longitude of the SED-fitted OB stars in the region, with the rotation curve models from Gaia Collaboration et al. (2023b) at distances of 1.5 and 2 kpc, using both a flat model (in green) and with the potential from Bovy (2014) (in blue).

This temporal link lends weight to the idea that a common origin was responsible for these two large-scale structures.

An argument against such an interpretation is that one would naively expect such an expansion signature to be present in the b direction as well as in l, yet this is not observed. This could be because of the restoring force of the Galactic disk inhibiting such expansion, or because the gas clouds expanding in the b direction did not form a new generation of stars, possibly because of the lower-density medium they expanded into that did not trigger further star formation.

Alternatively, the kinematic pattern I observe could be due to Galactic shear forces acting on the primordial giant molecular cloud, with the forming stars inheriting this motion. This interpretation has been argued to explain why the expansion of the Sco-Cen association is primarily in the Galactic Y direction (Wright & Mamajek, 2018). However, the timescales for Galactic shear to act are $\sim$ 70 Myr (Dobbs & Pringle, 2013), too short to explain the observed kinematics. Further observations and simulations appear necessary to explain this interesting kinematic pattern.

Regardless of the origin of this kinematic pattern, this encourages a deeper investigation and the motivation for the follow-up study of Chapter 4.

4 Large-scale expansion of OB stars in Cygnus

This chapter constitutes a follow-up of the previous chapter. Here I perform a 3D kinematic traceback on the new OB associations in Cygnus and find that both groups of three associations reached their most compact state ∼8 Myr ago. I consider whether this large-scale expansion pattern can be explained by feedback from a previous generation of stars or simply due to the intrinsic velocity dispersion within the primordial molecular cloud (as predicted by Larson's law). While both explanations are possible, the latter requires fewer assumptions and so is favoured, though feedback may still have played some role. This chapter is adapted from Quintana & Wright (2022).

4.1 Data

In this section, I reintroduce the new OB associations in Cygnus identified in Chapter 3 and outline the refinements made for this follow-up. In addition I introduce their related open clusters and how radial velocities (RVs) are determined for the groups of associations and open clusters.

4.1.1 Refinements of the new Cygnus OB associations

Chapter 3 (and Quintana & Wright 2021) describe the identification of six new kinematically-coherent OB associations inside the Cygnus region, labelled A to F. A further analysis led to the discovery of a bi-modal distribution in association B. This association is now divided between association B1 with 79 stars and B2 with 21 stars, respectively at lower and higher galactic longitudes and latitudes.

The SED fitting code has also been improved during the development of this thesis, as described in the latest sections of Chapter 2. I fitted again the new Cygnus OB associations with version 1.1 of the SED fitting code and found that ∼10 % of the OB members of the associations turn out to be cooler, A-type stars. I keep them because they remain kinematic members, and OB associations also contain low-mass stars (see e.g. Armstrong et al. 2020).

4.1.2 Related open clusters

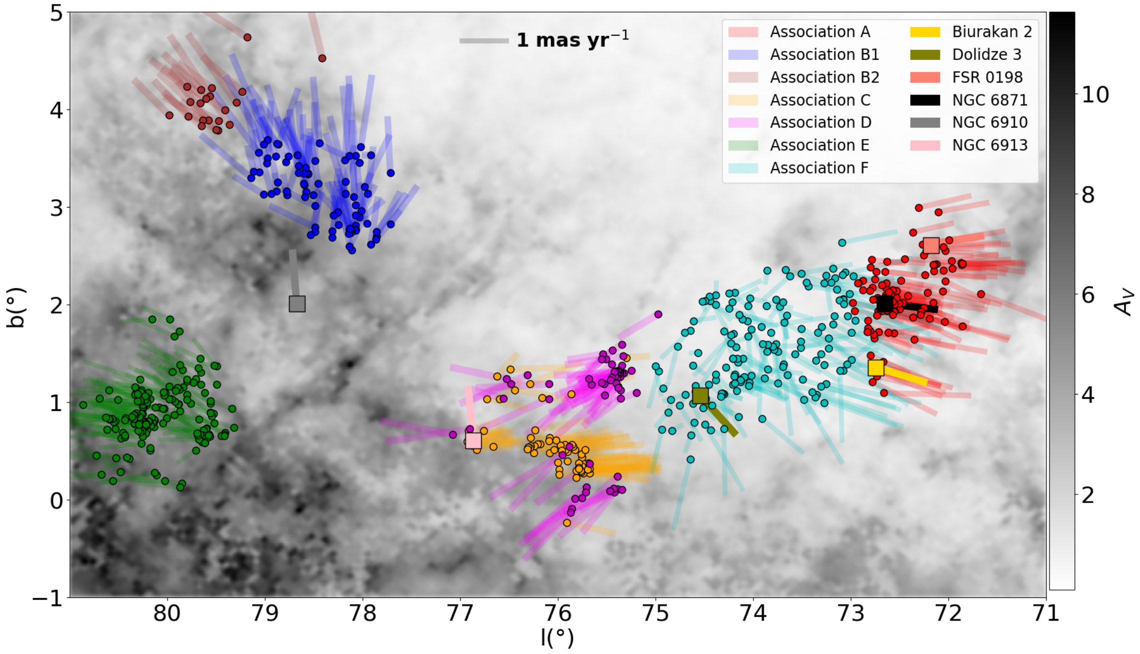


Figure 4.1: New OB associations in Cygnus and the six related open clusters in the region. The background map shows the integrated extinction up to 2 kpc from Green et al. (2019). The vectors indicate the proper motions of each star (and open clusters) subtracted by the median PMs of all the stars in the region. A representative 1 mas yr^{-1} proper motion vector is displayed on top of the plot: this is equivalent to a velocity of ~8.5 km s^{-1} at a distance of 1.8 kpc.

There are 31 open clusters (OCs) in the 60 deg^2 area of study (Cantat-Gaudin & Anders, 2020). This list is reduced to 17 when only selecting the OCs between 1 and 2.5 kpc. Out of those, I select six OCs whose PMs and RVs make them likely to be

Table 4.1: Galactic coordinates, PMs and distances of the six selected OCs in Cygnus from Cantat-Gaudin & Anders (2020).

Name	$l(\circ)$	$b(\circ)$	μ_α (mas yr^{-1})	μ_δ (mas yr^{-1})	d (pc)
NGC 6871	72.66	2.01	-3.13 ± 0.01	-6.44 ± 0.01	1841^{+4}_{-3}
NGC 6910	78.68	2.01	-3.41 ± 0.01	-5.36 ± 0.01	1741^{+6}_{-7}
NGC 6913	76.87	0.61	-3.41 ± 0.01	-5.77 ± 0.01	1719 ± 7
Biurakan 2	72.7	1.39	-3.17 ± 0.02	-6.84 ± 0.02	1751 ± 10
Dolidze 3	74.54	1.07	-2.87 ± 0.02	-5.61 ± 0.01	1907 ± 14
FSR 0198	72.18	2.61	-3.56 ± 0.03	-6.61 ± 0.03	1944^{+12}_{-11}

Table 4.2: Literature ages, RVs and total stellar masses for the six selected OCs in Cygnus. The labels for the references are: K05 = Kharchenko et al. (2005), P08 = Piskunov et al. (2008), C09 = Camargo et al. (2009), D14 = Dias et al. (2014), C17 = Conrad et al. (2017), C19 = Carrera et al. (2019), L19 = Liu et al. (2019) and K20 = Kaur et al. (2020).

Name	Age (Myr)	RV (km s^{-1})	Mass ($M_\odot$)
NGC 6871	7.0 ± 0.4 (L19)	-10.5 ± 2.2 (C17)	1148^{+801}_{-472} (P08)
NGC 6910	4.25 ± 1.50 (L19, K20)	-32.7 ± 2.1 (C17)	209^{+208}_{-104} (P08)
NGC 6913	5.0 ± 0.3 (L19)	-16.9 ± 0.6 (C17)	27^{+29}_{-14} (P08)
Biurakan 2	13.8 (K05)	-22.0 ± 9.5 (C17)	135^{+116}_{-63} (P08)
Dolidze 3	4.0 ± 0.2 (L19)	-7.7 ± 1.9 (C19)	200 ± 50
FSR 0198	10.0 ± 5.0 (C09)	-13.0 ± 3.7 (D14)	350 ± 100

related to the OB associations. All these OCs and associations are displayed in Fig. 4.1.

Tables 4.1 and 4.2 list the properties of the six related OCs. Piskunov et al. (2008) estimated masses for four of these clusters, though they should be considered as conservative since Le Duigou & Knödlseder (2002) provides larger masses for NGC 6910 and NGC 6913, respectively 580-1300 $M_{\odot}$ and 580-1090 $M_{\odot}$[1]. FSR 0198 and Dolidze 3 do not have a measured mass in the literature. Instead, I have fitted the mass function of SED-fitted members from Cantat-Gaudin & Anders (2020) to modelled mass function from Maschberger (2013), similar to the method used in Chapter 3, and estimated masses of 350 ± 100 $M_{\odot}$ and 200 ± 50 $M_{\odot}$, respectively.

4.1.3 Radial velocities

Applying a 3D kinematic traceback to the OB associations and related OCs requires velocity measurements in three dimensions. The availability of RVs is however limited compared to that of PMs, since spectroscopic observations are needed to measure them.

I gathered RVs from the literature by crossmatching (within 1 arcsec) the members of the new OB associations with SIMBAD. For all stars with a measured RV, I discarded those without any uncertainty on this value and whose quality measurement was too low to be considered reliable. This provided a sample of 93 stars with RVs across the 7 associations.

Unresolved close binaries can cause individual measured RVs to vary significantly from the binary system velocity. To overcome this I performed a Monte Carlo experiment to calculate the impact of unresolved binarity. I started by computing the median RV for each association and assigned this to all stars in the association. Since none of the stars in association B1 have measured RVs, I chose its RV to be identical to the nearby association B2. I then performed a Monte Carlo simulation to calculate the difference between an association's central velocity and a median that would be derived from N measured velocities for stars in that association. I assumed that each association has a 1D velocity dispersion equal to 2 km s^{-1} and sampled from this distribution. I then added randomly sampled instantaneous binary offset velocities, using a 100% binary fraction (as estimated in Cyg OB2, Kiminki & Kobulnicky 2012), along

---

[1] Being listed in the catalogue from Cantat-Gaudin & Anders (2020), and not in the list of asterisms from Cantat-Gaudin et al. (2018), these OCs are likely to be real.

Table 4.3: Radial velocities for the new OB associations. N_{stars} stands for the number of stars in the association whereas N_{RV} corresponds to the number of stars with a reliable measured RV. The fourth column lists the median RV of the association, and their uncertainty calculated as described in Section 4.1.3. References are: (1): Hayford (1932), (2): Abt (1973), (3): Huang & Gies (2006), (4): Gontcharov (2006), , (5): Kiminki et al. (2007), (6): Huang et al. (2010), (7): Chojnowski et al. (2017), (8): Gaia Collaboration et al. (2018), (9): Holgado et al. (2018), (10): Carrera et al. (2019)

Assoc.	$N_{\rm stars}$	$N_{\rm RV}$	RV (km $\rm s^{-1}$)	References
A	109	20	-12.50 ± 2.77	(1), (2), (3), (6)
B1	79	0		
B2	21	1	-21.00 ± 10.80	(4)
C	75	10	-19.20 ± 2.98	(1), (3), (6), (7), (10)
D	65	10	-7.00 ± 3.88	(1), (6), (10)
E	168	48	-12.35 ± 2.06	(5), (8), (9)
F	147	4	-13.48 ± 5.06	(3), (8), (10)

with mass ratios characterized by a power-law distribution and an index of 0.1, over the range 0.005 to 1, periods from 1 to 1000 days with a power-law distribution and an index of 0.2, and ellipticities from 0.0001 to 0.9 with a power-law distribution and an index of -0.6 (Kiminki & Kobulnicky, 2012). I added these velocity offsets to the computed velocities together with the measured uncertainties. This process was iterated 100,000 times, with each iteration providing an estimate of the true central velocity. The median of these iterations gave the best-fitting central velocity, and the 16th and 84th percentile values serve as lower and upper uncertainties, respectively. These results are displayed in Table 4.3, where it is seen that associations with sparsely-sampled RVs have a larger uncertainty on the median RV.

4.2 Kinematic traceback

This section is dedicated to the kinematic traceback of the OB associations and related OCs. To start, I calculate the Galactic space velocities UVW for each star, using the following equations:

$$U = \mathrm{RV}\ \cos(l)\ \cos(b) - V_l\ \sin(l) - V_b\ \cos(l)\ \sin(b) \tag{4.1}$$

$$V = \mathrm{RV}\ \sin(l)\ \cos(b) + V_l\ \cos(l) - V_b\ \sin(l)\ \sin(b) \tag{4.2}$$

$$W = \mathrm{RV}\ \sin(b) + V_b\ \cos(b) \tag{4.3}$$

where $V_l = 4.74\,\mu_l\,d$ is the transverse velocity in the l direction (and equivalent in the b direction), where μ_l (and μ_b) is in units of mas yr^{-1} and d in units of kpc.

These velocities are calculated for individual stars for the OB associations (as they are gravitionally unbound) while I use a single value for the OCs (as they are gravitationally bound). They are then corrected for the Sun's motion relative of the local standard of rest (LSR), with $(U_\odot, V_\odot, W_\odot) = (11.10, 12.24, 7.25)$ km s^{-1} from Schönrich et al. (2010).

An epicycle approximation is required for the kinematic traceback, as has been applied in Wright & Mamajek (2018) and Squicciarini et al. (2021). To that end I use the Oort constants $A = 14.82 \pm 0.84$ km s^{-1} kpc^{-1} and $B = -12.37 \pm 0.64$ km s^{-1} kpc^{-1}, along with the angular velocity of the local standard of rest, $\Omega_0 = 27.19 \pm 0.87$ km s^{-1} kpc^{-1}, each from Feast & Whitelock (1997). Furthermore, I use the local density value $\rho_0 = 0.1$ M$_\odot$ pc^{-3} from Holmberg & Flynn (2004). These parameters allow us to determine the epicycle frequency $\kappa = \sqrt{-4\,\Omega_0\,B}$ and the vertical oscillation frequency $\nu = \sqrt{4\,\pi\,\rho_0\,G}$ where G stands for the gravitational constant.

The epicycle equations from Fuchs et al. (2006) are written as follows:

$$X(t) = X(0) - \frac{V(0)}{-2\,B}\,[1 - \cos(\kappa\,t)] + \frac{U(0)}{\kappa}\ \sin(\kappa\,t) \tag{4.4}$$

$$\begin{split} Y(t) = Y(0) + 2\,A\left[X(0) - \frac{V(0)}{-2B}\right]t + \frac{\Omega_0}{-B\,\kappa}\,V(0)\ \sin(\kappa\,t) \\ + \frac{2\,\Omega_0}{\kappa^2}\,U(0)\,[1 - \cos(\kappa\,t)] \end{split} \tag{4.5}$$

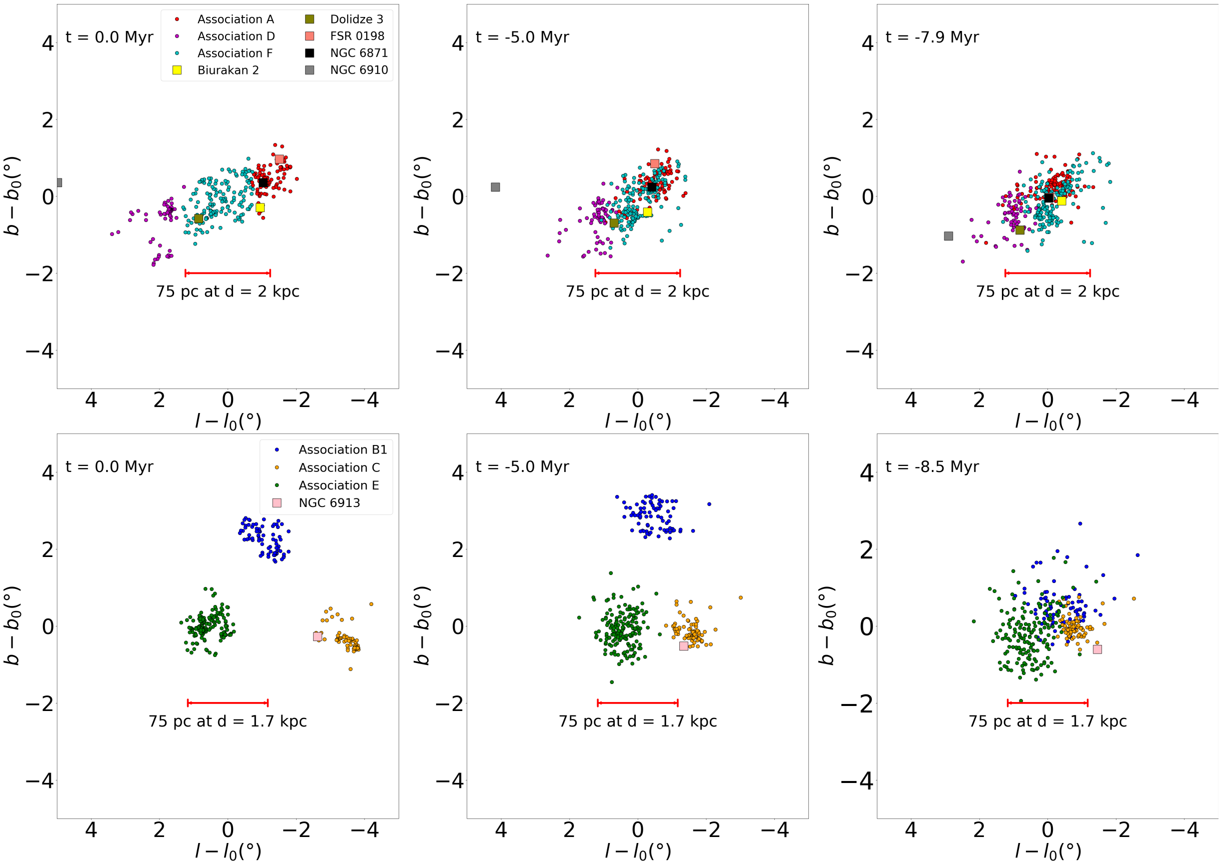


Figure 4.2: Relative galactic coordinates of all stars and OCs at various times in the past derived from traceback calculations in their reference frame. Here, l_0 and b_0 stand for the median galactic coordinate of each group.

$$Z(t) = \frac{W(0)}{\nu}\sin(\nu\, t) + Z(0)\cos(\nu\, t) \tag{4.6}$$

where 0 denotes the coordinates and velocities at the present time and t designates the time in the past I perform the traceback.

With these equations, I calculate XYZ coordinates at times t, up to 20 Myr in the past, at step times of 0.1 Myr. The projected positions of all stars and OCs are displayed as a function of time in Fig. 4.2 at certain points of the past. Line-of-sight distances are shown in Fig. 4.3, with the median distance to each association instead of individual stars because of the large uncertainties on individual distances.

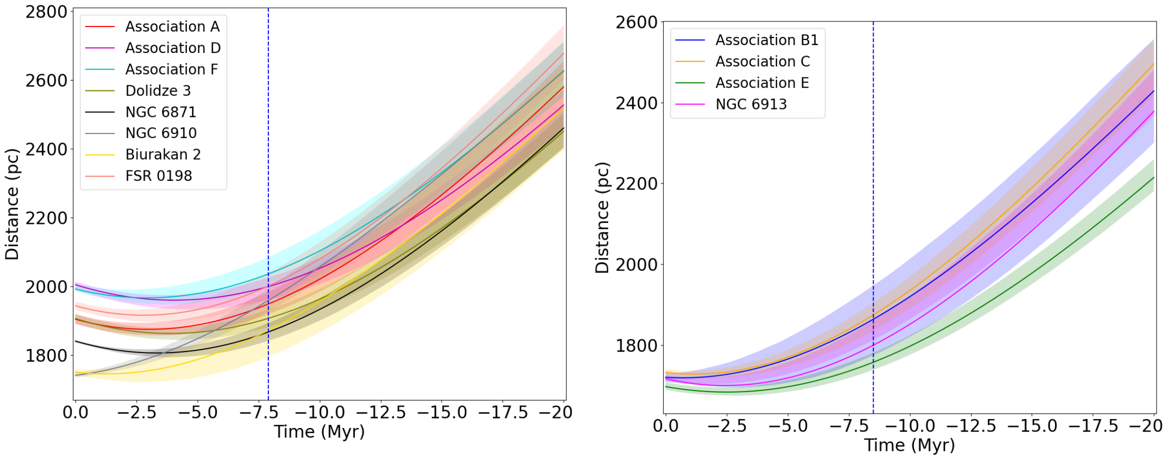


Figure 4.3: Line-of-sight distance as a function of look-back time for the associations and related OCs. Filled areas delimit the 16th and 84th percentiles of distance at each time step. Left panel: associations A, D and F with their related OCs. Right panel: associations B1, C and E with their related OC. The dashed blue line indicates the time of closest approach on the sky for each group.

As predicted from the kinematic signature of expansion identified in Chapter 3, both groups ADF and B1CE show stars closer together on the plane of the sky in the recent past than they are at the current time. This is verified not only in the l direction (where the kinematic signature of expansion was identified), but also in the b direction where the expansion signatures tended to be weaker (associations A, D and E) or non-existent (associations B1, C and F). Associations ADF are also closer together along the line-of-sight in the recent past, although associations B1CE move further apart in the recent past. It has to be noted, though, that the uncertainties on distance and RV are so large that this is not significant.

I subsequently calculate the distance, d, from each star to the central (median) position of each group (either group ADF or group B1CE). Because of the imprecision of the line-of-sight distances and RVs relative to the plane of the sky positions and velocities, I apply this calculation only in the plane of the sky. I sum the distances at each time step for each of the groups and determine the time of the most compact configuration as the time this sum is minimised. To compute the uncertainties, again I apply a Monte Carlo process with 1000 iterations, randomly dispersing all measured quantities in each iteration according to their uncertainties (including all constants used in the epicycle calculations) and using the 16th and 84th percentile values of the resulting time distribution as the upper and lower 1σ uncertainties.

The resulting times of most compact configuration are $7.9^{+3.0}_{-1.8}$ Myr for group ADF and $8.5^{+0.8}_{-2.8}$ Myr for group B1CE. The distribution of stars at these times are shown in the rightmost panels of Fig. 4.2.

4.3 Discussion

The two groups of three OB associations in Cygnus, along with multiple related OCs, have been shown to have a more compact configuration in the past, through 3D kinematics. This compact configuration occurs 7.9 and 8.5 Myr ago, respectively for groups ADF and B1CE, as illustrated in Fig. 4.2.

These calculated traceback ages are higher than the evolutionary ages estimated for these associations. In Chapter 3, I estimated that associations B, C and E are 3-5 Myr old while associations A, D and F are older: some of their stars aged up to ∼10 Myr in them. By contrast, the OCs that are part of these expansion patterns are 5-10 Myr old, with the exception of Biurakan 2, whose literature age does not have an uncertainty (such that the accuracy of this value is unsure). Given this information, it turns out that the identified large-scale expansion started before these stars and clusters formed. This means that the driving force of the expansion would have acted on the primordial molecular clouds and not the stars observed now. The traceback ages are also consistent with the expansion ages for the individual OB associations estimated in Chapter 3 (which are greater than the evolutionary ages for the associations). This suggests that the expansion of the individual OB associations was not driven by their own stellar dynamics or processes such as residual gas expulsion, as predicted by the linear expansion model of Blaauw (1964), but was rather seeded before formation, potentially by the same process responsible for the large-scale expansion patterns analysed here. This chronological pattern was also observed in λ Ori (Kounkel et al., 2018).

I suggested, in Chapter 3, that the cause of this large-scale expansion pattern could be feedback from a previous generation of stars. Not only would the large-scale coherent motions and the more compact configuration support this argument, but this was also invoked in similar studies (Kounkel, 2020; Großschedl et al., 2021). Nevertheless, this expansion pattern could alternatively be the consequence of the intrinsic turbulence present in the parent molecular cloud, as predicted by Larson's law (Larson, 1981). Both scenarios are discussed below.

4.3.1 Feedback

The first explanation is that the expansion pattern was engendered by two major feedback events that occurred in Cygnus ∼8 Myr ago. Since the kinematic trend of each group is distinct (see Chapter 3), these events would be separate events, albeit close in time. By sweeping up the molecular gas in this area, these feedback events could have potentially triggered star formation within them. This picture is similar to the classical feedback-driven scenario from Elmegreen & Lada (1977), as pointed out by Großschedl et al. (2021), although in this case the triggering mechanism is preferentially the 'compress and collapse' scenario.

To verify the validity of this scenario, I estimate how much momentum and kinetic energy can be found in the expanding motion of the stars. To do this I take inspiration from the method of Großschedl et al. (2021). With the current observations, I only have access to the stars rather than the primordial molecular cloud. I estimate the masses of the primordial molecular clouds through:

$$M_{\mathrm{cloud}} = \frac{M_{\mathrm{assoc}}}{\mathrm{SFE}} \tag{4.7}$$

M_{cloud} and M_{assoc} stand respectively for the cloud and OB association masses (the latter taken from Chapter 3) while SFE is the star-formation efficiency, that I assume is equal to 5 %. I apply the same equation for the OCs.

I calculate the velocity of each association and OC (U, V, W) relative to the mass-weighted mean velocity of each group of associations (U_0, V_0, W_0) from the following equation:

$$v_{\rm rel} = \sqrt{(U - U_0)^2 + (V - V_0)^2 + (W - W_0)^2} \tag{4.8}$$

Eq. (4.7) and Eq. (4.8) are then used to derive the momentum and kinetic energy injected into each system, assuming that they were previously at rest (a highly simplistic assumption).

$$p_{\rm cloud} = M_{\rm cloud}\, v_{\rm rel} \tag{4.9}$$

$$E_{\rm kin,cloud} = \frac{1}{2}\, M_{\rm cloud}\, v_{\rm rel}^2 \tag{4.10}$$

The total kinetic energy injected is 0.54×10^{50} erg for group ADF and 1.21×10^{50} erg for group B1CE, whereas the total momentum is 0.66×10^6 $M_\odot$ km s^{-1} for group ADF and 1.26×10^6 $M_\odot$ km s^{-1} for group B1CE.

By comparison, the total energy output from a supernova explosion is approximately 10^{51} erg (see e.g. Janka 2012). This level is similar to that computed for the groups in Cygnus. On the other hand, simulations have predicted that the total momentum injected into a surrounding molecular cloud by a supernova is between 2 and 4×10^5 $M_\odot$ km s^{-1} (Geen et al., 2015; Kim & Ostriker, 2015; Walch & Naab, 2015), thus between a factor 1.5 and 6 times smaller than I have calculated for the Cygnus feedback events. Furthermore, Kim & Ostriker (2015) simulated the feedback generated from ten supernovae and estimated that the final momentum would be in the range of 14 to $22 \times 10^5 M_\odot$ km s^{-1}, which is closer to my estimates. Either several supernovae would have contributed to the feedback events to explain these results, or the observed motions were produced by several feedback types, such as the combination of photo-ionization, radiation pressure, stellar winds and supernovae.

Similar kinematic patterns showing the expansion of large complexes of stars have been observed by Kounkel (2020) and Großschedl et al. (2021) in Orion, respectively through the 3D dynamics of stellar groups and the motion of star-forming clouds, and Drew et al. (2021) in Carina through the motion of OB stars. The large-scale expansion pattern in Orion identified by Kounkel (2020) was attributed to a supernova explosion which happened ∼6 Myr ago. Großschedl et al. (2021) observed the same expansion pattern in Orion, considered various sources of feedback and concluded that a combination of feedback sources was likely to be responsible. Großschedl et al. (2021) estimated the total kinetic energy in the expanding structures in Orion to be (3.5-9.6) $\times 10^{48}$ erg and the total momentum to be (0.7-1.3) $\times 10^5$ $M_\odot$ km s^{-1}. These estimates are significantly lower than my calculated kinetic energies and momenta, which means that both hypothetical feedback events in Cygnus would have to be considerably more energetic than the Orion event. This is not surprising given that the Cygnus star-forming complex is significantly larger and more massive than the Orion region.

4.3.2 Intrinsic turbulent motions

In Chapter 1, I introduced Larson's law as describing the size-linewidth relationship through a power-law between the 1-dimensional size of a molecular cloud and its intrinsic 1D velocity dispersion, as caused by turbulence (Larson, 1981; Solomon et al., 1987; Bolatto et al., 2008; Heyer & Dame, 2015; Miville-Deschênes et al., 2017). If I assume that an intrinsic distribution of velocities present in the primordial molecular cloud is passed onto the stars that form from that cloud, and if that velocity dispersion was high, without any significant restoring force, this would be observed as an expansion pattern in the stars formed from that molecular cloud.

I test whether Larson's law can explain the observed motions. To do this, I compare the velocity dispersion predicted by Larson's law for an appropriately sized molecular cloud and the velocity dispersion of the expanding stars in Cygnus. A size of approximately 75 pc can be deduced from Fig. 4.2 at the most compact time for both Cygnus groups. I find that a molecular cloud with a size of 75 pc would be expected to have a 1D velocity dispersion of 5.7 - 8.7 km s^{-1}, depending on the exact relationship used (Larson, 1981; Solomon et al., 1987; Miville-Deschênes et al., 2017). I compute the actual 3D velocity dispersions using the UVW velocities at the most compact time for both group ADF and B1CE, respectively finding 7.1 and 11.3 km s^{-1}, equivalent to 1D velocity dispersions of 4.1 and 6.5 km s^{-1}.

It turns out, from these results, that the observed velocity dispersions within the groups are perfectly consistent with arising from Larson's law in the primordial cloud. This would imply that this large-scale expansion pattern can be explained by the turbulence present in the molecular cloud that formed these OB associations and OCs. Feedback is thought to be one of the drivers of this turbulence (Miville-Deschênes et al., 2017), but no additional momentum is required to produce the observed expansion pattern in Cygnus.

This result has many implications. I repeat this calculation for Orion using the results from Großschedl et al. (2021). Their estimate of the primordial Orion cloud size is ~40 pc from their Fig. 7. Applying Larson's law gives a 1D velocity dispersion of 4.5-6.3 km s^{-1}. Performing a similar calculation to the Orion sub-regions listed in Table 4 from Großschedl et al. (2021) (using their UVW velocities in the local standard of rest as listed in the table), I find a 3D velocity dispersion of 4.8 km s^{-1}, equivalent to a 1D velocity dispersion of 2.8 km s^{-1}. Thus, the observed motions in Orion are also consistent with Larson's law and do not require an additional source of acceleration to reproduce what is observed.

Previous studies (Kounkel, 2020; Drew et al., 2021; Großschedl et al., 2021) invoked feedback as the main driver of these large-scale expansion patterns identified in massive star-forming regions. However, using Larson's law shows that turbulence can be the main driver instead, without the requirement for an additional source of momentum. This has implications for the understanding of these phenomena.

4.4 Summary and implications

In this chapter, I have studied in greater depth the two large-scale expansion patterns in Cygnus identified in Chapter 3, applying a 3D kinematic traceback on the OB associations and related OCs. I found that both groups reached their most compact state ~8 Myr ago. By examining the velocity field of the young stars, I showed that the driver of such pattern is likely to be the turbulence within the primordial molecular clouds, as predicted by Larson's law.

This result has implications for studies of expansion in other star-forming complexes, many of which have sought to explain such large-scale expansion events as being due to the effects of feedback. My calculations show however that the observed expansion patterns can more simply be explained as being inherited from the gas from which they formed, which itself has kinematics that follow Larson's law. Future studies should explore the possibility of turbulence being the expansion driver, and contrast it with stellar feedback.

5 Mapping the distribution of OB stars and associations in Auriga

This fifth chapter presents the results of the analysis of the OB stars and associations in Auriga. I describe that, similarly to Cygnus, the historical Auriga OB associations lack kinematic coherence when analysed with *Gaia* data, before detailing the selection and clustering process used to identify new OB associations. I analyse them physically and kinematically, connect them to the surrounding open clusters and star-forming regions, and discuss them in the wider context of Galactic structure. Notably, the new Auriga OB associations are found to follow an age gradient whose position coincides with the motion of the Perseus spiral arm over the last ∼20 Myr. This chapter was published in Quintana et al. (2023).

5.1 The Auriga region

This section outlines a new picture of the existing OB associations in Auriga using recent data, more specifically parallaxes and proper motions from *Gaia* EDR3 (Gaia Collaboration et al., 2021). The star-forming regions and open clusters in the vicinity, likely to be connected to the OB associations, are also presented. I focus on a 150 deg^2 area, limited by the Galactic coordinates l = [165°, 180°] and b = [-5°, 5°].

5.1.1 Historical associations

Two OB associations, Aur OB1 and OB2, are located within this area. While their members have been listed in Humphreys (1978), I use the most recent list from Melnik & Dambis (2020), since they already crossmatched them with *Gaia* DR2. This catalogue compiles 36 stars in Aur OB1, 20 stars in Aur OB2 and 10 stars in NGC 1893. Unfortunately, only 6 out of these 10 stars in NGC 1893 have corresponding equatorial coordinates in SIMBAD that allow a correspondence in *Gaia* EDR3. Since NGC 1893 is usually considered to be part of Aur OB2 (see e.g. Marco & Negueruela 2016; Lim et al. 2018), I include them in Aur OB2 and thus increase its number of members to 26 stars.

A crossmatch within 1 arcsec is performed between these 62 stars and *Gaia* EDR3, with all stars having an identifiable counterpart. I however only use the astrometry for the 48 stars with $RUWE < 1.4$ following the condition from Lindegren et al. (2018). I also take the geometric distances from Bailer-Jones et al. (2021). A view in position, proper motion and distance of the Aur OB1 and OB2 members is shown in Figs. 5.1, 5.2 and 5.3.

From Figs. 5.1 and 5.2, the existing members of Aur OB1 and OB2 exhibit a large spread in proper motions (over 2-3 mas yr^{-1} or 10-15 km s^{-1} at 1 kpc) and consequently lack the kinematic coherence expected for OB associations, whose velocity dispersions are typically a few km s^{-1} (Wright, 2020). Fig. 5.3 unveils a similar picture with respectively members from Aur OB1 located at distances between 0.6 to > 2 kpc and members from Aur OB2 at distances between 1.7 and > 4 kpc (albeit the latter possesses a core group of stars at $\sim$2 kpc). The presence of stars at 3-4 kpc within these OB associations, previously noted by Marco & Negueruela (2016), alongside the absence of correspondence between the stars at $\sim$2 kpc in Aur OB2 and NGC 1893 located further away at $\sim$2.9 kpc (Mel'Nik & Dambis, 2009; Melnik & Dambis, 2020), reinforces this view.

Such a result is unsurprising since a similar trend was observed for the Cygnus OB associations (see Chapter 3): historically, OB associations were defined through the photometry of their brightest members and their on-sky spatial distribution (see e.g. Humphreys 1978). Without any strong kinematic coherence, nor a narrow distribution of distances, the historical members from Aur OB1 and OB2 are unlikely to have been born together and therefore to be genuine OB associations.

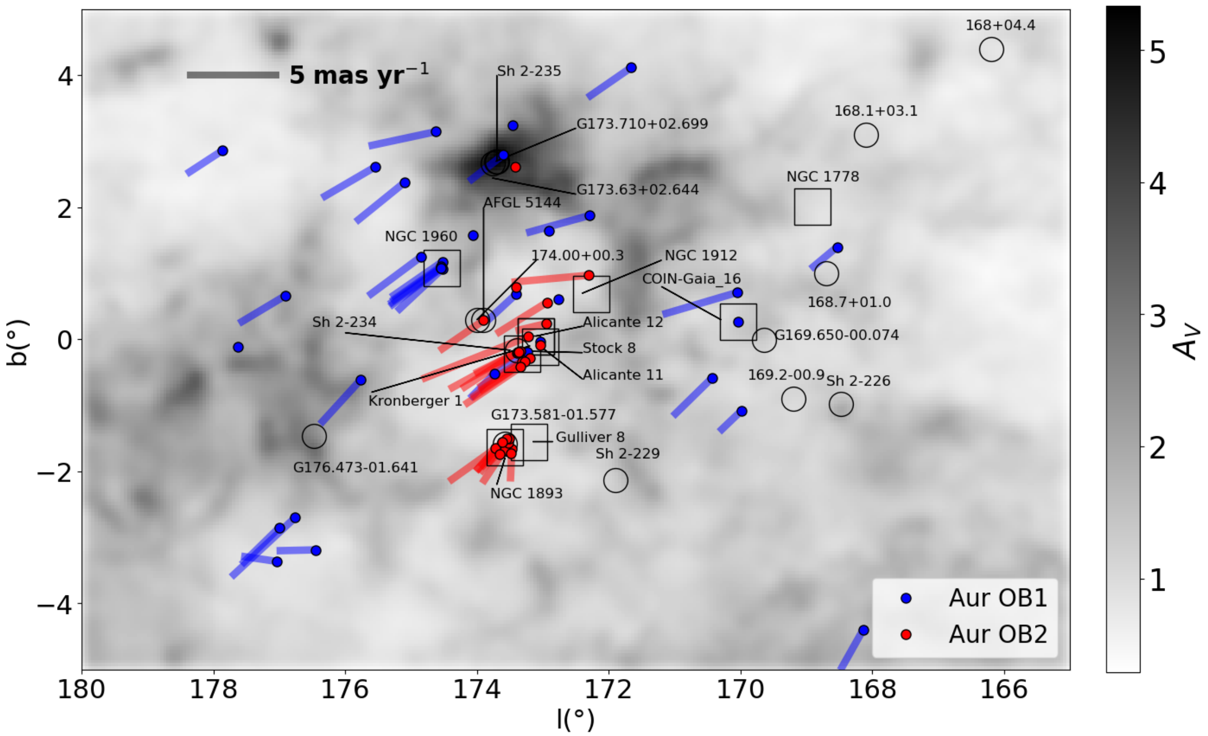


Figure 5.1: Historical members of Aur OB1 and OB2 associations in Galactic coordinates. The 48 stars with reliable astrometry have their Galactic proper motions displayed as vectors, whose scale length is shown in the top left. The other stars do not have reliable proper motions and therefore are shown as points. In this figure are also included open clusters as empty squares (Cantat-Gaudin & Anders, 2020), and HII and star-forming regions as empty circles (Paladini et al., 2003; Mellinger, 2008; Anderson et al., 2015). The background extinction map, from Green et al. (2019), shows the integrated visual extinction at a distance of 2 kpc.

5.1.2 Open clusters and star-forming regions

The reality of OB associations also lies in their relation with other objects. This requires me to gather information about the surrounding open clusters and star-forming regions. Cantat-Gaudin & Anders (2020) identified several tens of open clusters in this area, five of which are probably related to the historical OB associations, as suggested by Straižys et al. (2010), Marco & Negueruela (2016) and Pandey et al. (2020). I summarize the

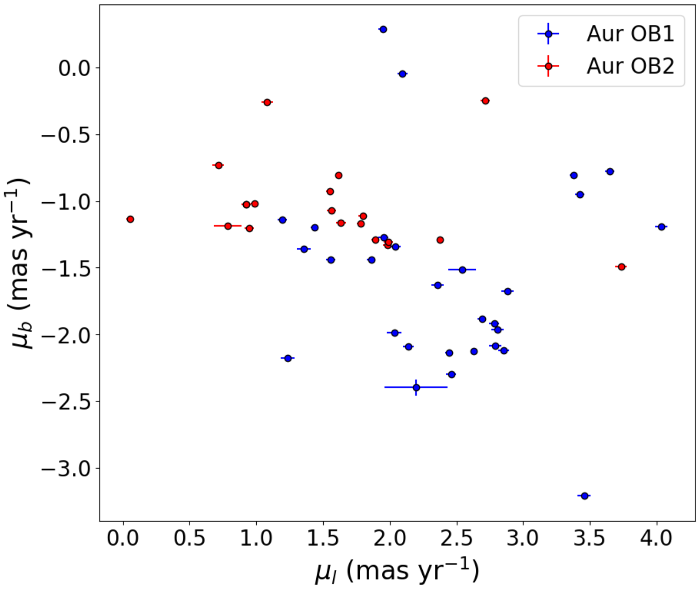


Figure 5.2: Galactic proper motions of the 48 stars in Aur OB1 and OB2 with reliable astrometry ($RUWE < 1.4$).

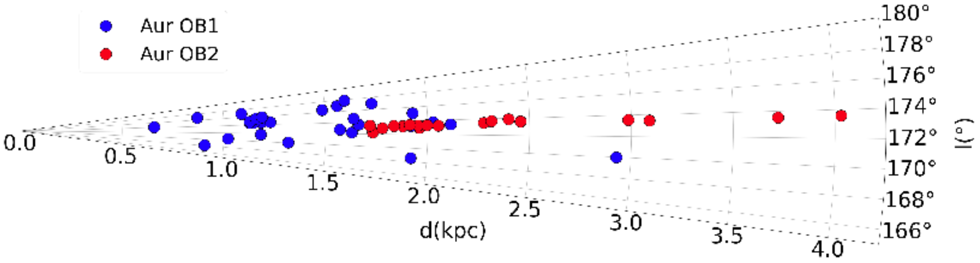


Figure 5.3: Galactic longitude shown as a function of geometric distance from Bailer-Jones et al. (2021) for the 48 stars in Aur OB1 and OB2 with reliable astrometry ($RUWE < 1.4$).

properties of these OCs in Table 5.1. In Table 5.1 are also included Alicante 11 and 12, though they are not listed in Cantat-Gaudin & Anders (2020), because of their relation with Stock 8. Marco & Negueruela (2016) found a common distance of ∼2.8 kpc for these three OCs, even if this is overestimated compared with other estimations in e.g. Jose et al. (2008) and Melnik et al. (2020). Table 5.1 also lists other OCs in the region whose relevance will be detailed in Section 5.2.5. The clusters are displayed in Fig. 5.1 together with the historical OB associations.

Table 5.1: Properties of the open clusters in Auriga likely related to the OB associations. Galactic coordinates and distances were taken from Cantat-Gaudin & Anders (2020). The references for the OCs ages are respectively: Jeffries et al. (2013) and Joshi et al. (2020) for NGC 1960, Marco & Negueruela (2016) for Stock 8, Alicante 11 and 12, Tapia et al. (1991), Marco et al. (2001), Sharma et al. (2007) and Lim et al. (2014) for NGC 1893, Subramaniam & Sagar (1999), Dias et al. (2021) for Gulliver 8, Jacobson et al. (2002), Pandey et al. (2007), Kharchenko et al. (2005) and Dib et al. (2018) for NGC 1912, Barbon & Hassan (1973), Kharchenko et al. (2013), Dib et al. (2018) and Cantat-Gaudin et al. (2020) for NGC 1778, Cantat-Gaudin et al. (2020) for COIN-Gaia_16 (here abbreviated CG16), Dib et al. (2018) and Cantat-Gaudin et al. (2020) for Kronberger 1 (here abbreviated K1).

OC	Assoc.	$l\,(^\circ)$	$b\,(^\circ)$	d (kpc)	Age (Myr)
NGC 1960	Aur OB1	174.542	1.075	1.16 ± 0.01	18-26
Stock 8	Aur OB2	173.316	-0.223	2.11 ± 0.01	4-6
Alicante 11	Aur OB2	173.046	-0.119	2.11 ± 0.01	4-6
Alicante 12	Aur OB2	173.107	0.046	2.11 ± 0.01	4-6
NGC 1893	Aur OB2	173.577	-1.634	3.37 ± 0.05	1-5
Gulliver 8	-	173.213	-1.549	1.11 ± 0.01	22-39
NGC 1912	-	172.270	0.681	1.10 ± 0.01	250-375
NGC 1778	-	168.914	2.007	1.64 ± 0.01	150-282
CG16	-	170.038	0.270	$1.53^{+0.02}_{-0.01}$	26
K1	-	173.106	0.049	2.12 ± 0.06	6-8

Several HII regions (Paladini et al., 2003; Anderson et al., 2015), along with star-forming regions such as Sh 2-235 and AFGL 5144 (Mellinger, 2008) are located within this area. They are also shown in Fig. 5.1.

Looking in more detail at Fig. 5.1, the main feature is found at the centre of the region ($l \sim 173^\circ$ and $b \sim 0^\circ$). Most of the Aur OB2 members lie there (Melnik & Dambis, 2020), as well as its three related OCs Stock 8, Alicante 11 and 12 (Table 5.1) and the HII regions Sh 2-234 and 174.0+00.3. Close to this feature is the star-forming region AFGL 5144 at $l = 173.7°$ and $b = 0.3°$ (Mellinger, 2008), consistent with the young age of the OCs (Marco & Negueruela, 2016).

Another important feature of Fig. 5.1 is the region of highest extinction, wherein lies the star-forming region Sh 2-235 at $l = 173.7°$ and $b = 2.7°$ (Mellinger, 2008), itself close to the HII regions G173.710+02.699 and G173.63+02.664. Few members of Aur OB1 and OB2 can be found there, hinting at the possible presence of hidden features within this area, as it was theorised that OB associations could be the expanded and unbound remnants of embedded clusters (Lada & Lada, 2003).

5.2 Identification of new OB associations

In this section, I will describe the method applied to identify OB stars and associations. In particular, for this dataset, a different clustering process is performed compared to in Chapter 3 and this will be detailed.

5.2.1 Data and selection process

Firstly, I queried the *Gaia* EDR3 database from the *Gaia* archive. I imposed the conditions $RUWE < 1.4$ and $\frac{\varpi}{\sigma_{\varpi}} > 2$[1]. Furthermore, I have only selected stars with BP-RP < 2.5. This colour limit roughly corresponds to a star with $\log(T_{\mathrm{eff}}/K) = 4$ and $A_V = 6$, as tested with the model SEDs, approximately the maximum extinction found in this region at a distance of 3 kpc (Green et al., 2019). I also selected stars with $d < 3.5$ kpc, using the geometric distances from Bailer-Jones et al. (2021), in order to encompass the volume spanned by both the historical OB associations and their related open clusters. This query produced an initial sample of 1,684,696 sources.

---

[1] This time, the correction on parallax uncertainty from El-Badry et al. (2021) was not applied, due to the line-of-sight distribution of OB stars within the new associations suggesting that the *Gaia* parallax uncertainties are overestimated in this area

Subsequently, I crossmatched this list with 2MASS, IGAPS and UKIDSS, before checking the photometry for each of the twelve photometric bands. The criteria were identical to those used in Chapter 3, and I kept only the sources with a least one good photometric blue band (g, $G_{\rm BP}$ or G) and at least one good NIR photometric band, reducing the sample to 1,525,978 sources.

Faints stars (likely non-OB stars) were removed through an absolute NIR magnitude cut. Again the NIR bands were chosen due to their lower sensitivity to interstellar extinction (Fitzpatrick et al., 2019). This translated into requiring $M_K < 1.07$ (if K-band photometry is available), $M_H < 1.10$ (otherwise if H-band photometry is available), or $M_J < 1.07$ (if only J-band photometry is available). These thresholds were chosen as the absolute magnitudes of main-sequence A0 stars (Pecaut & Mamajek, 2013) and trimmed the sample down to 85,245 sources.

The final cut consisted of discarding background giants, by applying a cut in a near-IR colour-colour diagram, as performed in Chapter 3. The resulting working sample contains 29,124 sources.

5.2.2 General results

The version 1.2 of the SED fitting code (with the scaled extinction from *Bayestar*, as described in Chapter 2) has been applied to these 29,124 candidate OB stars. The results of this process are shown as histograms of the fitted physical parameters in Fig. 5.4.

5434 stars have been fitted with $\log(T_{\rm eff}/K) > 4$ (OB stars, 18.66 %) and 115 stars with $\log(T_{\rm eff}/K) > 4.3$ (O stars, 0.39 %). $\log(M/M_{\odot})$ has a median value of 0.31 (with a standard deviation of 0.12 dex) whereas the median value of $\log(L/L_{\odot})$ is 1.43 (with a standard deviation of 0.44 dex), indicating the presence of many late B-type stars and cool giants despite the previous cuts. This highlights the value of a full SED fit when identifying hot and luminous stars. The majority of stars can be found within 4 kpc, which is consistent with our selection from Bailer-Jones et al. (2021), with an increasing number at larger distances (as a bigger volume is probed). The extinction

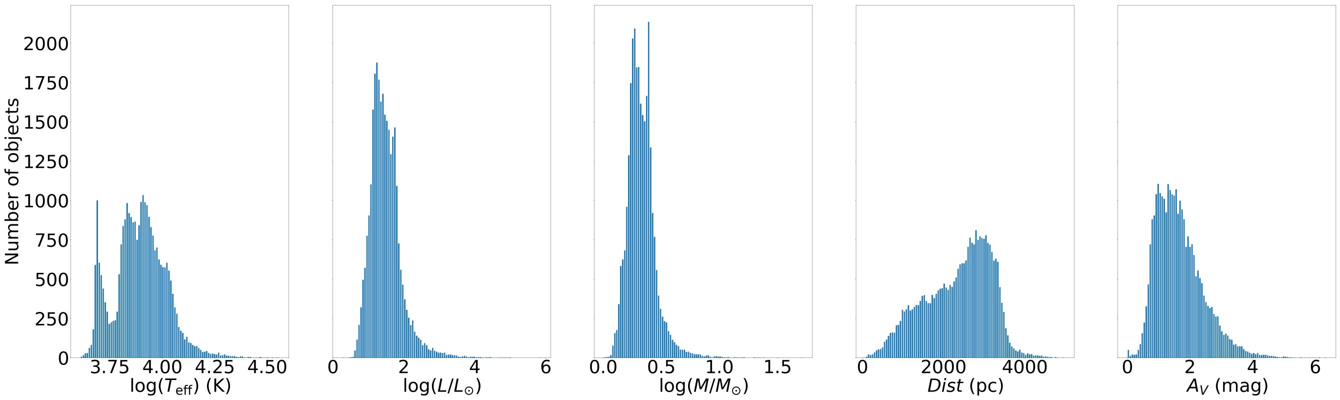


Figure 5.4: Median values of the SED-fitted parameters of the 29,124 candidate OB stars of the working sample.

is relatively low, with a peak at ∼1.5 mag and most of the stars with an extinction smaller than 3 mag.

5.2.3 Incompleteness

Similarly to Chapter 3, I estimated the level of incompleteness in the working sample by calculating the fraction of stars lost during the selection process described in Section 5.2.1, as a function of magnitude. Cuts on data quality (astrometry and photometry) can trim the number of true OB stars. The astrometric cuts include the removal of 2-parameter sources in *Gaia* EDR3 (without any measured parallax and proper motion), plus stars with a large error on measured parallax and $RUWE$. The photometric cuts include the removal of bad blue and NIR photometry, together with stars with too high BP-RP values. I have displayed the completeness level as a function of G magnitude in Fig. 5.5 for the SED-fitted OB stars (stars with $\log(T_{\rm eff}/K) > 4$) and luminous evolved stars ($\log(L/L_\odot) > 2.5$).

Another way to verify the completeness of the sample is to compare with another catalogue of OB stars. The list from Zari et al. (2021), who mapped luminous hot stars throughout the entire sky, is an ideal candidate for such purpose. 14,973 OBA stars

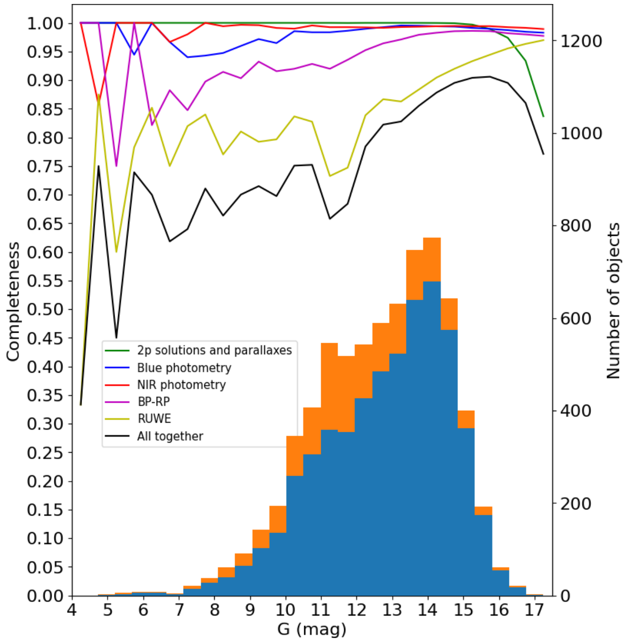


Figure 5.5: Completeness as a function of G magnitude for the 5617 SED-fitted OB stars in the sample divided according to the different steps performed to trim the sample. The black curve corresponds to the product of all completeness curves, while the blue and orange histograms show the number of sources before (blue) and after (orange) the completeness correction is applied. The bin size is 0.5 mag.

from their sample are located within the Auriga region. By comparison, my sample in Auriga comprises 5617 SED-fitted OB stars. I have crossmatched both catalogues within 1 arcsec, and divided the results in three samples. The list A corresponds to the successful crossmatches and contains 2155 stars. The list B corresponds to the unsuccessful crossmatches from my catalogue and contains 3462 stars. Finally, the list C corresponds to the unsuccesful crossmatches from the Zari et al. (2021) and contains 12,818 stars. List B includes 3279 (out of 3462 stars) with $M_K > 0$ mag, which is the threshold used by Zari et al. (2021) (instead of my more liberal cut of $M_K < 1.07$ mag). Most of the remaining stars from list B did not pass the photometric cuts defined in Zari et al. (2021) (their equations 2 to 5). List C, on the other hand, includes 9052 stars with a geometric distance from Bailer-Jones et al. (2021) higher than 3.5 kpc, 988 stars with bad astrometry, 88 stars with bad photometry and 26 stars removed through the NIR colour-colour diagram, all the steps from the selection process outlined in Section 5.2.1. Finally, in list C, there are 2663 stars found in my final working sample that I did not fit as OB stars (including 1942 stars fitted with T_{eff} greater than 8000 K, which corresponds to the threshold of minimum temperature chosen by Zari et al. 2021). Henceforth, the differences between both catalogues can be attributed to the selection process itself.

5.2.4 Clustering analysis with HDBSCAN

In Chapter 3, the clustering analysis used to identify OB associations in Cygnus was based on a Kolmogorov-Smirnov (KS) test on the Galactic coordinates and proper motions of the SED-fitted OB stars. This plane-of-the-sky approach was used because the historical Cygnus OB associations were mostly located between 1.5 and 2 kpc, while I limited my sample to distances between 1 and 2.5 kpc. Here, however, I am probing a much larger volume (from 0 to 3.5 kpc) due to the larger spread in distance of the historical Auriga OB associations. The distance component cannot be neglected anymore and must therefore be included in the clustering analysis.

To that end, I chose to use the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN, McInnes et al. 2017). Just like other clustering tools, HDBSCAN outlines clusters by defining their cores through the number of neighbours within a radius ϵ. Its main advantage over equivalent algorithms, including DBSCAN for which it is an extension, is its flexibility. HDBSCAN indeed allows the user to identify clusters at different density thresholds, thereby finding more reliable groups by removing the bias introduced by a fixed value of ϵ.

Several parameters affect the results of HDBSCAN, but the main ones can be listed as follows:

- `cluster_selection_method` corresponds to the selection method, and there are two of them, namely `Excess of mass (EOM)` and `Leaf`. The first one has the advantage of decreasing the noise by identifying larger clusters (see e.g. Kerr et al. 2021). By contrast, the second one will select more homogeneous and smaller structures. The latter method has already been applied for OB associations (see e.g. Santos-Silva et al. 2021), and since it is more appropriate, I have chosen the `Leaf` method for my clustering analysis.

- `min_cluster_size` sets the minimum size for a cluster to be identified, i.e. a cluster is only selected if it has more members than `min_cluster_size`.
- `min_samples` sets the number of samples within a neighborhood for a point to be considered like a core point. Contrary to `min_cluster_size`, which only affects the number of identified clusters, `min_samples` changes their membership: consequently, it is the most important parameter for the clustering analysis.

For this analysis I use the Cartesian coordinates X, Y and Z, alongside the transverse velocities V_l and V_b, including the SED-fitted distance for both the positions and the velocities. This translates into a 5D clustering analysis, with three parameters in units of pc and two parameters in units of km s^{-1}.

The positions and velocities are normalised as follows:

$$X_{\text{norm}} = \frac{X - min(X)}{max(X) - min(X)} \tag{5.1}$$

$$Y_{\text{norm}} = \frac{Y - min(Y)}{max(X) - min(X)} \tag{5.2}$$

$$Z_{\text{norm}} = \frac{Z - min(Z)}{max(X) - min(X)} \tag{5.3}$$

$$Vl_{\text{norm}} = \frac{V_l - min(V_l)}{max(V_l) - min(V_l)} \tag{5.4}$$

$$Vb_{\text{norm}} = \frac{V_b - min(V_b)}{max(V_l) - min(V_l)} \tag{5.5}$$

Each parameter of the same units is normalised with respect to the parameter with the largest extent sharing this unit, i.e. X, Y and Z were normalised with respect to X (in this instance) in order to overcome the stretching along the line-of-sight, while

V_b was normalised with respect to V_l. As a result, all the normalised parameters have values between 0 and 1, and can be compared directly since they share the same units.

To be consistent with the definition of OB associations (Humphreys, 1978; Wright, 2020), I set `min_cluster_size` to 15, and use a value of `min_samples` of 10, which after some testing, produced groups with properties similar to OB associations. I applied HDBSCAN with these parameters on the 5617 SED-fitted OB stars with $\log(T_{\rm eff}/K) > 4$ or $\log(L/L_\odot) > 2.5$. Again, similarly to Chapter 3, these thresholds include massive stars that became cooler when they evolve. This process produced 14 initial groups, listed in Table 5.2.

However unbiased HDBSCAN might look, this clustering tool is likely to detect unreal groups, because its overconfidence will lead it to consider random dense fluctuations in a sample to be part of a group (Hunt & Reffert, 2023). Hence the next step consisted of assessing the probability of existence for every group. To that end, I applied a method based on the approach from Santos-Silva et al. (2021) and applied a bootstrapping process. I randomly varied the PMs and SED-fitted distance of each star within their uncertainties, and applied HDBSCAN again. For each iteration, a new set of associations was produced, which I then compared to the original associations. When a 'bootstrapped' association had 5D parameters within 1σ from the median of one of the original associations, I marked it as corresponding to the same association. Whenever such a match occurred, I also compared the individual members of the bootstrapped associations with the original associations. I repeated this process 10,000 times, computing the fraction of iterations in which a given association appears, together with a fraction of these iterations in which a given star appears in each association. These fractions were then respectively taken as the probability that a given association is genuine and a membership probability for each star in each association. In these associations are also included stars that were not part of the original associations, but appeared in more than 50 % of iterations in the bootstrapped associations.

Next it was necessary to select a probability threshold above which an association could be considered reliable. I performed a Monte Carlo simulation in order to estimate the number and properties of OB associations that would be identified from a random distribution of stars. I randomly sampled 100 times the Galactic coordinates, PMs and SED-fitted distances of the 5617 OB stars. HDBSCAN was run for each iteration, identifying new associations in the process, and for every iteration I applied a bootstrapping process (with 1000 iterations) to calculate their probabilities. These simulations resulted in a total of 1154 'randomized' associations, i.e., an average of ∼12 per simulation. Typically, the probability of existence for these associations is very low. Only 188 have probabilities >50%, 77 have probabilities >80% and 46 have probabilities >90%, equivalent to ∼2, ∼1 and <1, on average, per simulation. In the real data, I identified 9 groups with probabilities >50%, 7 with probabilities >80% and 4 with probabilities >90%. This comparison implies that the 4 associations with probabilities >90% are likely to all be real (especially since their probabilities are all >99%), whereas the 5 associations with probabilities of 50–90% may include 2 contaminants.

From these simulations, it was also clear that false-positive associations (>80%) are almost entirely found nearby ($d \lesssim 1.5$ kpc). Therefore, I choose to discard all the nearby (<1.5 kpc) OB associations with a probability <90 %, keeping only the very high probability groups (now named associations 1-4) and the very distant group with a moderately-high probability (association 5). There are four associations whose existence is moderately probable, but without being certain these have been discarded. Further data (e.g. RVs), more precise astrometry or expanded membership amongst lower-mass stars would be required to confirm their existence.

This clustering analysis has thus produced 5 new high-confidence, spatially- and kinematically-coherent OB associations. They are displayed in Galactic coordinates in Fig. 5.6, in Galactic transverse velocity in Fig. 5.7 and in Galactic longitude as a function of distance in Fig. 5.8. These new OB associations are distributed over a range of distances from 1 kpc to almost 3 kpc, with many super-imposed on each other on the plane of the sky. *Gaia* data was therefore crucial to find these OB associations.

Table 5.2: Properties of the new OB associations in Auriga. N corresponds to the initial number of stars in the association (before any bootstrapping is applied), while N_{g} is the number of likely members (with a membership probability of 50 % or more, after bootstrapping) and N_{tot} is the total number of stars in the associations (including those appearing during the bootstrapping with a probability of 50 % or more). d_m is the median distance of the group. Probability gives the probability that the association is real.

Association	N	N_{g}	N_{tot}	d_m (pc)	Probability(%)
	26	20	21	738	86.58
	18	18	25	906	83.20
1	198	186	215	1056	99.98
2	41	41	43	1085	99.99
	17	16	16	1475	57.56
3	99	89	119	1514	99.93
4	130	127	138	1923	99.87
	15	13	13	1956	4.15
	37	19	19	2188	48.99
	23	11	11	2508	9.35
	21	9	9	2677	18.24
5	90	39	39	2760	82.31
	83	13	13	2803	63.10
	15	7	7	2951	8.35

5.2.5 Comparison with historical associations and open clusters

Beyond the clustering analysis, one way to verify the reality of the newly-identified OB associations is to connect them with existing structures. I thus crossmatched their members with the Aur OB1 and OB2 members listed in Melnik & Dambis (2020), as well as the open cluster members from Cantat-Gaudin & Anders (2020). The results of the crossmatch are shown in Table 5.3.

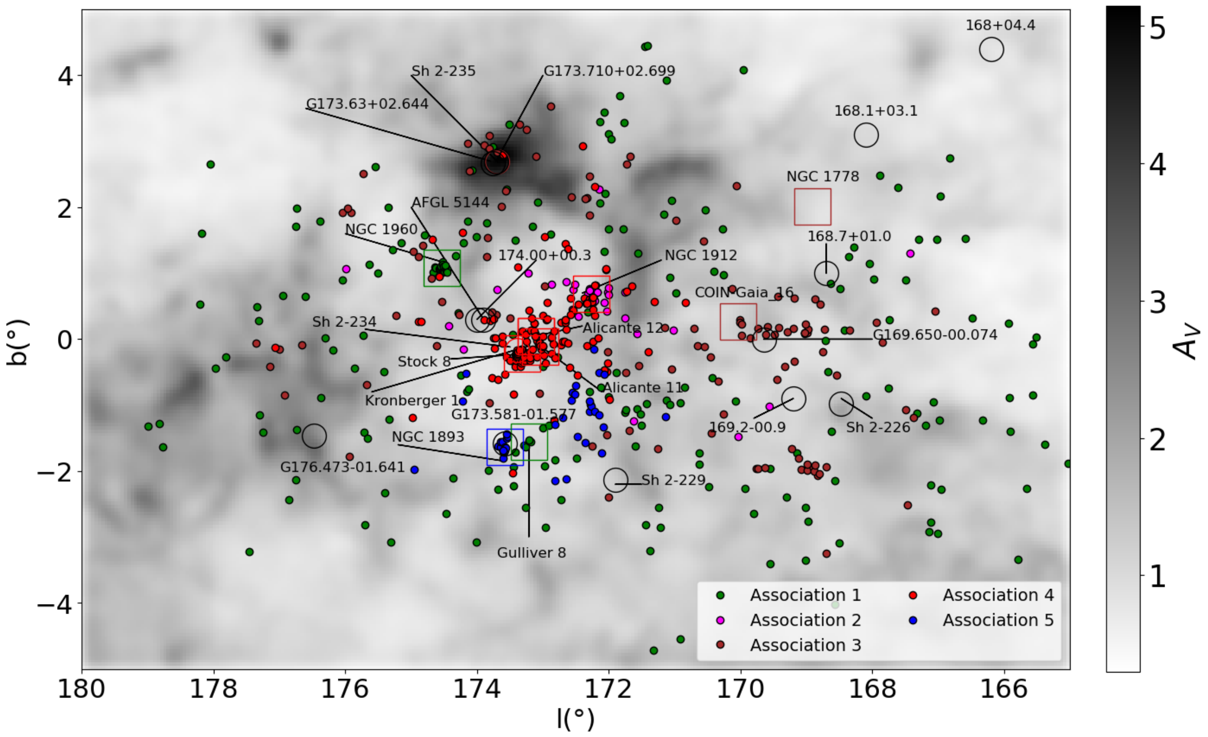


Figure 5.6: Galactic coordinates of the members of my 5 new OB associations in Auriga. Both the background extinction map and the features highlighted are the same as in Fig. 5.1.

Table 5.3: Comparison between the new OB association members and both the historical associations members from Melnik & Dambis (2020) and the open clusters from Cantat-Gaudin & Anders (2020). N_{hist} is the number of stars in the historical association and N_{OC} corresponds to the number of stars in the open cluster.

Assoc.	N_{hist}	Hist. assoc.	N_{OC}	OC
1	7	Aur OB1	25, 4	NGC 1960, Gulliver 8
2			26	NGC 1912
3	1	Aur OB1	9, 4	NGC 1778, COIN-Gaia_16
4	1, 8	Aur OB1, OB2	49, 5	Stock 8, Kronberger 1
5	2	Aur OB2	15	NGC 1893

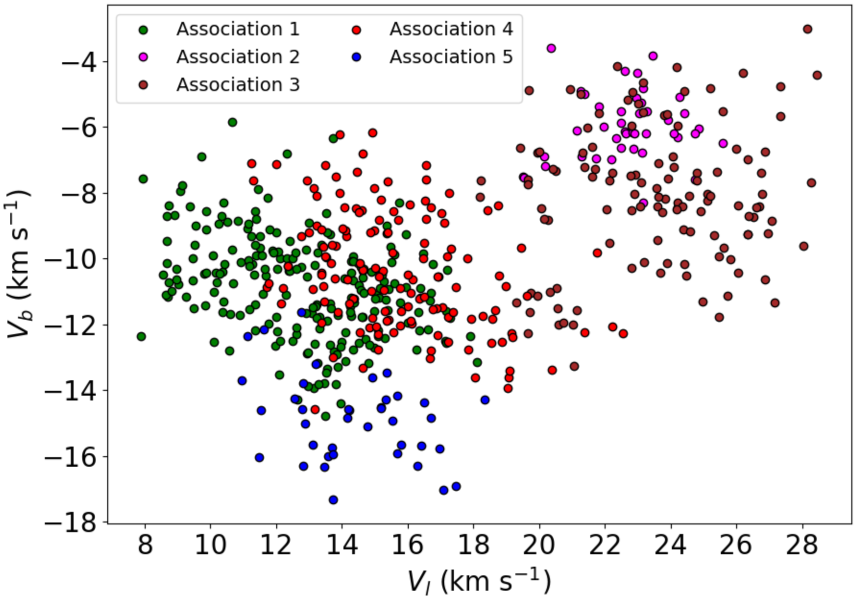


Figure 5.7: Galactic transverse velocities of the 5 new OB associations in Auriga.

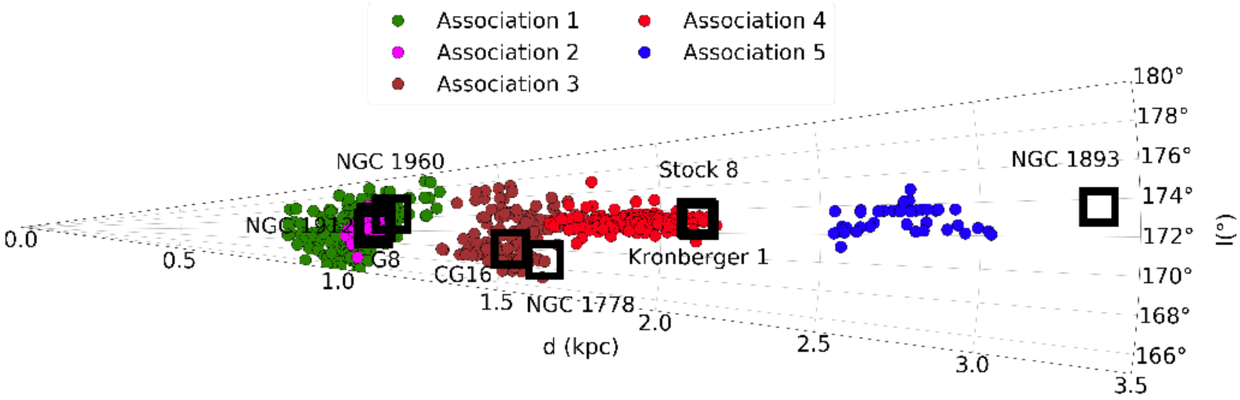


Figure 5.8: Galactic longitude as a function of SED-fitted distance for the 5 new OB associations in Auriga. The median uncertainty on distance for stars in each association are respectively equal to ∼30 pc for associations 1 and 2, ∼50 pc for association 3, ∼70 pc for association 4 and ∼160 pc for association 5. Related OCs have been included, with their position and distance taken from Table 5.1. Gulliver 8 and COIN-Gaia_16 have respectively been abbreviated to G8 and CG16 for readability.

Association 1 contains stars from Aur OB1 and unsurprisingly from NGC 1960, given the relation between the historical association and the open cluster. A few other members of Aur OB1 are spread over the other new OB associations. Members from Aur OB2 are shared between associations 4 and 5, which also have significant overlaps with Stock 8 and NGC 1893. This comparison suggests that NGC 1893 is located closer than previous estimations (Lim et al., 2018; Cantat-Gaudin & Anders, 2020) at a distance of $\sim$2.8 kpc, consistent with the distance from Melnik et al. (2020).

5.3 Analysis of the new OB associations

In this section, I conduct a physical and kinematic analysis of the new OB associations in Auriga, including HR diagrams. before studying their expansion and star formation history.

5.3.1 Physical properties of the individual associations

The first step in characterizing the OB associations consists of estimating their number of O- and B-type stars. A B-type star is defined as having an SED-fitted $\log(T_{\text{eff}}/K) > 4$ and $\log(T_{\text{eff}}/K) < 4.3$ and O-type stars as having a SED-fitted $\log(T_{\text{eff}})/K > 4.3$.[2] These thresholds are identical to Section 5.2.2. I estimated the uncertainties on the resulting numbers through a Monte Carlo experiment: for every iteration, I randomly sampled the effective temperatures of stars within their uncertainties. The observed number of B- and O-type stars for each association is shown in Table 5.4. This unveils a picture of a dominance of B-type over O-type stars, also illustrated in the HR diagrams in Fig. 5.9.

The second step was to calculate the total stellar initial mass of each association, with a method similar to Chapter 3. I identified the range of masses over which

---

[2] A more conservative threshold on T_{eff} was chosen in Chapter 3, but here I favour the threshold from Martins et al. (2005).

Table 5.4: Properties of the new OB associations. The parameters are listed in the first column, with the subscripts 'm' and 'σ' respectively used for the median and dispersion. The total initial stellar mass is corrected for observational incompleteness, as described in the text.

Parameters	Units	Assoc. 1	Assoc. 2	Assoc. 3	Assoc. 4	Assoc. 5
RA(ICRS)$_m$	deg	81.30	82.13	80.16	82.02	80.70
DE(ICRS)$_m$	deg	34.97	35.84	36.57	34.77	33.94
l_m	deg	170.72	172.24	170.70	173.09	172.82
b_m	deg	-0.16	0.70	0.11	-0.03	-1.48
d_m	pc	1056	1085	1514	1923	2760
σ_d	pc	102.2	25.2	76.0	103.8	-
V_{l_m}	km s^{-1}	12.98	22.85	23.69	15.37	14.18
σ_{V_l}	km s^{-1}	2.52	1.09	2.96	2.20	2.10
V_{b_m}	km s^{-1}	-10.75	-6.16	-8.10	-10.58	-14.84
σ_{V_b}	km s^{-1}	1.41	0.95	2.00	2.05	1.17
Observed number of B stars		194 ± 3	40^{+2}_{-1}	107^{+3}_{-2}	115^{+3}_{-4}	32 ± 2
Observed number of O stars		12^{+3}_{-2}	0 ± 0	4 ± 1	13 ± 2	3^{+2}_{-1}
Total stellar initial mass	$M_\odot$	6051^{+426}_{-387}	1219^{+182}_{-167}	3075^{+298}_{-276}	3500^{+315}_{-306}	879^{+163}_{-136}
HR diagrams age	Myr	0-30	-	0-20	0-5	0-10
Related OCs age	Myr	18-26	250-375	26	4-8	1-5
Traceback age	Myr	$20.9^{+1.1}_{-1.2}$	$369.9^{+8.3}_{-22.2}$	$11.7^{+7.2}_{-3.0}$	$1.6^{+1.3}_{-0.9}$	-
Age	Myr	~ 20	-	10-20	0-5	0-10

my sample completeness is expected to be unbiased by the age of my stars. The corresponding mass range corresponds to 2.5 $M_\odot$ (the mass of an A0 star) for the lower threshold and 7.1 $M_\odot$ (the post main-sequence turn-off mass at an age of 50 Myr Ekström et al. 2012) for the upper threshold. I then corrected the number of stars following the incompleteness levels computed in Section 5.5 and shown in Fig. 5.5.

I performed a Monte Carlo simulation where I randomly sampled stellar masses, using the mass function from Maschberger (2013). During the simulation I counted both the number of stars in the selected mass range and the total number and mass of stars. Only when the total number of observed stars in the selected mass range was reached was the simulation finished. I repeated this process 10,000 times, using the uncertainties on the individual SED-fitted stellar masses. This allowed me to derive an uncertainty for the total stellar mass of each association, as listed in Table 5.4. These masses range from ~900 to ~6000 $M_\odot$. The most massive is association 1, with an estimated initial stellar mass of ~6000 $M_\odot$ and currently containing about 200 B-type members. It has to be noted that the applied version of the SED fitting process (version 1.2) assumes all stars to be single, which will underestimate the total masses of the OB associations by ~25% (see Chapter 7).

5.3.2 Kinematic properties of the individual associations

I also calculated kinematic properties of the OB associations. In Table 5.4 are listed their median equatorial and galactic coordinates, distances and transverse velocities. Again I applied the method from Ivezić et al. (2014) to estimate the intrinsic dispersion in velocity and distance using the observational uncertainties.

Velocity dispersions typically range in the interval 1-3 km s^{-1}, while the distance dispersions are equal to a few tens of pc, with a maximum at $\sim$ 100 pc for associations 1 and 4.[3] These results are consistent with the properties of OB associations (Wright, 2020).

5.3.3 HR diagrams of association members

I produced HR diagrams for the OB association members, using the SED-fitted effective temperatures and luminosities, and they are displayed in Fig. 5.9. Consistent with Table 5.4, this plot unveils a picture dominated by late B-type stars, which makes a

---

[3] I could not estimate σ_d for association 5, because at its distance the observational uncertainties are greater than its extent.

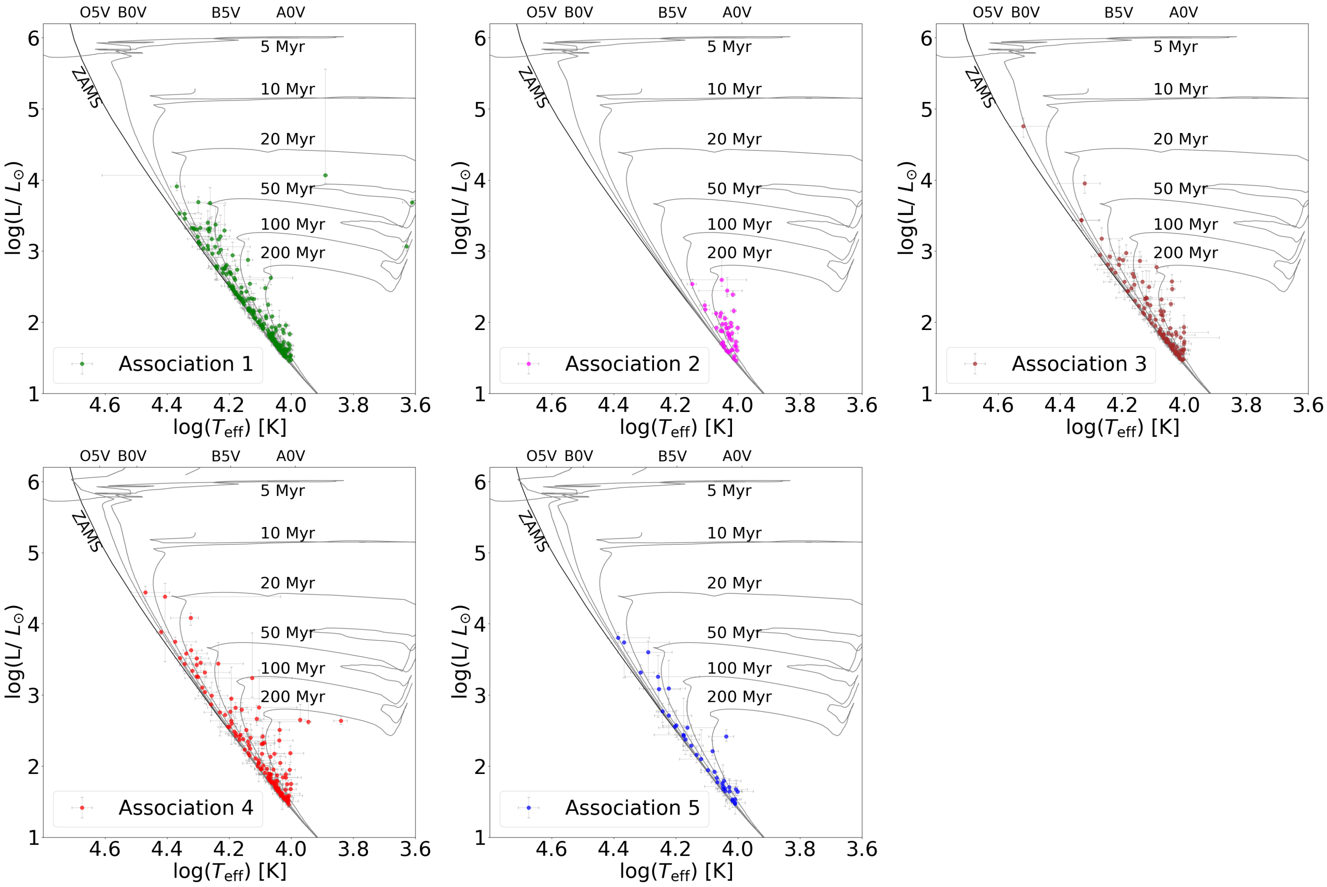


Figure 5.9: HR diagrams for the new associations in Auriga. I have included the isochrones from the rotating evolutionary models of Ekström et al. (2012), alongside the positions of some spectral type on the top horizontal axis.

direct age assignment more difficult, as most OB stars are still very close to the main sequence at the typical age of an OB association. This is especially true for association 2, and to a lesser extent for association 1, which contains a few stars close to the 50, 100 and 200 Myr isochrones (suggesting it would be much older than other known OB associations, though these may be contaminants). Associations 3, 4 and 5 are probably younger, < 20 Myr, due to the presence of hotter stars.

Given the presence of contaminants, it is worth discussing which factors affect the positions of stars within a HR diagram. Extinction has an influence on the position of stars, and it is derived for every individual star from the SED fitter, through an interpolation of the sightline from the extinction map. Henceforth, the uncertainty on extinction stems directly from the uncertainty in distance. Uncertainty in extinction then propagates through the uncertainty in effective temperature and luminosity, resulting in the displayed HR diagram.

Instead of individual reddening, one could consider using a single extinction value for every OB association member. The effect of such choice would be minor, because the variation in extinction across a single association is small, typically with a standard deviation in A_V of 0.2 to 0.6 mag. While this causes minor changes to the position of stars, the extreme outliers still remain after fitting the stars with the median extinction of their respective association.

The position of stars in the HR diagrams is more affected by binarity (that I will explore in Chapter 7), together with the presence of possible contaminants. An illustration of this is association 1, for which the majority of stars lie close to the ZAMS, consistent with being under 20 Myr old. This would imply that the stars in the association sitting on the 50, 100 and 200 Myr isochrones may be contaminants. Another explanation is that these stars are not members of the association and are contaminants. However, an inspection revealed that there is little to no correlation between member probability (derived from the clustering analysis) and position on the HR diagram.

5.3.4 Expansion and traceback age

In Chapter 3, I estimated the kinematic age of the new OB associations in Cygnus using a linear fit between l and μ_l. This approach was feasible because of their youth (typically below 10 Myr), but for the OB associations in Auriga, both Table 5.4 and Fig. 5.9 suggest typical ages of a few tens of Myr for most associations.

Instead I used the same method as in Chapter 4 and performed a 3D kinematic traceback on these OB associations. The ingredients of this method are composed again of the epicycle approximation from Fuchs et al. (2006), corrected for the local standard

Table 5.5: Average RV for each OB association in Auriga, with N_{RV} being the number of stars with a reliable measured RV. The references numbers correspond to, respectively: (1): APOGEE, (2): Fehrenbach et al. (1992), (3): Grenier et al. (1999), (4): Gontcharov (2006), (5): Turner et al. (2011), (6): Chojnowski et al. (2017), (7): Gaia Collaboration et al. (2018), (8): Zhong et al. (2020)

Association	$N_{\rm RV}$	RV (km s^{-1})	References
1	47	2.13 ± 1.26	All but (6)
2	5	6.59 ± 4.31	(8)
3	8	-4.64 ± 3.01	(1), (6), (8)
4	31	4.12 ± 1.63	(1), (7), (8)
5	4	1.46 ± 7.45	(8)

of rest from Schönrich et al. (2010), together with a combination of the observed *Gaia* PMs and literature RVs.

The sample of RVs comes from both APOGEE and SIMBAD, removing stars with unreliable measurement or without any indicated uncertainty. I took the weighted mean of the two RV values for stars in both these samples. This resulted in a list of 95 stars with reliable RVs.

I computed the median RV for each association, with the aim of tracking back whole associations rather than individual stars because of the influence of the unresolved close binaries. To that end, a Monte Carlo simulation was again applied to derive the uncertainties on RVs (see Chapter 4 for a detailed description of the method). The results are shown in Table 5.5.

With these RVs obtained, I had to choose the upper limit for the 3D kinematic traceback, constrained through the ages of the related open clusters and from the HR diagrams of the associations (Fig. 5.9). These limits were set, respectively, at 50 Myr in the past for associations 1 and 5, 400 Myr for association 2, 30 Myr for association 3 and 20 Myr for association 4.

I performed the traceback for each association using a time step of 0.1 Myr. At every iteration, I calculated the on-sky median absolute deviation (MAD) of the association members in Galactic coordinates[4]. The traceback age of the association then corresponded to the time where its MAD was at its minimum, i.e. the most compact configuration of the association. I repeated this process 1000 times in order to estimate uncertainties on these ages. The results are shown in Fig. 5.10.

Table 5.4 lists the OB association ages for the different methods. The final, assigned age was based on the combination of (i) the time in the past when the system reached its most compact state from the 3D kinematic traceback, (ii) the age of any open cluster or star-forming region linked to the association (Section 5.2.5), and (iii) any age constraints arising from the HR diagram (Fig. 5.9). The traceback method was sometimes able to place reasonable constraints on the age of the system, (e.g., for associations 1 and 3), but for other associations the best constraint came from the open clusters and star-forming regions the association was connected to (e.g., for associations 4 and 5), or the HR diagram (e.g., for association 3). There is very little constraint on the age of association 2.

5.4 Discussion

This section is dedicated to a discussion of the findings of this chapter. I will notably discuss how the new Auriga OB associations provide insights into the recent star formation and Galactic structure in this region. The main results of this chapter are summarized below:

- A modern picture of Aur OB1 and OB2, using parallaxes and proper motions from *Gaia* EDR3, shows that their members lacks the kinematic coherence that would make them genuine OB associations. A new exploration of the OB associations in the region is therefore required.

---

[4] A 2D MAD was decided rather than a 3D MAD due to the large error bars on RVs causing line-of-sight distance uncertainties as the traceback goes further into the past.

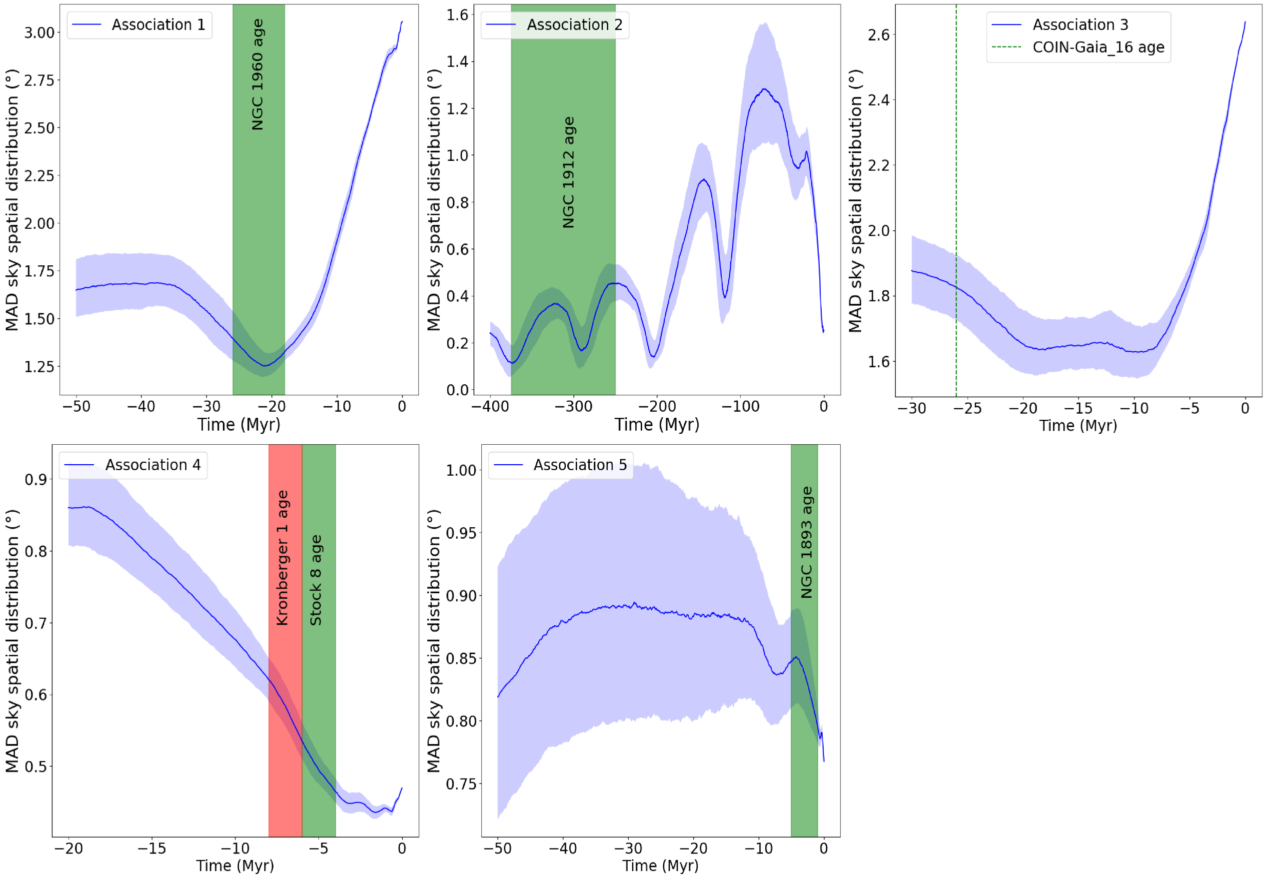


Figure 5.10: 2D MAD (on-sky in Galactic coordinates) distribution of members of each association as a function of traceback time. The ages of the related open clusters were taken from Table 5.1.

- Applying version 1.2 of the SED fitter, I reliably identified more than 5000 OB stars across the Auriga region.
- I identified 5 high-confidence new OB associations in Auriga, which I studied both physically and kinematically. I estimated their age through a combination of 3D kinematic traceback, connection to open clusters and star-forming regions with a known age, and the position of their members in the HR diagrams. ∼90 % of the SED-fitted OB stars ended up unclustered, and I will comment on that in this section.

5.4.1 The new Auriga OB associations

The 5 identified OB associations are highly reliable. Their total stellar masses range from ∼1000 to ∼6000 $M_\odot$, with their velocity dispersions ranging from ∼1 to ∼3 km s^{-1} in both directions, consistent with the definition of OB associations (Wright, 2020). This view is reinforced by their likely connection with the open clusters in this area (Table 5.3).

Due to its common members with Aur OB1, and its relation with NGC 1960, association 1 can be considered as the replacement of the historical Aur OB1 association, and becomes the main foreground OB association in this region. Similarly, the members of the historical Aur OB2 are now shared between associations 4 and 5, themselves related to Stock 8 and NGC 1893. Such a result confirms the suggestion of dividing Aur OB2 into a foreground and background association (Marco & Negueruela, 2016), located respectively at distances of 1.92 ± 0.10 kpc and ∼2.8 kpc (Table 5.4).

The HII region Sh 2-235 is located at a distance of 1.36 ± 0.27 kpc (Foster & Brunt, 2015). The median distance of association 3 is equal to 1.51 ± 0.08 kpc (Table 5.4), thus this would suggest that they are related. Furthermore, association 3 contains HD 36483: this O9.5IV star (Sota et al., 2011) may be responsible for ionizing the HII region.

Sh 2-234 could be likewise related to association 4, as it occupies the centre of the region of study alongside the three related OCs Stock 8, Alicante 11 and 12 (Fig. 5.6). Its distance of 2.19 ± 0.10 kpc (Foster & Brunt, 2015), close to the median distance of association 4 at 1.92 kpc, reinforces this picture. Such a relation has already been suggested by Marco & Negueruela (2016), with the star LS V +34 23 ionizing the HII region. Nevertheless, this star failed the quality checks listed in Section 5.2.1, so I cannot comment whether it belongs to association 4. There are however three O-type stars in association 4, i.e. LS V +34 15, LS V +34 21 and LS V +35 25, whose respective spectral types are O5.5V (Negueruela et al., 2007), O9IV (Roman-Lopes & Roman-Lopes, 2019) and O9.5V (Georgelin et al., 1973). Each of them could contribute to the ionization of Sh 2-234. Unfortunately, there is no kinematic relation between association 4 and Sh 2-234, since I estimated the RV of the former to be 4.12 ± 1.63 km s^{-1} whilst the latter has a RV of -21.4 ± 0.2 km s^{-1} (Anderson et al., 2015). These RVs are too different for the association and the HII region to be related beyond the ionizing O-type stars.

5.4.2 Expansion and age of the associations

Analysing the new Auriga OB associations unveiled a complex picture, where their estimated ages range from very young ($<$ 10 Myr for associations 4 to 5) to older associations (e.g. association 3). Generally, the different age indicators for these OB associations were consistent with each other (Table 5.4). Association 2 is the exception with most of its stars close to the ZAMS in the HR diagram (Fig. 5.9), but its related OC, NGC 1912, has been estimated to be several hundreds of Myr old (Table 5.1), which is much older than most OB associations (Wright, 2020), possibly suggesting the OC is not related to the OB association..

Most of these OB associations, as shown in Fig. 5.10, were found to be in a more compact state in the past, although associations 4 and 5 reached this state very recently. This is a signature of expansion (Wright & Mamajek, 2018; Miret-Roig et al., 2022), and would be a first step to support the *clustered* model of star formation from Lada & Lada (2003). A second step would be to calculate the extent of the associations at their most compact, to see if they are consistent with being clusters, however the lack of precision in the RVs prevents me from determining an accurate linear size for these groups.

5.4.3 OB stars unassigned to groups

Out of the original 5617 SED-fitted OB stars, only 554 OB stars were found to be clustered among the 5 newly identified OB associations. This means that ∼90 % of the OB stars inside this area remain unclustered, in spite of the commonly-held view that most stars form in clusters (Lada & Lada, 2003). Such an observation could be explained by several causes, that I will outline below.

Firstly, in Section 5.2.4, I imposed a minimum number of 15 OB stars per association, following a general definition (Humphreys, 1978; Wright, 2020). In practice, this means that I am missing groups containing only a handful of OB stars, which would be dominated by stars with lower masses. Furthermore, I kept only 5 of the 14 originally identified groups, and also rejected stars initially belonging to the associations but with low probability membership (see Table 5.2). This is particularly true for most distant groups (> 2 kpc), as the parallaxes from *Gaia* EDR3 have a lower precision at such distances.

OB associations are also, by definition, young structures due to their unbound nature (up to several tens Myr old, Wright 2020). In this sample are contained many late B-type stars, and a B9 stars with an initial stellar mass of 2.75 $M_\odot$ will have a lifetime of ~700 Myr (Ekström et al., 2012). Thus, even if many late B-type stars were born clustered, their formation group would have dispersed into the Galactic field well before the end of the star's life.

Another explanation is that, while some of these stars were born clustered, they ended up ejected from their parent cluster or association, becoming runaways. Especially, since the majority of massive stars belong to multiple systems (Preibisch et al., 2001; Sana & Evans, 2011; Sana, 2017), it is possible that when the primary star finishes its life in a supernova explosion, the process pushes the secondary star beyond the group it formed in.

5.4.4 Distribution of OB associations and Galactic structure

Association 4 is the youngest identified OB association, located at a distance of ~2 kpc (Table 5.4). Its position coincides with the current position of the Perseus spiral arm from Reid et al. (2019). In Fig. 5.11 are displayed the positions of the new OB associations, relative to the Local Standard of Rest (LSR) as a function of time. This picture reveals a potential age gradient.

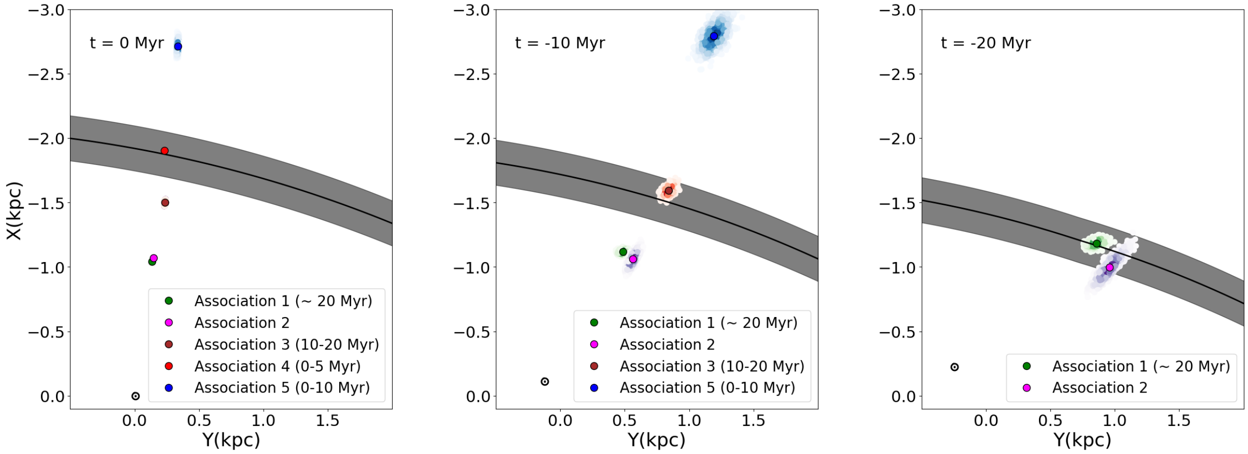


Figure 5.11: Median positions in Cartesian coordinates (in the X-Y plane) of the new OB associations from the present time to 20 Myr in the past, in steps of 10 Myr, and displayed relative to the Local Standard of Rest. The blur around the median positions represent the uncertainties from the traceback. The position of the Perseus spiral arm (with its thickness) is taken from Reid et al. (2019), and the position of the Sun is shown with a Solar symbol with respect to the LSR.

To verify whether this age gradient is related to the motion of the Perseus spiral arm, I modelled the positions of the spiral arm and our new OB associations over the last 20 Myrs. The position of of the Perseus spiral arm originates from Reid et al. (2019), and I also used the spiral arm pattern speed of $\Omega_p = -28.2 \pm 2.1$ km s^{-1} from Dias et al. (2019). I projected the position of the Perseus spiral arm using the following equation from Dias et al. (2019):

$$\theta(t) = \theta_0 - \Omega_p\, t \tag{5.6}$$

where θ_0 stands for the current angular position of the spiral arm and $\theta(t)$ its position at a time t. This position is then converted into Cartesian coordinates.

The position of the OB associations, compared with the Perseus spiral arm, are shown in Fig. 5.11 in steps of 10 Myr and back to 20 Myr in the past. At 10 Myr ago, association 3 (whose estimated age ranges at 10-20 Myr, see Table 5.4) is found within the Perseus spiral arm. At 20 Myr ago, using again Eq. (5.6), a similar observation can be made for association 1, which has an estimated age of 20 Myr. Association 2 crosses the spiral arm at about the same time, though its related OC and traceback suggest it to be older (Fig. 5.10). This could suggest that the association is actually younger and therefore unrelated to NGC 1912 or that the association did not form in a spiral arm. Finally, association 5 remains too distant from the Perseus spiral arm. It may have thus formed outside a spiral arm or in the Outer spiral arm, which lies beyond the Perseus spiral arm.

This shows that OB association are likely to have formed within spiral arms and therefore that not only are OB associations tracers for the current positions of spiral arms (e.g., Wright 2020), they can also act as probes of the recent star formation history of the region, and potentially the spiral arm progress within.

6 Kinematics of OB stars in Cassiopeia

This chapter is dedicated to the initial analysis of the Cassiopeia region. Here I first give a general picture of the area, with its historical OB associations, open clusters, HII and star-forming regions. I then present the census of SED-fitted OB stars across the region.

The distribution of the SED-fitted OB stars in Cassiopeia is shaped by a linear correlation between Galactic longitude and proper motion in Galactic longitude, similarly to that observed in Cygnus. This time, however, it could be partially explained by the effects of Galactic rotation, though some residuals remain once its contribution is subtracted. While it is tempting to attribute these residuals to a local episode of star formation that shaped a large-scale expansion pattern, its spread in distance and temperature would argue against such interpretation, in spite of attempts to isolate the stars responsible for the gradient.

A kinematic traceback on the OB stars, both in 2D and 3D, reveals that they were more compact in the past and that they reached their most compact ∼17 Myr ago, but identifying the individual OB associations will be required to better analyse the recent star formation history of the region.

The work on Cassiopeia is ongoing and therefore unfinished, so I present here only the first, relatively complete, part of this work.

6.1 The Cassiopeia region

In this section I provide a modern picture of the Cassiopeia region. This includes a brief description of the historical Cassiopeia OB associations, where I explain why they need to be revisited, together with the open clusters and star-forming regions in the area.

6.1.1 Historical OB associations

Ten OB associations have been identified in Cassiopeia (Roberts, 1972; Humphreys, 1978). I introduced them in Chapter 1, and focused on the seven OB associations Cas OB1, Cas OB2, Cas OB4, Cas OB5, Cas OB7, Cas OB8 and Cas OB14. I chose them because they are all located relatively close together on the plane of the sky, within a 224 deg^2 area that I will be studying, limited by l = [105°,133°] and b = [-4°, 4°]. As recently as in Melnik & Dambis (2020), their historical members have been used to analyse the kinematics of OB associations with *Gaia* DR2 data.

Nevertheless, as I showed in Chapters 3 and 5 using *Gaia* EDR3 data, both the historical Cygnus and Auriga OB association members lack the kinematic coherence expected for such structures, which motivated me to identify new OB associations from the ground-up. Even if some historical OB associations showed some level of kinematic coherence (e.g. Cyg OB2), there are still some foreground and background contaminants that justify this.

Based on these observations, it is clear that the historical Cassiopeia OB associations likewise suffer from such issues, and therefore also need to be revisited. The method and techniques used are similar to those employed in Chapters 3 and 5.

6.1.2 Open clusters and star-forming regions

Like in Chapter 5, revisiting the census of OB associations in the area benefited from connecting them to existing structures such as open clusters and star forming regions, and thus to gather the current knowledge of these systems.

Cantat-Gaudin & Anders (2020) list more than one hundred OCs in the area studied within a distance of 4 kpc. Ten of those are thought to be related to the historical Cassiopeia OB associations, as mentioned in Chapter 1. I have listed some of their properties in Table 6.1.

This area also contains a wealth of star-forming and HII regions. The catalogue from Foster & Brunt (2015) lists 33 HII regions within 4 kpc in Cassiopeia, while

Table 6.1: Properties of the open clusters in Cassiopeia thought to be related to the OB associations. Galactic coordinates and distances have been taken from Cantat-Gaudin & Anders (2020), where the error bars correspond to the 16th and 84th percentiles on the most probable distance. The abbreviations for the age references are: P89 = Pandey et al. (1989), BH96 = Barbon & Hassan (1996), P01 = Pigulski et al. (2001), P04 = Piskunov et al. (2004), K05 = Kharchenko et al. (2005), P05 = Pandey et al. (2005), H08 = Hancock et al. (2008), H10 = Huang et al. (2010), D12 = Davidge (2012), K13 = Kharchenko et al. (2013), B19 = Bossini et al. (2019), D21 = Dias et al. (2021)

Open cluster	OB association	$l\,(^\circ)$	$b\,(^\circ)$	d (kpc)	Age (Myr)	Age references
UBC 84	Cas OB1	124.149	-1.127	2.50 ± 0.02	-	-
NGC 7510	Cas OB2	110.908	0.072	3.18 ± 0.02	6-50	BH96, P04, K05, K13, D21
NGC 103	Cas OB4	119.795	-1.386	$3.16^{+0.03}_{-0.02}$	107-367	K13, D21
King 12	Cas OB5	116.128	-0.145	3.08 ± 0.03	10-14	P89, K05, H08, K13, D21
NGC 7788	Cas OB5	116.430	-0.784	3.04 ± 0.02	16-40	P89, K05, D12, K13, B19, D21
NGC 7790	Cas OB5	116.595	-1.008	3.36 ± 0.02	60-126	P89, K05, D12, K13, D21
NGC 581	Cas OB8	128.047	-1.796	$2.49^{+0.01}_{-0.02}$	22-29	H10, K13, D21
NGC 654	Cas OB8	129.086	-0.359	2.91 ± 0.01	14-20	K05, P05, K13, D21
NGC 659	Cas OB8	129.378	-1.528	$3.17^{+0.02}_{-0.03}$	22-58	K13, D21
NGC 663	Cas OB8	129.494	-0.958	2.87 ± 0.01	14-32	P01, K05, K13, D21

Mellinger (2008) includes 3 star-forming regions in this area. Both are shown in Fig. 6.1.

6.2 Identifying and characterizing OB stars

In this section, I explain how I selected the sample of candidate OB stars within Cassiopeia, highlighting the differences with the selection in Cygnus (Chapter 3) and Auriga (Chapter 5). I also comment on the general distribution of the identified OB stars.

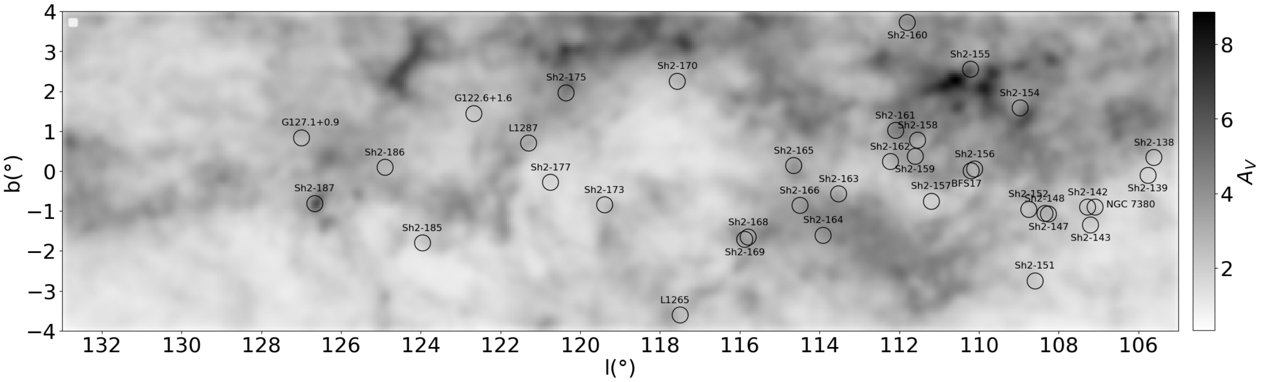


Figure 6.1: Star-forming and HII regions from Mellinger (2008) and Foster & Brunt (2015) shown in Galactic coordinates. The background extinction map, *Bayestar*, comes from Green et al. (2019) and has been scaled up by 22 %, following Section 2.8.

6.2.1 Data and selection process

I selected *Gaia* EDR3 sources (Gaia Collaboration et al., 2021) located within the Cassiopeia region, encompassing the same area as in Section 6.1.1, shown in Fig. 6.1. I discarded the sources with bad astrometric solutions, i.e., that do not fulfill the conditions $\frac{\varpi}{\sigma_\varpi} > 2$ and $RUWE < 1.4$, following the prescription from Lindegren et al. (2018). I downloaded from the *Gaia* archive the resulting list of 4,933,460 sources.

I further imposed a colour cut of BP-RP < 3 and a cut on the geometric distances, that were extracted from Bailer-Jones et al. (2021), requiring them to be less than 4 kpc, as the wide majority of members of the historical Cassiopeia OB associations (taken from Melnik & Dambis 2020) are located closer than this threshold. This reduced the list down to 4,267,258 objects.

This sample was crossmatched within 1" with IGAPS, 2MASS and UKIDSS. The same photometry checks and cuts were applied as for the data in Cygnus and Auriga. This time, however, the difference is that I only removed stars without valid photometry in a blue band (g, G or G_{BP}), trimming the catalogue to 4,256,246 objects. I divided it between the sources with at least one valid NIR band (3,253,106 sources) and without any valid NIR photometry (1,003,140 sources). This choice allows us to

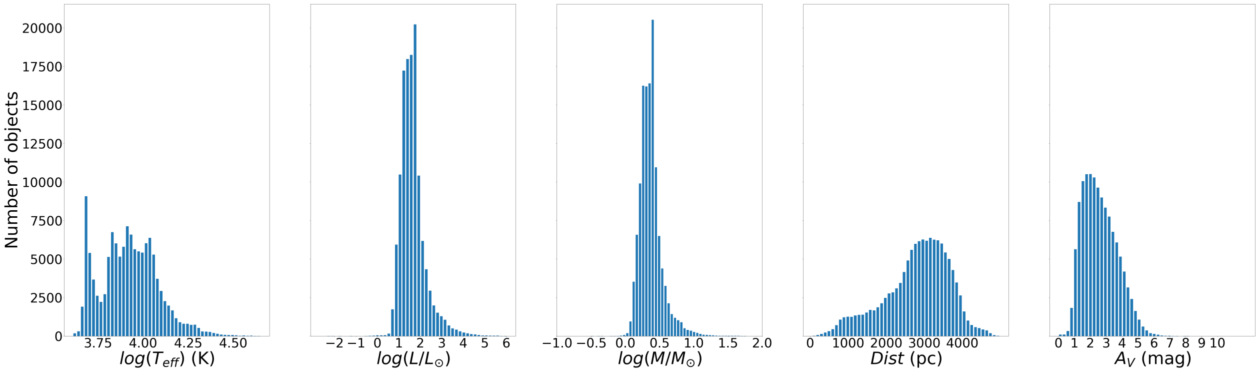


Figure 6.2: Median fitted parameters of the 124,712 stars in the working sample.

increase the completeness of the sample.

These two subsamples were finally filtered as follows:

- I applied an absolute NIR magnitude cut to the NIR sample, as well as another cut in the NIR colour-colour diagram, as described in Section 5.2.1 for the Cygnus sample. This reduced this subsample down to 120,336 sources.

- For the blue sample, I selected stars with $M_G < 1$, corresponding to an A0 star (Pecaut & Mamajek, 2013). M_G was calculated using the geometric distance from Bailer-Jones et al. (2021) along with the scaled extinctions from Green et al. (2019). I increased the extinction, A_G, by 25 % to reduce the risk of accidentally removing valid OB stars. This process reduced this sample to 4376 sources.

The final working sample of study for the Cassiopeia region, combining these two subsamples, contains 124,712 stars.

6.2.2 General results

I have applied version 1.2 of the SED fitting code to this working sample. The median physical parameters that I derived are displayed in Fig. 6.2.

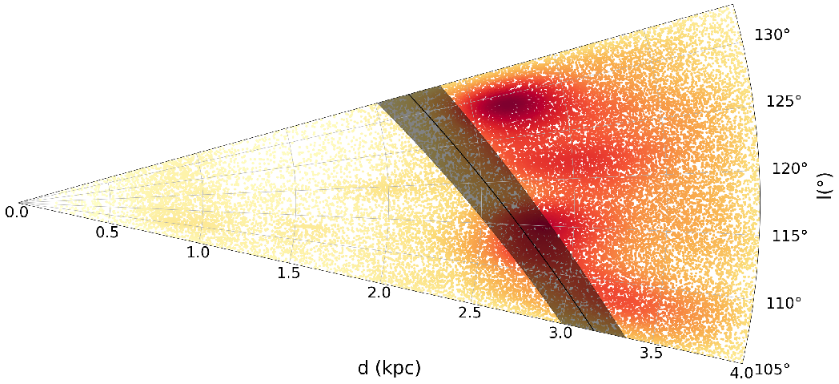


Figure 6.3: Galactic longitude as a function of SED-fitted distance for the 40,107 SED-fitted OB stars in Cassiopeia, colour-coded by density. The position and thickness of the Perseus spiral arm, displayed in black, have been taken from Reid et al. (2019).

The sample is dominated by early A-type and late B-type stars, with the median value of $\log(M/M_{\odot})$ equal to 0.36 (with a standard deviation of 0.16 dex) and the median value of $\log(L/L_{\odot})$ is 1.57 (with a standard deviation of 0.58 dex). The bulk of stars are located between 2.5 and 3.5 kpc, and below a reddening of 5 mag with a median value of ∼2.5 mag. This list also contains 39,149 stars fitted with $\log(T_{\text{eff}}/K) > 4$ (31.4 %, OB stars) and 2126 with $\log(T_{\text{eff}}/K) > 4.3$ (1.7 %, O-type stars).

For this work, I again define an OB star as having either $\log(T_{\text{eff}}) > 4$ or $\log(L/L_{\odot}) > 2.5$, in order to include stars at a later stage of their evolution that became cooler. 40,107 stars in the working sample fulfill these conditions. Their Galactic longitude, as a function of SED-fitted distance, is shown as a density plot in Fig. 6.3, compared with the position of the Perseus spiral arm from Reid et al. (2019).

Fig. 6.3 reveals that, while the Perseus spiral arm intercepts some of the overdensities of OB stars, other overdensities, and notably the biggest one at high Galactic longitudes, are located outside the estimated position of the spiral arm. It was previously noted by Zari et al. (2021) that the distribution of Galactic hot stars was more complex than thought before, as they did not all sit within the spiral arms, and the same conclusion can be drawn here. The position of the spiral arms provided by Reid et al. (2019) is also an approximation, fit to data that extends beyond Cassiopeia, so it is not unreasonable to observe slight differences in localised regions. Given the lifetime of late B-type stars that dominate the sample, it could be anticipated that these overdensities correspond to an older population of hot stars which formed within the spiral arm a few tens of Myr ago, before moving out of it. Identifying OB associations is vital to confirm this scenario which would be similar to what I observed in the Auriga region (Chapter 5).

6.2.3 Incompleteness

There is an inherent incompleteness in the sample of OB stars because of potential candidates getting discarded during the selection process described in Section 6.2.1. I calculated the level of incompleteness for each of these. In contrast with the sample in Chapter 5, the sample here has a higher completeness because stars without valid NIR photometry were not discarded.

A plot of the completeness as a function of G is shown for the SED-fitted OB stars (stars with $\log(T_{\rm eff}/K) > 4$ or $\log(L/L_{\odot}) > 2.5$) in Fig. 6.4. It is clear that a reasonable completeness was achieved in identifying OB stars, with the completeness above 70% at most magnitudes and above 80% at the peak of the distribution, i.e. between $G = 14$ and $G = 15$ mag.

The next step was to compare my distribution of OB stars with another Galactic catalogue of high-confidence OB stars. An ideal candidate for this purpose is the list from Zari et al. (2021). I crossmatched within 1 arcsec the 40,107 SED-fitted OB stars in Cassiopeia with the 46,025 OBA stars from Zari et al. (2021). Similarly to Chapter 5, I divide the results into three samples: the successful crossmatches (16,094 stars, list A), the unsuccessful crossmatches from my list (24,013 stars, list B) and the unsuccessful crossmatches from the list of Zari et al. (2021) (29,931 stars, list C).

21,879 stars from list B were too faint to be included in the catalogue from Zari et al. (2021) (i.e. they have $M_{Ks} > 0$ mag), while the remaining stars did not pass the photometric cuts based on their selection process. On the other hand, 18,384 stars

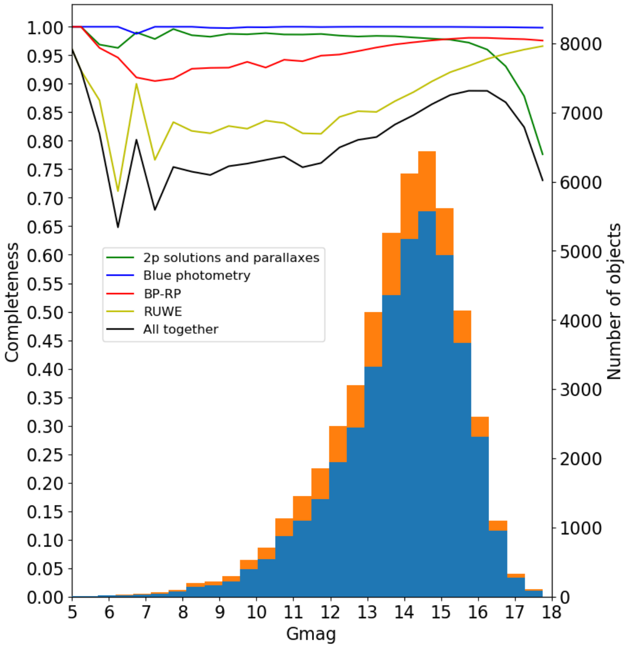


Figure 6.4: Completeness as a function of G magnitude for the 40,107 SED-fitted OB stars in Cassiopeia. The different completeness curves stem from the successive steps of the selection process, and their product is shown as a black curve. In blue and orange are shown respectively the number of sources before (blue) and after (orange) the application of the completeness correction.

from list C had a geometric distance from Bailer-Jones et al. (2021) greater than 4 kpc (the threshold used in Section 6.2.1), whereas 7086 stars were found in the SED-fitted sample (including 5228 with $T_{\rm eff} > 8000$ K) and the others did not pass the photometric and astrometric cuts.

Hence, again, the differences between the two lists can be attributed to the differences in the selection process. Compared with Chapter 5, the statistics are even similar, as ∼38 % of the SED-fitted OB stars in Auriga were found in the list from Zari et al. (2021), whilst it is the case for ∼40 % of the SED-fitted OB stars in Cassiopeia.

6.3 Kinematics of OB stars

In this section, I study the general kinematics of the identified OB stars in Cassiopeia, and especially focus on the gradient which deserves to be analysed in detail.

6.3.1 Large-scale kinematics

The distribution of SED-fitted OB stars in Cassiopeia reveals a pattern in their kinematics that is highlighted in the left panel of Fig. 6.5. This correlation between Galactic longitude and proper motion in Galactic longitude could constitute an indicator of large-scale expansion, as previously observed in Carina by Drew et al. (2021), and in Cygnus (Chapter 3 and Quintana & Wright 2021). The main gradient from Fig. 6.5 appears over l = [110°,130°], with a slope of ~0.17 mas yr^{-1} deg^{-1}. Another smaller gradient is evident at l = [105°,110°].

Using their sample of identified B-type stars in *Gaia* DR3, Gaia Collaboration et al. (2023b) fitted a Galactic rotation curve model to the l vs μ_l plot across all Galactic longitudes, with a sinusoidal shape. Gaia Collaboration et al. (2023b) used a Bayesian model and MCMC sampling to predict the proper motions from the Galactic rotation curve. To that end, they calculated the model velocities $\mathbf{v} = (-V_\phi \sin(\phi), V_\phi \cos(\phi), 0)$, where the azimuthal component of the velocity in Galactocentric cylindrical coordinates, V_ϕ, is defined as:

$$V_\phi = -(V_{\text{circ},\odot} + dV_{\text{circ}}/dR \times (R - R_\odot)) \tag{6.1}$$

where $V_{\text{circ},\odot}$ stands for the circular velocity at the Sun's position, dV_{circ}/dR corresponds to the Sun's peculiar velocity (Gaia Collaboration et al., 2023b), and $R_\odot$ is the Galactocentric distance of the Sun, equal to 8.277 kpc from GRAVITY Collaboration et al. (2022).

They then computed $\mathbf{v}$ - $\mathbf{v}_\odot$ with $\mathbf{v}_\odot = (U_\odot, V_\odot + V_{circ,\odot}, W_\odot)$ and used this to predict the proper motions. From this method, they derived values for the model

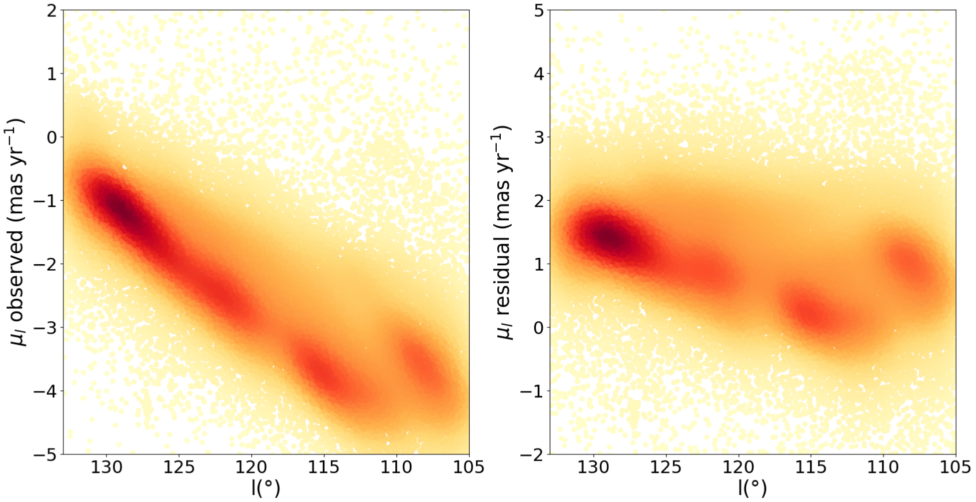


Figure 6.5: Proper motion in Galactic longitude as a function of Galactic longitude for the 40,107 SED-fitted OB stars in Cassiopeia. The left panel shows the observed μ_l and the right panel the residual μ_l after subtracting the predicted motion due to Galactic rotation, as described in Section 6.3.1.

parameters $V_{\mathrm{circ},\odot}$, $U_\odot$, $V_\odot$, $W_\odot$ and dV_{circ}/dR, that they found were consistent with the results from Bland-Hawthorn & Gerhard (2016).

I applied this model to my own list of OB stars in Cassiopeia (providing it with the individual Galactic coordinates and SED-fitted distances). This resulted in a list of predicted proper motions from the Galactic rotation curve, $\mu_{l,\mathrm{rot}}$. I subtracted this quantity from the observed μ_l, such that $\mu_{l,\mathrm{res}} = \mu_l - \mu_{l,\mathrm{rot}}$. The residuals are plotted on the right panel of Fig. 6.5.

The right panel of Fig. 6.5 reveals that a residual dependence still exists between $\mu_{l,\mathrm{res}}$ and l with the main slope having a gradient of $\sim$0.09 mas yr^{-1} deg^{-1}, hence about half that of the total gradient.

I performed a few tests, varying some of the aforementioned model parameters, to verify whether the entire kinematic pattern seen in Fig. 6.5 could be explained by Galactic rotation. Only an unrealistic combination of rotation curve parameters could

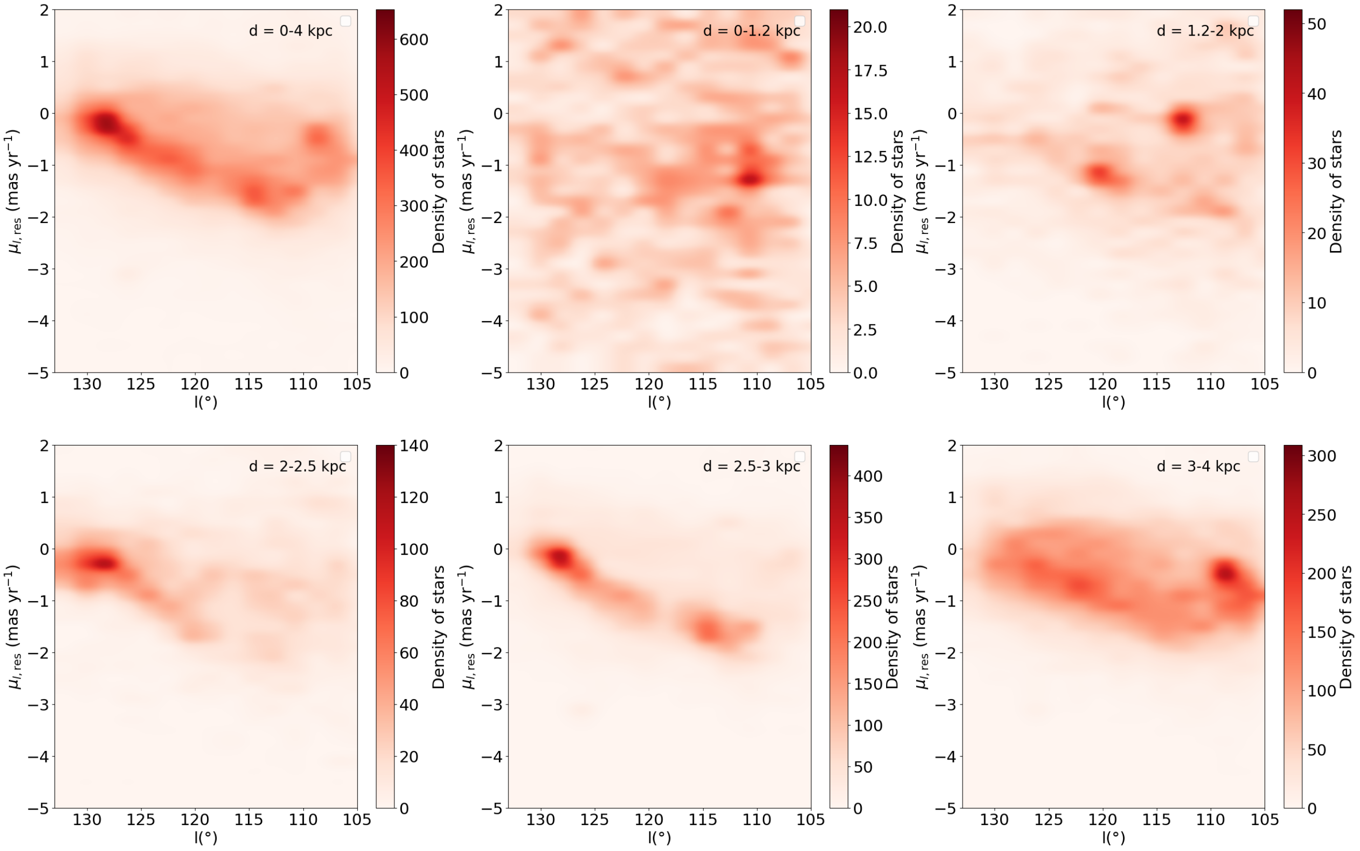


Figure 6.6: Density plot of the residual proper motions in Galactic longitude as a function Galactic longitude for the 40,107 SED-fitted OB stars in Cassiopeia. The top left panel displays the distribution for all distances while the other panels show the distribution at different distance subsets.

completely explain the observed kinematics.

These results show that there is a contribution from Galactic rotation to the gradient observed in Fig. 6.3.1, contrary to the Cygnus region where the same model predicted a flat distribution (see end of Chapter 3). Nevertheless, a gradient is still observed after subtracting this contribution, which requires deeper investigation.

6.3.2 The kinematic pattern as a function of distance

Fig. 6.5 shows a projected view of the kinematic pattern between $\mu_{l,\mathrm{res}}$ and l in Cassiopeia. To observe this distribution in more detail, I separated the OB stars into several distance subsets in Fig. 6.6, whose results are described below:

- Within 1.2 kpc, the distribution of μ_l is broader, consistent with the stars being closer. An overdensity of OB stars is visible at l ~110° and $\mu_{l,\mathrm{res}} \sim$ -1 mas yr^{-1}, that was hidden in the density of the more distant stars.
- Between 1.2 and 2 kpc, two other overdensities are highlighted. The first one is located roughly at l ~112° and $\mu_{l,\mathrm{res}} \sim$ 0 mas yr^{-1}, while the second one can be found at l ~120° and $\mu_{l,\mathrm{res}} \sim$ -1 mas yr^{-1}.
- Between 2 and 2.5 kpc, the main gradient starts to become apparent. A clear overdensity lies at l ~130° and $\mu_{l,\mathrm{res}} \sim$ -0.5 mas yr^{-1}. This roughly corresponds to the position of Cas OB8, as its assigned unique distance was found to be equal to 2.3 kpc (see Section 1.5.3).
- Between 2.5 and 3 kpc, the same overdensity at the position of Cas OB8 is still clearly marked. Most of the OB stars in this distance subset sit along the gradient. This is also where the overdensity of OB stars intercepts the Perseus spiral arm (Fig. 6.3).
- Beyond 3 kpc, the gradient is still marked. This is also where the smaller gradient, between l ~105° and l ~110°, and between $\mu_l \sim$ -1 mas yr^{-1} and $\mu_{l,\mathrm{res}} \sim$ 0 mas yr^{-1}, becomes apparent.

There is a linear gradient still present in the residual proper motions, especially at distances where the distribution of SED-fitted OB stars intercept the Perseus spiral arm. This would favour the interpretation of this gradient as being a relics of the star formation process, resulting in a large-scale expansion pattern. Nonetheless, this gradient appears between 2 and 4 kpc, much larger than one would expect for a local episode of star formation.

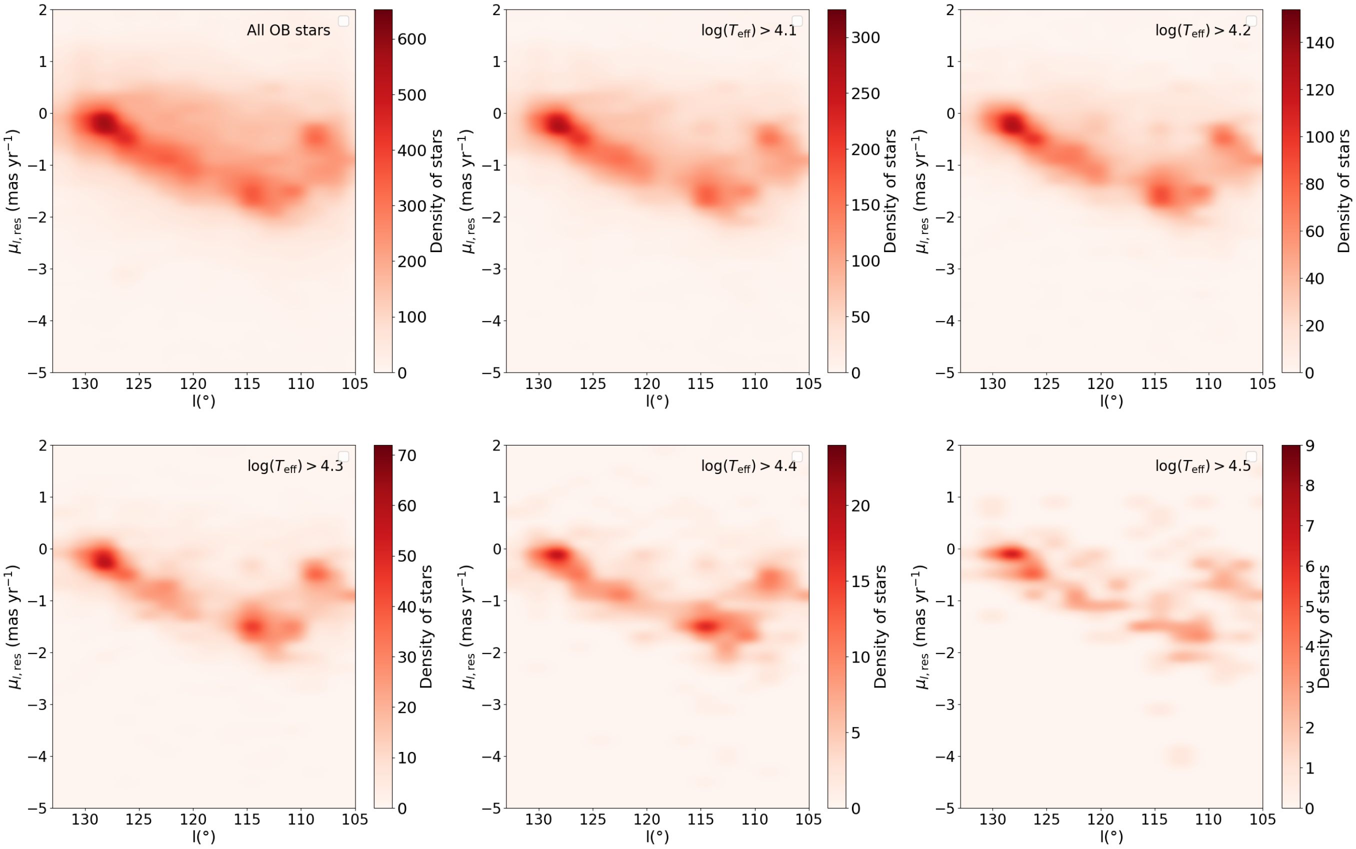


Figure 6.7: Density plot of the residual proper motions in Galactic longitude as a function Galactic longitude for the 40,107 SED-fitted OB stars in Cassiopeia. The top left panel displays the distribution for all effective temperature while the other panels show the distribution at different effective temperature subsets.

6.3.3 The kinematic pattern as a function of effective temperature

For this kinematic pattern to be the result of local star formation, it should be shaped by the youngest and thereby hottest stars of the sample. To verify this, I plotted again the projected view of the kinematic pattern between $\mu_{l,\rm res}$ and l in Cassiopeia, this time limiting the sample to hotter stars. This is shown in Fig. 6.7.

As the sample is reduced to the hottest stars, the distribution becomes more and more narrowed down to the gradient pattern, particularly for stars with $\log(T_{\rm eff}/K) > 4.3$ and $\log(T_{\rm eff}/K) > 4.4$. This suggests that the youngest stars in the region are strongly contributing to the gradient pattern observed. This should however be contrasted by their lower number and therefore lower density. Even if the cooler stars are more spread out in Fig. 6.7, they still dominate the stars of the kinematic pattern, which would argue against the star formation interpretation as late B-type stars can live up to hundreds of Myr old (Ekström et al., 2012).

6.3.4 Isolating the gradient

Sections 6.3.2 and 6.3.3 revealed that the kinematic pattern could not be entirely attributed to a population of young stars sitting over a short range of distances. Nevertheless, the fractional error on *Gaia* parallaxes become larger at greater distances, so one could wonder if such spread can be attributed to the effect of the observed uncertainties.

I used Topcat (Taylor, 2005) to select all the 34,927 SED-fitted OB stars above 2 kpc sitting across the gradient in the μ_l vs l diagram, together with an background around this gradient, as shown in Fig. 6.8. Both samples have a similar size, containing respectively 15,482 and 15,498 stars.

As the gradient sample will include both stars part of the pattern and some which are not, I isolated the gradient component. To that end, I binned both the gradient and background samples with a step size of 0.01 mas in parallax, weighted the distribution by their area (that I determined by assuming a rough elliptical shape), and subtracted their density. The resulting distribution of observed parallaxes for the gradient-background sample is shown in Fig. 6.9.

Fig. 6.9 shows that this sample exhibits some spread in its distribution of observed parallaxes. The median value is equal to 0.31 mas, while its 25th and 75th percentile values are respectively equal to 0.27 and 0.34 mas.

Could this spread be solely be attributed by the uncertainties in observed parallaxes? I randomly generated a uniform sample of stars between 2.5 and 3.5 kpc, that I

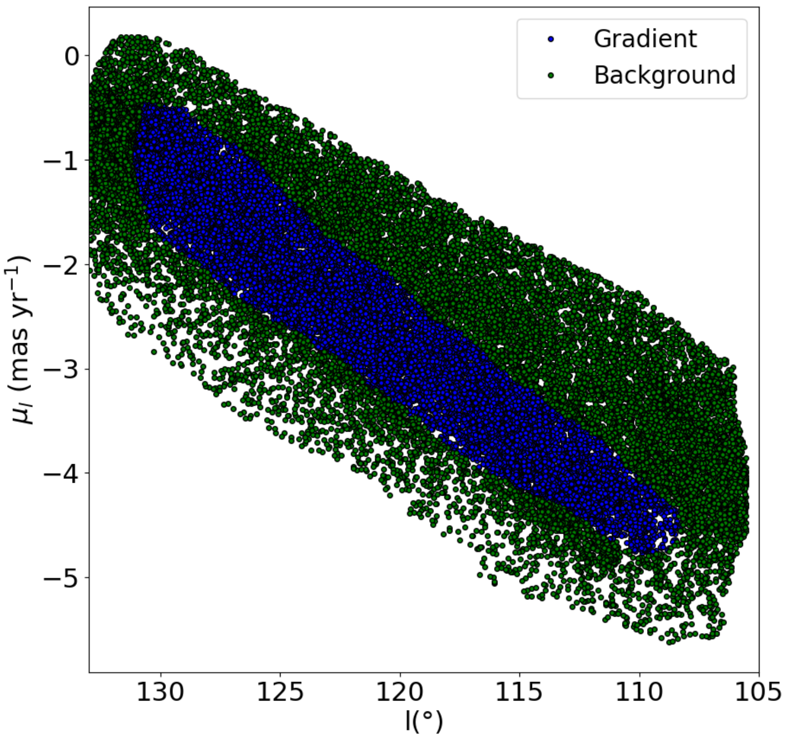


Figure 6.8: Highlights of the gradient (in blue) and the background (in green) samples across the plot of proper motion in Galactic longitude as a function of Galactic longitude.

then converted into parallaxes. Subsequently, I randomly shuffled the observed parallax uncertainties and associated them to these simulated parallaxes: these uncertainties corresponded to the 1σ value of a random Gaussian distribution.

The resulting distribution is also shown in Fig. 6.9, with respective 25th, 50th and 75th percentile values of 0.31, 0.33 and 0.36 mas. While there is some overlap with the observed distribution (including a similar peak between 0.30 and 0.35 mas), the latter still remain wider, containing stars at lower parallaxes. A larger range of distances would therefore be needed to explain the observed distribution, which would again argue against the local episode of star formation interpretation. The evidence appears to suggest that the gradient we have observed is unlikely to be due to a single episode of star formation, and therefore is more likely due to large scale Galactic rotation.

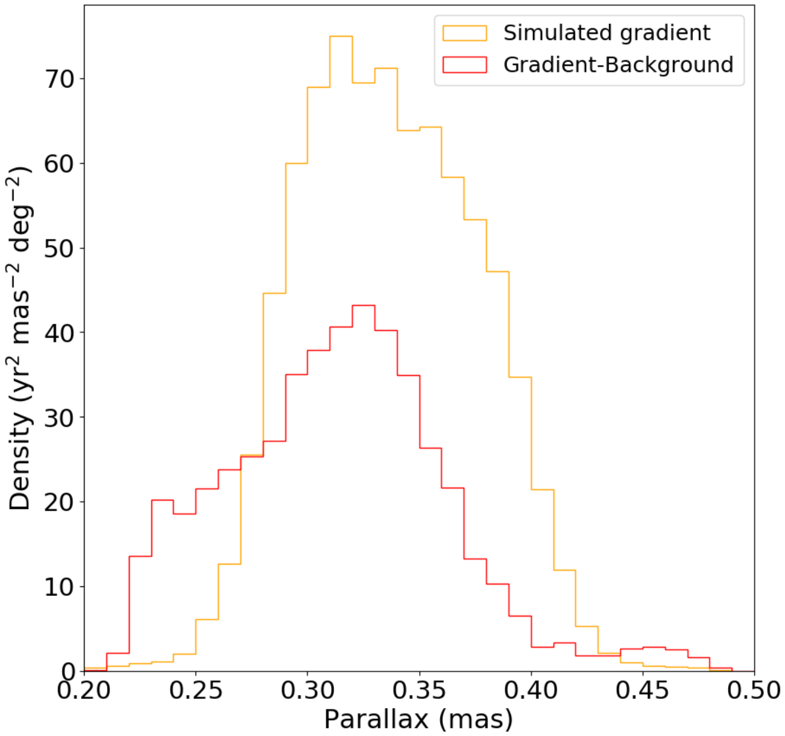


Figure 6.9: Parallax distribution as a function of density. for the observed stars in the gradient-background sample (in red) and for a simulated population of stars between 2.5 and 3.5 kpc sitting across the gradient (in orange).

6.4 Kinematic traceback

In this section I analyse the dynamical history of the identified OB stars by performing a kinematic traceback on them.

6.4.1 2D kinematic traceback

Fig. 6.6 shows that the observed gradient can mostly be attributed to OB stars more distant than 2 kpc. It is thus worth combining position and proper motions for these 34,927 OB stars with $d > 2$ kpc in Cassiopeia to perform a 2D linear traceback. To that aim, I apply the following equations:

$$l(t) = l(0) - 0.277\,\mu_l\,t \tag{6.2}$$

$$b(t) = b(0) - 0.277\,\mu_b\,t \tag{6.3}$$

where t is the time in the past in Myr where I perform the traceback, 0 stands for the present time and 0.277 is the conversion factor for the PMs from mas yr^{-1} to deg Myr^{-1}.

I applied a 2D kinematic traceback on these stars up to 30 Myr in the past, with a time step of 0.1 Myr. I worked out the time in the past these stars reached their most compact state by calculating the MAD on the plane-of-the-sky. I repeated this process 1000 times, randomly varying the PMs within their uncertainties. The result is displayed in Fig. 6.10.

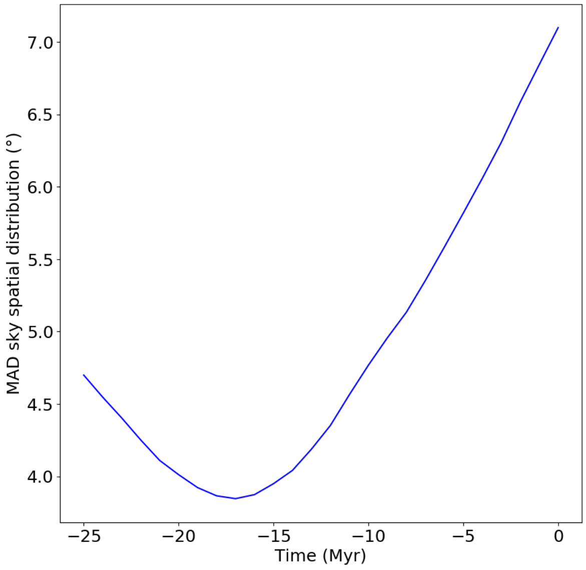


Figure 6.10: 2D MAD as a function of look-back time, from the 2D linear kinematic traceback on the 34,927 SED-fitted OB stars more distant than 2 kpc.

The stars reached their most compact state $17.0^{+0.3}_{-0.2}$ Myr ago. A more detailed view of this traceback is illustrated in Fig. 6.11, where I showed the position of these stars colour-coded by density by snapshots of 5 Myr, from the present time to 25 Myr in

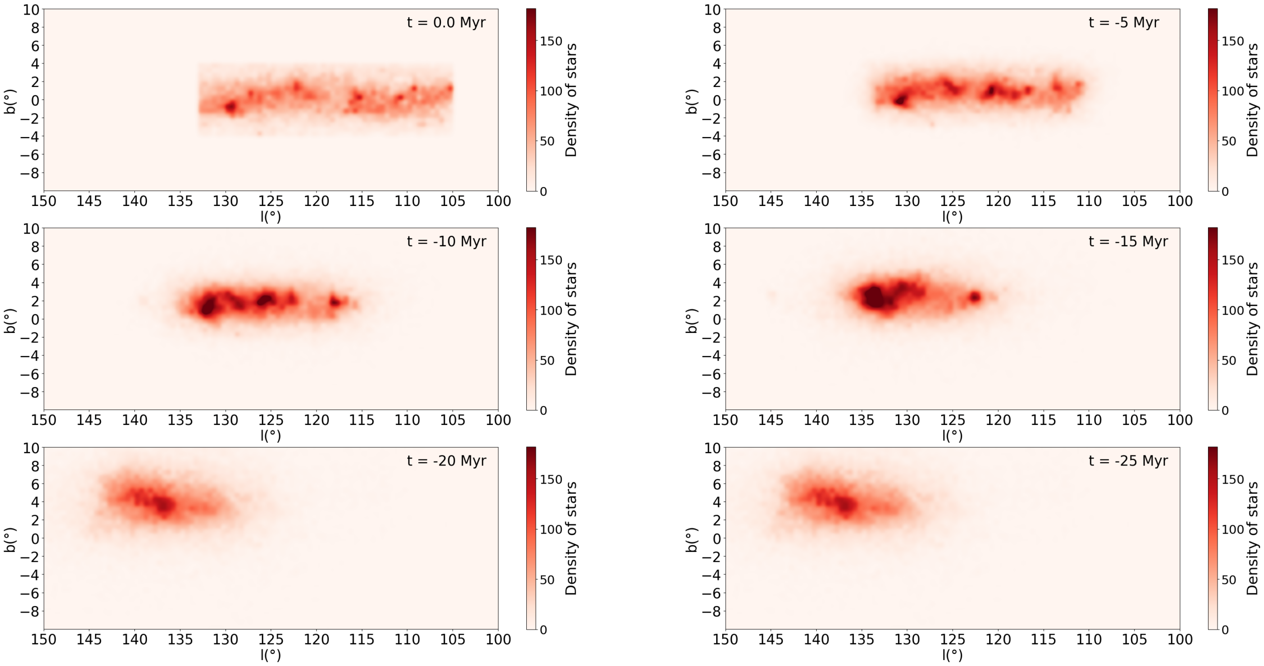


Figure 6.11: Density plot of the 34,927 OB stars as a function of look-back time. A common density scale has been chosen for every panel, with a bin size equal to 0.5 deg.

the past. Unveiled is a picture of stars converging in a common area at high longitudes, especially highlighted 15 and 20 Myr ago.

In Fig. 6.11 appears as well a smaller overdensity at 10 and 15 Myr ago, corresponding to the smaller gradient at lower longitudes shown in Fig. 6.5. It is overshadowed by the bigger one, that is therefore not apparent in the general MAD displayed in Fig. 6.10.

It has however to be emphasized that this kinematic traceback can only provide a simplified picture of the motion of the stars, as the third dimension has not been considered. An attempt at providing a more realistic view is described in Section 6.4.2.

6.4.2 3D kinematic traceback

A more precise traceback view on the kinematics requires the incorporation of RVs. This thereby requires to gather these measurements, that I take from two different sources:

1. 846 out of the 40,107 SED-fitted OB stars have a measured RV from the literature (as taken from SIMBAD), and including an uncertainty. This number is further reduced to 806 by inspecting the references and keeping only stars whose RV measurement method can be considered reliable (e.g. discarding stars with an imprecise or non-spectroscopic measurement).
2. There are 373 stars from the working sample in APOGEE DR17 (García Pérez et al., 2016; Abdurro'uf et al., 2022), and this number reduces to 349 taking only the stars with a measured RV.

A crossmatch between the first and the second list gives 286 stars in common. Out of those, 271 are in APOGEE DR16 (Jönsson et al., 2020), therefore I exclude them from the SIMBAD list. I take the weighted average between the two measurements for the 15 remaining corresponding stars, which leads to a sample of 864 OB stars with RVs in Cassiopeia.

These RVs are shown as a function of SED-fitted distance in Fig. 6.12. Immediately apparent is the presence of a bias in the distribution of RVs, as there are more nearby stars with RVs, while the number of OB stars is higher at larger distances (see Fig. 6.2). In Fig. 6.12 is also highlighted a rough correlation between RV and distance. To verify its origin, I used again the Galactic rotation curve model from Gaia Collaboration et al. (2023b) with the potential from Bovy (2014). I chose b = 0°, averaged over all Galactic longitudes in Cassiopeia by a step of 1°, and adopted a distance step of 100 pc. The resulting line shown in Fig. 6.12 exhibits a dependence between these quantities. Consequently, this proves that the relation between RV and distance can be attributed to Galactic rotation.

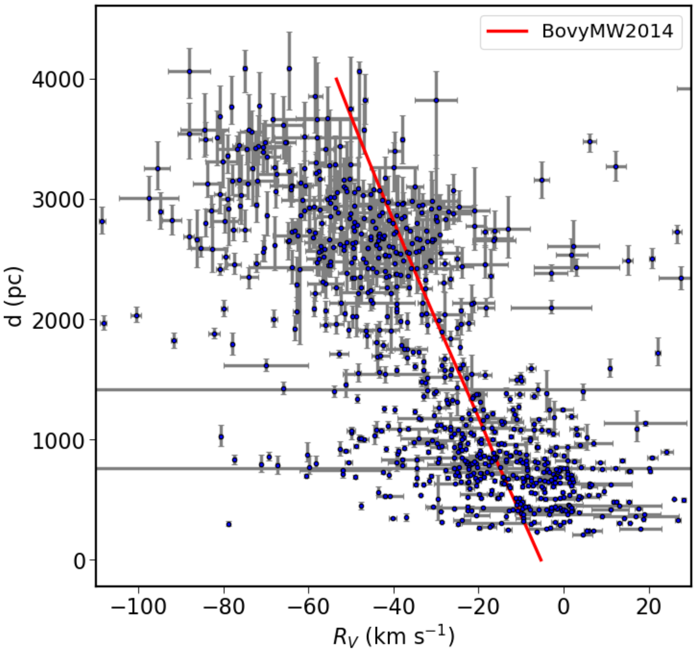


Figure 6.12: RV as a function of SED-fitted distance for the 864 OB stars in Cassiopeia with measured RVs. The red line shows the average prediction in this area from the Galactic rotation curve model from Gaia Collaboration et al. (2023b) that uses the potential from Bovy (2014).

Based on Fig. 6.12, I only kept stars whose RV is between -80 and 0 km s^{-1} in order to focus on the two peaks of the distribution, reducing the sample to 707 stars. I then applied a 3D kinematic traceback on these stars, again using the epicycle approximation from Fuchs et al. (2006) and the corrections for the LSR from Schönrich et al. (2010).

The results of this process are shown in Fig. 6.13, where the position of the stars on the plane-of-the-sky is displayed up to 30 Myr in the past by time steps of 10 Myr.

Fig. 6.13 reveals the presence of two subsets getting more compact in the past, consistent with the picture shown in Fig. 6.11 for the 2D linear traceback. Nevertheless, the 3D traceback provides an incomplete view, as only a minority of the OB stars have measured RVs.

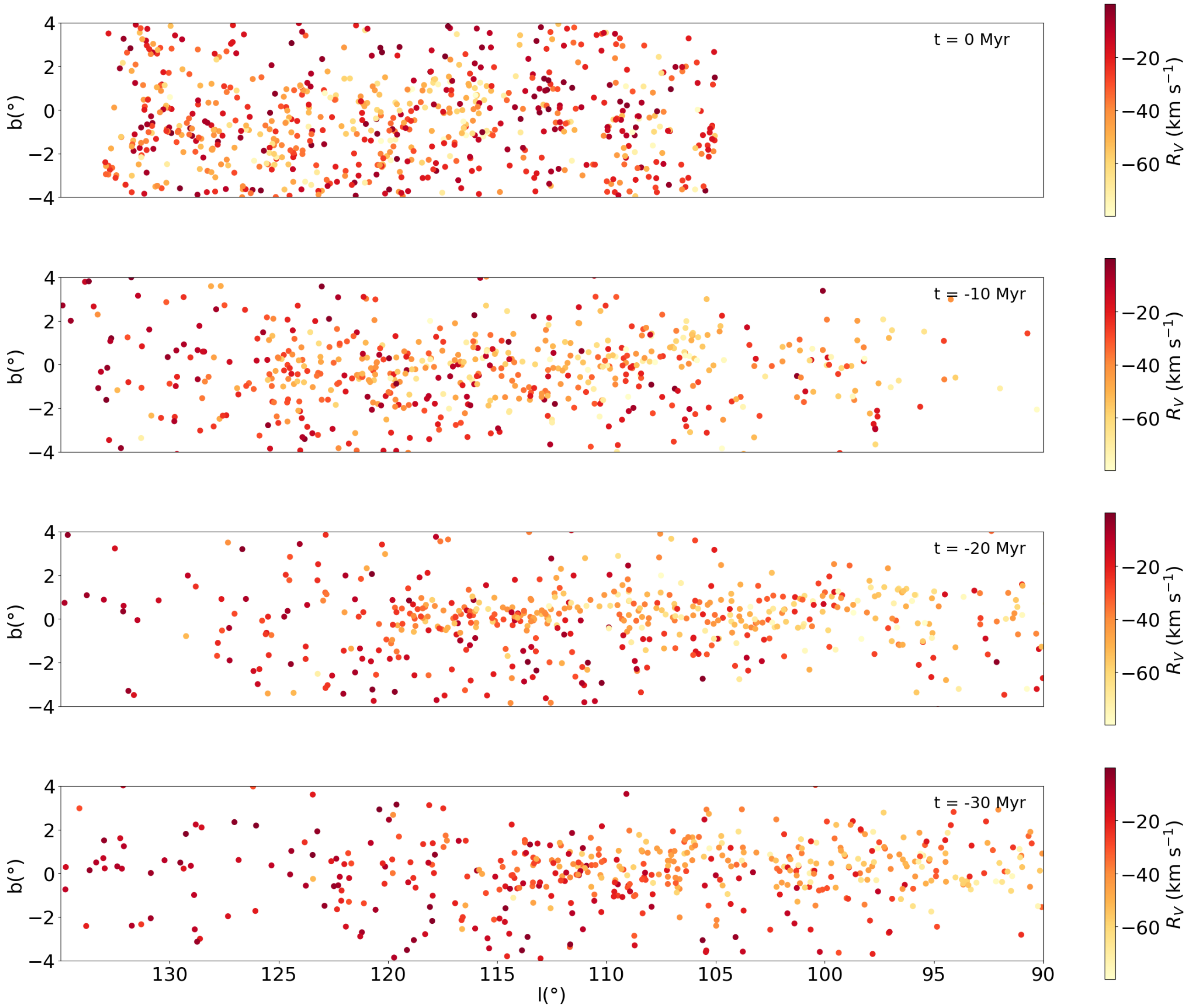


Figure 6.13: Galactic coordinates as a function of time for the 707 OB stars in Cassiopeia with RVs between -80 and 0 km s^{-1}, from the 3D kinematic traceback. Each star has been colour-coded by their measured RV.

To overcome this issue, obtaining more RVs would be ideal, but this depends on the availability of data from future spectroscopic surveys. Another, more direct approach would be to identify OB associations in Cassiopeia, as this would allow me to obtain average RVs, as I did for Cygnus and Auriga in Chapters 4 and 5.

6.5 Next steps of the work

Analysing the kinematics of the general distribution of OB stars within Cassiopeia already offered promising perspectives, as they were also found to be shaped by a linear correlation between Galactic longitude and proper motions in Galactic longitude. However, not only does Galactic rotation contributes at least partially to this gradient, the spread in distance nuances the interpretations as being a large-scale expansion pattern shaped by a star formation event. Furthermore, while the kinematic tracebacks provided a glance of the recent star formation history in Cassiopeia, the lack of available measured RVs prevented us from performing precise calculations.

The next steps of this work will then naturally consist of identifying OB associations across the Cassiopeia region. This will allow me to obtain better 3D kinematics, which in turn will help to unravel the recent star formation history of the region. Similarly to Auriga, given the extent of the probed volume, I will connect the OB associations to the surrounding open clusters, HII and star-forming regions, determine their kinematic ages, and verify whether they form an age gradient coincident with the position of the Milky Way spiral arms.

7 Introducing binarity into the SED fitter

This chapter describes ongoing work focused on introducing binarity into my SED fitter. I outline the motivations behind this modification, along with the details of this refined model. I then present the first applications of this new model, highlighting the changes compared to the previous versions of the SED fitting code, and discuss the results.

7.1 Motivation

The SED fitting code described in Chapter 2, and applied throughout this thesis to several massive star-forming regions (Chapters 3, 5 and 6), made the assumption that the observed SED originated due to light from a single star. In reality, as mentioned in Chapter 1, stars are often found in multiple systems. It is particularly the case for the massive OB stars which constitute the main targets of this thesis (Preibisch et al., 2001; Sana & Evans, 2011; Sana, 2017).

An accurate (and more honest) model should therefore account for the possibility that the observed SED is produced by light from a binary system, considering both the primary and secondary stars. Furthermore, we often do not know whether a given star is single or a binary, and so an ideal model would take this possibility into account. Implementing this model will have an effect on the derived physical parameters, and may affect the modelled ages of individual stars.

7.2 Existing knowledge of binary systems

The prevalence of binary and multiple systems can be described using several parameters. The first one of relevance is the multiplicity frequency MF, which is the probability that a star belongs to a multiple system. The review from Duchêne &

Table 7.1: Multiplicity frequency for different primary mass ranges, adapted from Duchêne & Kraus (2013). Abbreviations for the references are: A83 = Abt (1983), A90 = Abt et al. (1990), R97 = Reid & Gizis (1997), D04 = Delfosse et al. (2004), K05 = Kouwenhoven et al. (2005), A07 = Allen et al. (2007), M09 = Mason et al. (2009), R10 = Raghavan et al. (2010), D11 = De Rosa et al. (2011), S11 = Sana & Evans (2011), C12 = Chini et al. (2012), D12 = Dieterich et al. (2012) and DK13 = Duchêne & Kraus (2013).

Primary mass range	MF	References
$M_1 \leq 0.1 M_\odot$	22^{+6}_{-4} %	A07, DK13
$0.1 M_\odot \leq M_1 \leq 0.5 M_\odot$	26 ± 3 %	R97, D04, D12
$0.7 M_\odot \leq M_1 \leq 1.3 M_\odot$	44 ± 2 %	R10
$1.5 M_\odot \leq M_1 \leq 5 M_\odot$	≥ 50 %	A83, K05, D11
$8 M_\odot \leq M_1 \leq 16 M_\odot$	≥ 60 %	A90
$M1 \geq 16 M_\odot$	≥ 80 %	M09, S11, C12

Kraus (2013) gives tabulated values of MF as a function of the mass of the primary star, M_1, extracted from various studies. I have adapted it as shown in Table 7.1.

From the tabulations of Table 7.1, I perform a linear interpolation between M_1 and MF, in order to obtain a relation between these two quantities, as illustrated in Fig. 7.1. This will allow me to use this prior information in my model to determine the probability that the observed SED has been produced by a binary system (see Section 7.3).

A second, crucial parameter to describe binary systems is the mass ratio q. It is defined as $q = M_2/M_1$, where M_2 stands for the mass of the secondary star. A plethora of studies suggest that q typically follows a power-law $f(q) = q^\gamma$, as displayed in Fig. 7.2, where γ is the power-law index. To illustrate the relation between M_1 and γ, I gathered the different power-law indexes for the mass ratio distributions listed in Duchêne & Kraus (2013) for the different mass ranges, whose references are indicated in the caption of Fig. 7.2. I then sampled a probability density function, written as:

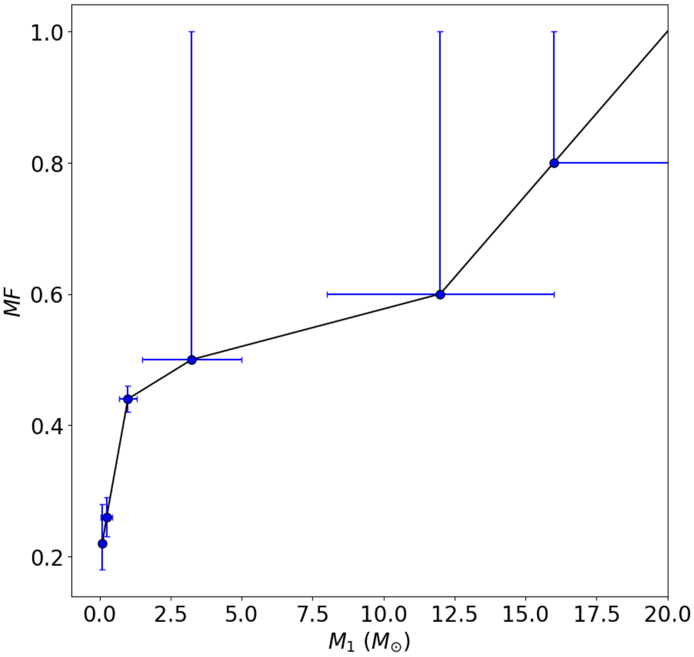


Figure 7.1: Relation between multiplicity frequency and primary star mass from the tabulated values of Table 7.1. The linear interpolation is shown in black. The horizontal error bars correspond to every mass range from the left panel, where the dots are their centre (apart from 0.1 and 16 $M_\odot$ where they are respectively the upper and lower threshold). The vertical error bars stand for the uncertainties in MF (for the higher primary star mass ranges, they are only lower thresholds).

$$P(x, a) = a\, x^{a-1} \tag{7.1}$$

where $a > 0$ and $0 \leq x \leq 1$.

I generated 100,000 values of x using Eq. (7.1) that follow the distribution with the power-law index γ, such that $a = \gamma + 1$. I plotted the resulting histograms for each power-law index and mass range in Fig. 7.2.

Fig. 7.2 illustrates the complexity of modelling mass ratio distributions as a function of primary star mass. Various studies derived distinct values for γ within the same mass range. They also found that both the orbital period (P) and separation (a) are related to γ.

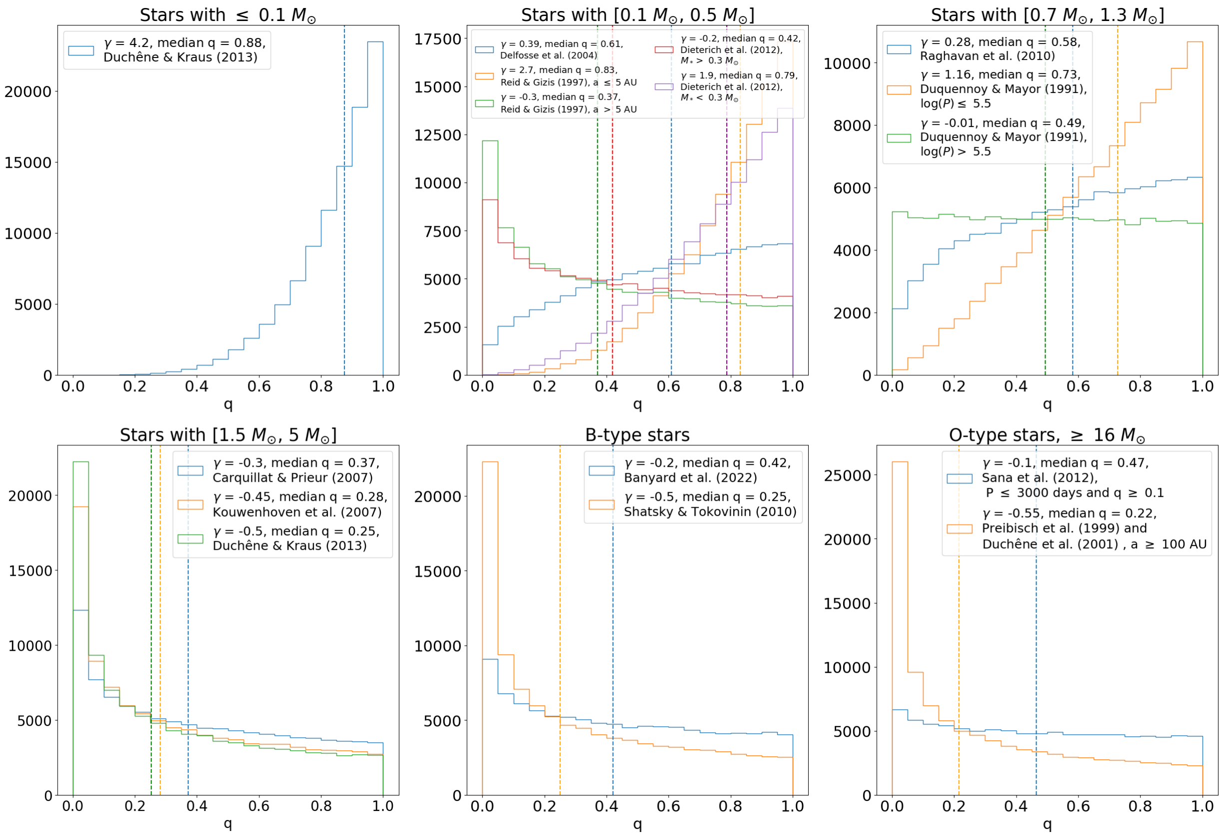


Figure 7.2: Mass ratio distributions for different mass ranges drawn from different power-law indexes, where the median values of q have been displayed in dashed lines. References are: Duchêne & Kraus (2013) for stars with $\leq 0.1\ M_\odot$. Reid & Gizis (1997), Delfosse et al. (2004) and Dieterich et al. (2012) for stars with [0.1 $M_\odot$, 0.5 $M_\odot$], Duquennoy & Mayor (1991) and Raghavan et al. (2010) for stars with [0.7 $M_\odot$, 1.3 $M_\odot$], Carquillat & Prieur (2007), Kouwenhoven et al. (2007) and Duchêne & Kraus (2013) for stars with [1.5 $M_\odot$, 5 $M_\odot$], Shatsky & Tokovinin (2002) and Banyard et al. (2022) for B-type stars, and Preibisch et al. (1999), Duchêne et al. (2001) and Sana et al. (2012) for O-type stars.

7.3 Implementation

The goal here is to allow my SED fitter to model an observed SED as either a single or a binary star for each iteration, depending on which better fits the data[1]. For this I introduce two new model parameters, the binary threshold P, which determines whether a star is modelled as a single star or a binary, and the mass ratio q, which gives the relation between and the primary and secondary star mass. Both parameters are outlined below:

- The binary threshold parameter, P, is a free parameter capable of modelling any star as either a single star or a binary star (depending on which fits the data better), but should be guided by the prior knowledge that more massive stars are more likely to be found in a binary. Its values range between 0 and 1. For each iteration, if P is lower than MF (interpolated from the parameter M_1, see Table 7.1 and Fig. 7.1), then the observed SED is modelled as a binary system, otherwise it is modelled as a single star.

- The mass ratio parameter, q, is a free parameter that ranges between 0 and 1 (under the definition that the primary star is the most massive within a binary system), although I will discuss the impact of such a choice in Section 7.4.3.

The updated version of the SED fitting code now contains six model parameters: $\log(M/M_{\odot})$, Fr(Age), d, P, q and ln(f). Two paths are possible, controlled by the values of P and M_1 on each iteration, as outlined above. The case of a single star will remain identical to the process explained in Chapter 2. In the binary case, on the other hand, the physical parameters of the secondary star will also be derived, following these equations:

$$M_2 = q\,M_1 \tag{7.2}$$

$$\mathrm{Fr}(\mathrm{Age}_2) = \frac{\mathrm{Age}_1}{\mathrm{Max}(\mathrm{Age}_2)} \tag{7.3}$$

---

[1] This improved version can however only account for non-interacting models

where Age_1 and Age_2 stand for the modelled ages of the primary and secondary star, respectively. The maximum age of the secondary star is constrained through an interpolation of the stellar evolutionary models from Ekström et al. (2012).

Once both the mass and the fractional age of the the secondary star are obtained, they are converted into effective temperature and luminosity (using the evolutionary models from Ekström et al. 2012). I convert the magnitudes into flux for both the primary and secondary stars through $F_x = 10^{-M_x/2.5}$ where M_x designates the magnitude for a photometric band x. The fluxes for the primary and secondary star are summed together, and converted back into magnitudes to produce the total model SED, which is compared to the observed SED.

The whole process produces a set of physical parameters for both the primary and secondary stars for every observed SED. Besides the properties of the primary star, I now also record the properties of the modelled secondary star along with the binary probability (i.e. the number of iterations where the SED was fitted as binary divided by the total number of iterations of the MCMC process).

An illustration of the binary SED fitting code is shown in Fig. 7.3, where both the modelled primary and secondary stars have been displayed in an HR diagram with 3σ confidence intervals. For this binary system, there is good agreement between the fitted temperatures and the temperatures derived from the spectral types (although my SED fitter modelled the secondary star as hotter than the primary star).

7.4 Results

In this section I summarize the first applications of the binary model on several samples, and emphasize the differences compared to fits with the single star SED model. Throughout this section, the single model corresponds to version 1.2 of the SED fitting process described in Section 2.8.

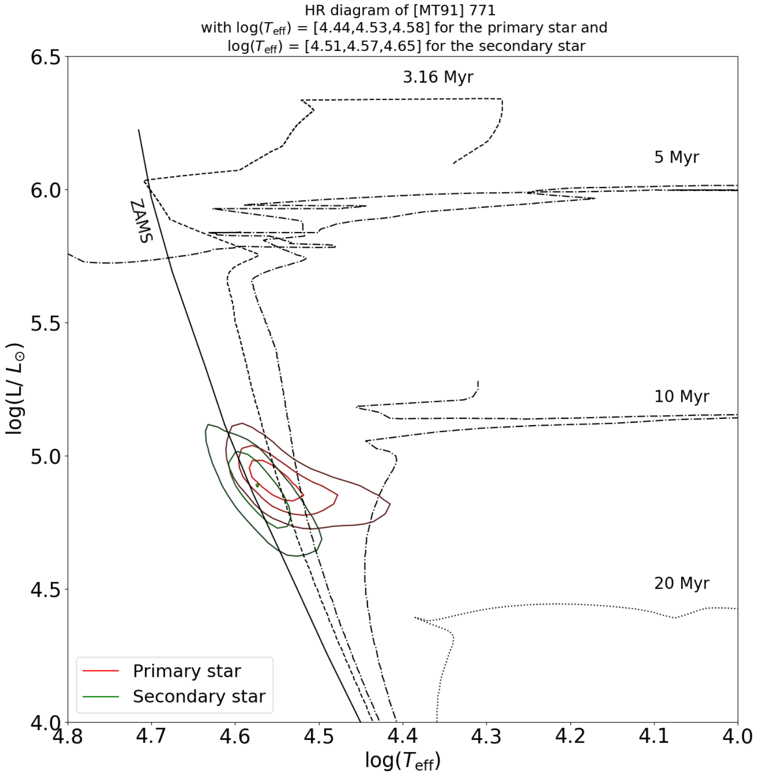


Figure 7.3: HR diagram of the star [MT91] 771 for the modelled primary (in red) and secondary (in green) stars, with their posterior distribution shown as 3σ contours. [MT91] 771 is a binary system of spectral type O7V+O9V (Wright et al., 2015), corresponding to respective $\log(T_{\rm eff})$ of 4.57 and 4.53, respectively, from Pecaut & Mamajek (2013). The fitted effective temperatures of $\log(T_{\rm eff}/K) = 4.57^{+0.08}_{-0.06}$ and $\log(T_{\rm eff}/K) = 4.53^{+0.05}_{-0.09}$ are in good agreement with these values. Isochrones have been taken from Ekström et al. (2012).

7.4.1 Model comparison for individual stars

The first application of my binary model consists of comparing the SED-fitted stellar parameters between those produced by the single star and the binary star SED models. To that aim, a suitable candidate is the catalogue from Wright et al. (2015), as it provides a list of massive stars in Cyg OB2, including information on their binary companions where known.

The catalogue from Wright et al. (2015) contains 167 stars, but this number was reduced to 105 when I applied the usual astrometric data tests and kept only the stars with valid photometry in a blue band. I performed the SED fitting process with both the single star and binary star models to all stars in this list, deriving two sets of physical parameters.

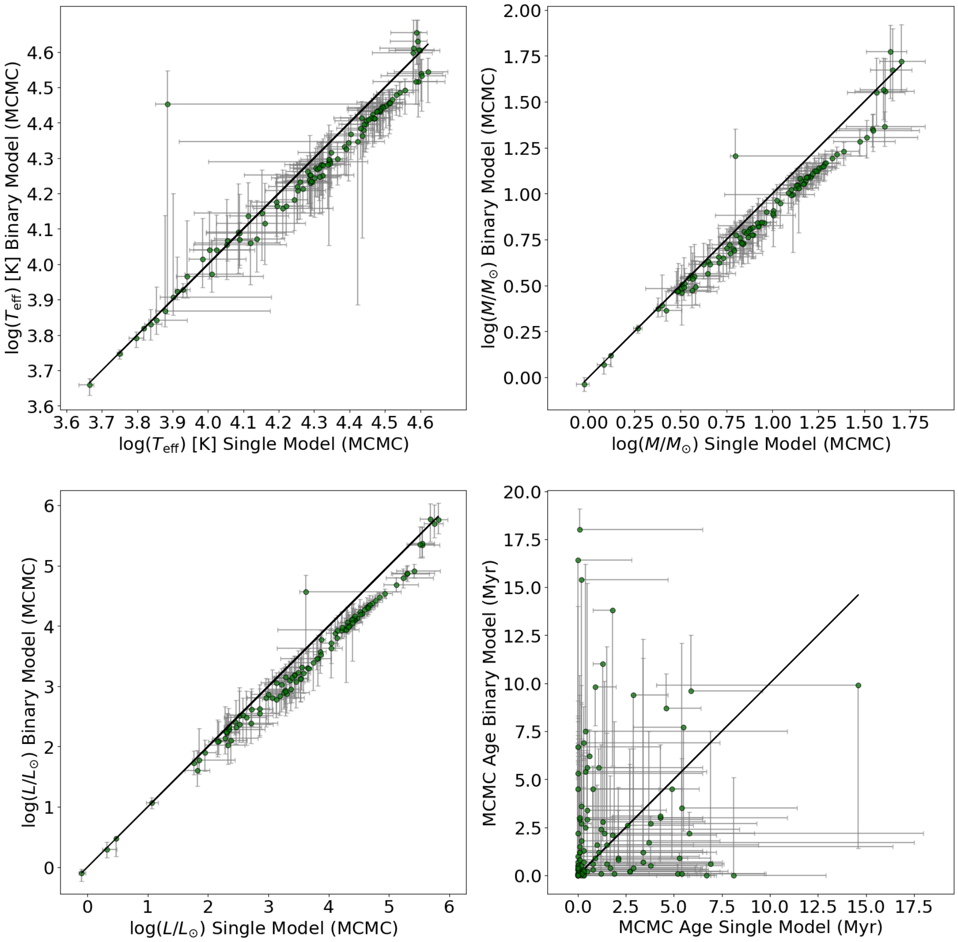


Figure 7.4: Comparison between the SED-fitted parameters of the single (abscissa) and binary (ordinate) models. The black line corresponds to a 1:1 line for the single model.

The comparison in effective temperature, luminosity, mass and age between the single and binary star models, is shown in Fig. 7.4. The general trend for the three first parameters is that for stars with a low temperature, mass or luminosity (i.e., lower mass stars) there is general agreement between the single and binary star models (consistent with such stars being more likely to be single). For more massive stars (with higher temperatures, masses and luminosities) there is a growing tendency for the primary star to be fitted with a lower value of the relevant parameter when fitted with the binary star model than when fitted with the single star model. This trend can however be reversed for the hottest stars, for which the binary model sometimes fits slightly larger temperatures. The comparison in age does not exhibit any strong correlation.

To better quantify the difference between the models, I considered an extreme case where the observed SED is fitted as a binary for every iteration, which I call the 'binary all' star model. The same comparison between the SED-fitted parameters of the single star model and the latter model is shown in Fig. 7.5.

Fig. 7.5 illustrates a general trend of fitting lower effective temperature, luminosity and mass for primary stars of all temperatures with the binary all star model, contrasted with the single star model. To quantify this, for both the classical binary and binary all star models, I worked out the relative difference in effective temperature between that fitted with the single star model and the binary star models, binned with a size of 0.1 dex in $\log(T_{\rm eff})$. I repeated this process 100,000 times, randomly varying the temperatures within their uncertainties, and chose the 16th and 84th percentiles of the resulting distribution as the error bars in the relative difference.

This comparison is displayed in Fig. 7.6. Given the size of the error bars, there is no significant difference between the single and binary star models up to $\log(T_{\rm eff}/K) = 4.1$. The difference then becomes higher for the binary star model, with temperatures typically lower by 10-15 % compared with the single star model. The binary all model, on the other hand, exhibits even larger variations, with the highest difference observed in the bin with $\log(T_{\rm eff}/K) = 4.1 - 4.2$ where the difference reaches $\sim$25 %.

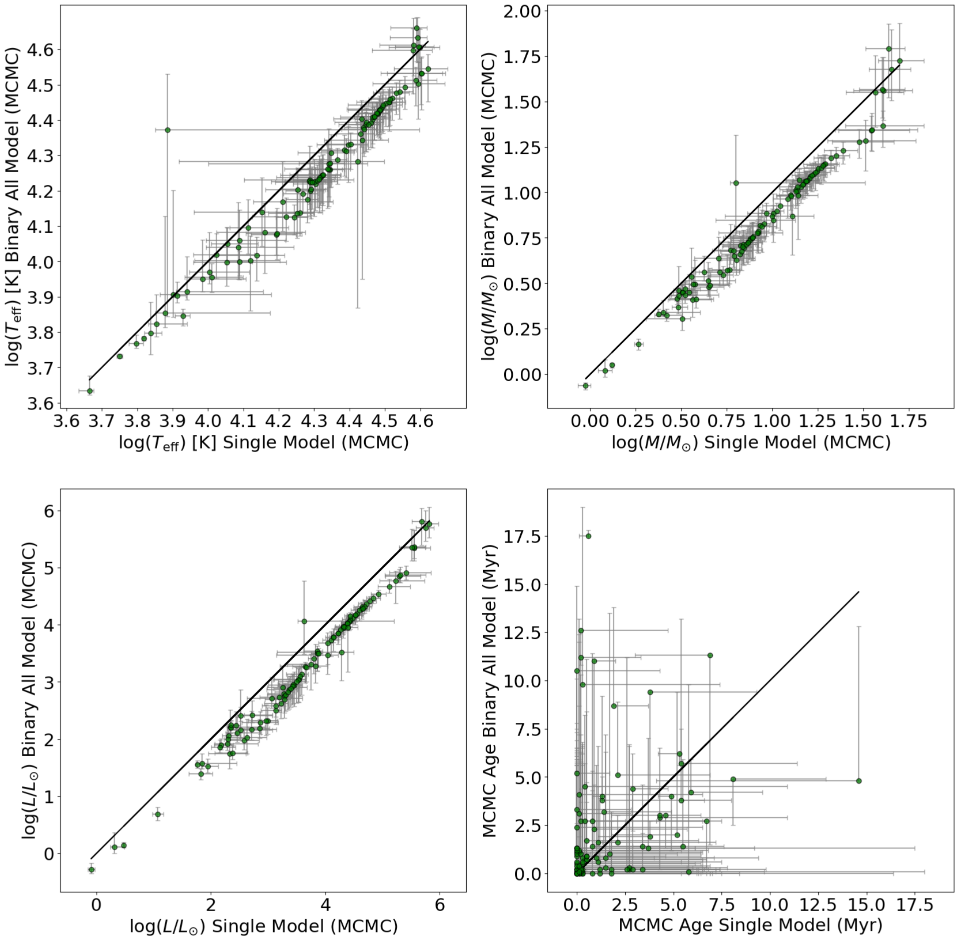


Figure 7.5: Same than Fig. 7.4 but with the binary all star model in ordinate instead.

7.4.2 Model comparison for whole associations

7.4.2.1 Total stellar mass

Throughout this thesis, I have derived physical properties of OB associations (Chapters 3 and 5). I noted that assuming every star to be single could affect the results: therefore, the binary model provides an opportunity to revisit them.

To explore how large this difference would be I selected association E from the Cygnus region (Chapter 3), as it contains the largest number of O-type stars of any

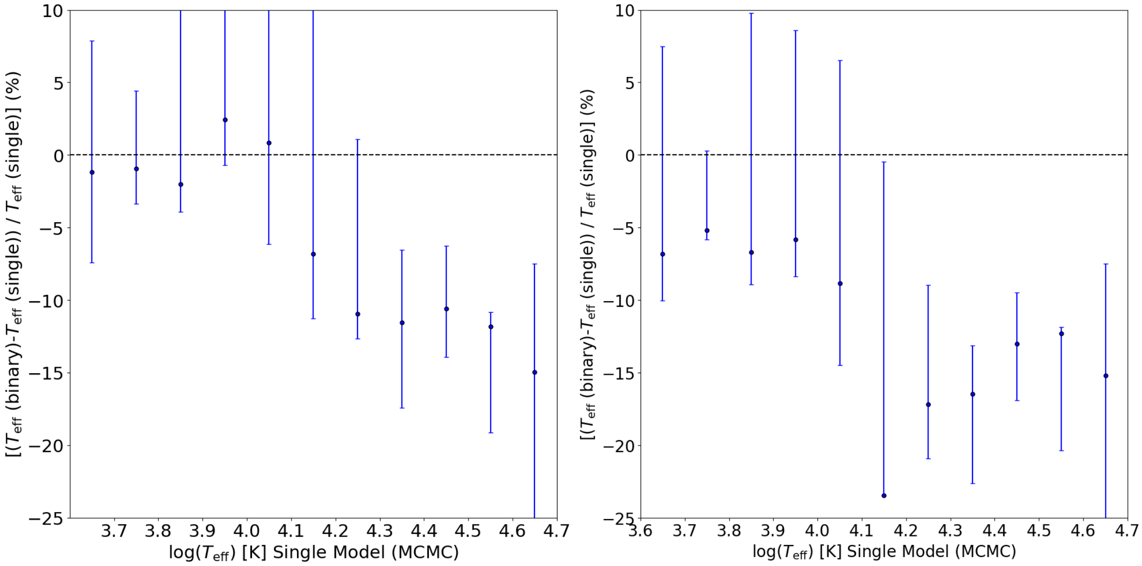


Figure 7.6: Relative difference between the single model and the binary (left panel) and binary all (right panel) star models, as a function of the effective temperature of the single model.

OB association studied in this thesis. I applied the SED fitter with both the single star model and the binary star model to its 168 members.

Binarity will affect the total stellar initial mass of a stellar group. Let M_{tot}, $M_{\mathrm{tot,P}}$ and $M_{\mathrm{tot,S}}$ be the total initial stellar mass of all stars, the primary stars and the secondary stars, respectively, of the association. I then study this case by considering four models:

1. **Single star model (1)**: Every SED is fitted using the single star model and assumed to correspond to a single star, i.e. $M_{\mathrm{tot}} = M_{\mathrm{tot,P}}$. This is the model I used for both the Cygnus (Chapter 3) and Auriga (Chapter 5) OB associations.

2. **Single star model with binary corrections (2)**: This model is similar to Model 1, but includes a simple correction for the presence of binary companions. I assume a binary fraction of 0.5 and an average mass ratio of 0.5 for non-OB stars, a binary fraction of 0.75 and an average mass ratio of 0.35 for

B-type stars, and a binary fraction of 1 and an average mass ratio of 0.35 for O-type stars. These values are based on the distribution displayed in Fig. 7.2.

3. **Binary star model (3)**: The mass of the secondary star for non-OB primaries are derived in the same way as Model 2, above. For the OB star primaries, the mass of the secondary star is derived from the posterior distribution of the binary star model fit. If the interpolated binary probability is above the fitted value of P, then the mass of the secondary star for this iteration is included, otherwise it is set to zero. This allows the mass of the secondary star to account for the fraction of iterations where the star is modelled as a single star. The final mass of the secondary star is then equal to the median value of the resulting distribution.

4. **Binary all star model (4)**: As described in Section 7.4.1, every SED is fitted as though it is a binary star, therefore $M_{\mathrm{tot}} = M_{\mathrm{tot,P}} + M_{\mathrm{tot,S}}$ with both $M_{\mathrm{tot,P}}$ and $M_{\mathrm{tot,S}}$ calculated with the same method.

For each model, M_{tot} is then estimated as in Chapter 3, sampling the mass functions from Maschberger (2013) to extrapolate the mass of the unobserved low-mass stars. The results are displayed in Table 7.2. Table 7.2 shows that by implementing the binary model (3), M_{tot} is increased by $\sim$25 % compared to the single star only model. This result has however to be contrasted with the models (2) and (4), for which a slightly larger M_{tot} has been derived (though they agree within their error bars). While both models have been based on approximations, the comparison is still worth noting. The estimation of $M_{\mathrm{tot,S}}$ is mostly based on the SED-fitted mass ratios which may require refinements. This parameter is explored in Section 7.4.3.

7.4.2.2 Position on the HR diagram

The analysis of OB associations revealed the presence of contaminants at higher ages sitting in their HR diagrams, as I discussed in Chapter 5 in the case of the Auriga OB

Table 7.2: Total stellar initial masses for association E in Cygnus, from the primary, secondary and all stars, respectively, derived from each model.

Model	$M_{\mathrm{tot,P}}\,(M_\odot)$	$M_{\mathrm{tot,S}}\,(M_\odot)$	$M_{\mathrm{tot}}\,(M_\odot)$
(1)	4347 ± 338	-	4347 ± 338
(2)	4347 ± 338	1898^{+222}_{-221}	6245 ± 404
(3)	4567^{+377}_{-340}	1139^{+172}_{-164}	5706^{+378}_{-377}
(4)	3975^{+355}_{-315}	2275 ± 253	6250^{+436}_{-404}

associations. One explanation that I invoked was that assuming every observed SED to be produced by the light for a single star could lead to erroneous ages.

Since the binary model fits slightly different stellar parameters than the single star model, this will have an impact on their position in the HR diagram. I illustrate this by comparing the HR diagram between the two models in Fig. 7.7. There is highlighted the influence of the binary star model on the primary star, as it mostly fitted with a younger age (i.e. 3.57 Myr for the binary star model vs 4.99 Myr for the single star model), although the introduction of more free parameters increased the size of its error bars.

7.4.3 Mass ratio distribution

The binary star model for SED fitting, described in Section 7.3, incorporates two new MCMC parameters, including the mass ratio q. So far, I decided to uniformly sample it between 0 and 1, even if Fig. 7.2 shows different distributions at different primary star masses.

Fig. 7.2 exhibits a tendency of high and low q for low- and high-mass stars, respectively. Specifically, for B- and O-type stars, Fig. 7.2 shows median q values of 0.25-0.42 and 0.21-0.46, respectively. By contrast, B- and O-type stars in association E fitted with the binary star model have median values of 0.52 and 0.50, thus close to a uniform distribution. This motivates a deeper investigation of the SED-fitted mass

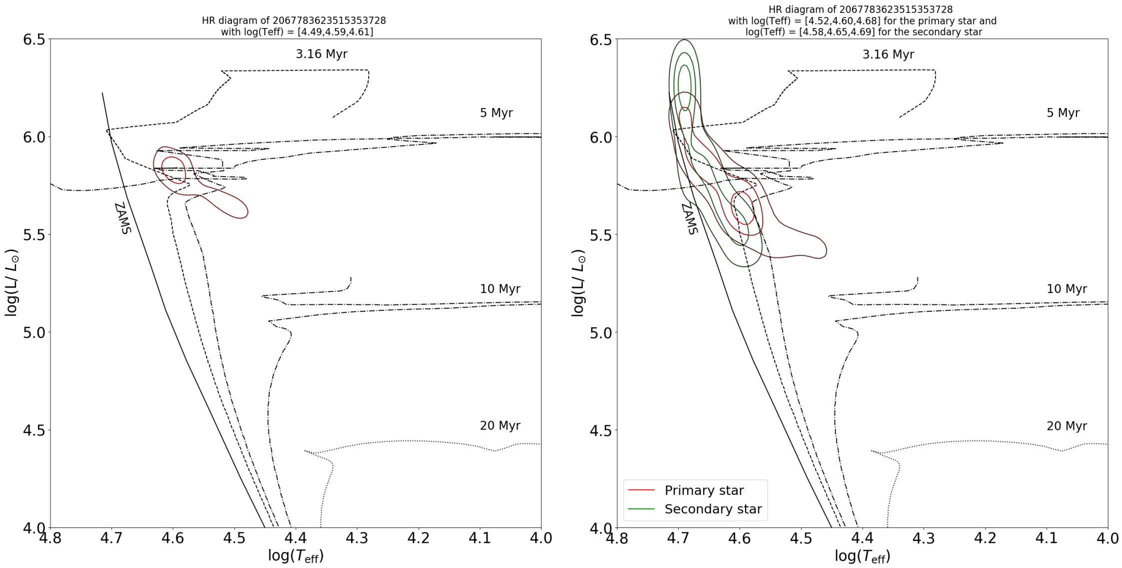


Figure 7.7: HR diagram of 2067783623515353728, with the modelled stars shown as 3σ contours. The left panel shows the results for the single star model and the right panel for the binary star model. The median ages for the primary star are 4.99 and 3.57 Myr, for the single and binary star model respectively.

ratios.

To that aim, I selected the first six stars of association E, and applied an extended version of the SED fitting process on them, with 10,000, 100,000 and 200,000 burnin iterations. In Fig. 7.8, I represented their resulting posterior distribution in q, only for the iterations where the SED is fitted as a binary star (as q becomes meaningless otherwise).

Some stars in Fig. 7.8 exhibit peaks that suggest a non-uniform distribution. This is particularly the case for the first, second, third and fifth star, whose main peaks are located around $q \approx 0.3$. Fig. 7.8 shows that increasing the number of burn-in iterations does not have an effect on the convergence of q.

To check where this modulation could come from, I performed an experiment for the third star in association E as an example. I applied the SED fitter, using 10,000 burnin iterations, with both the single star and binary all star models, and plotted

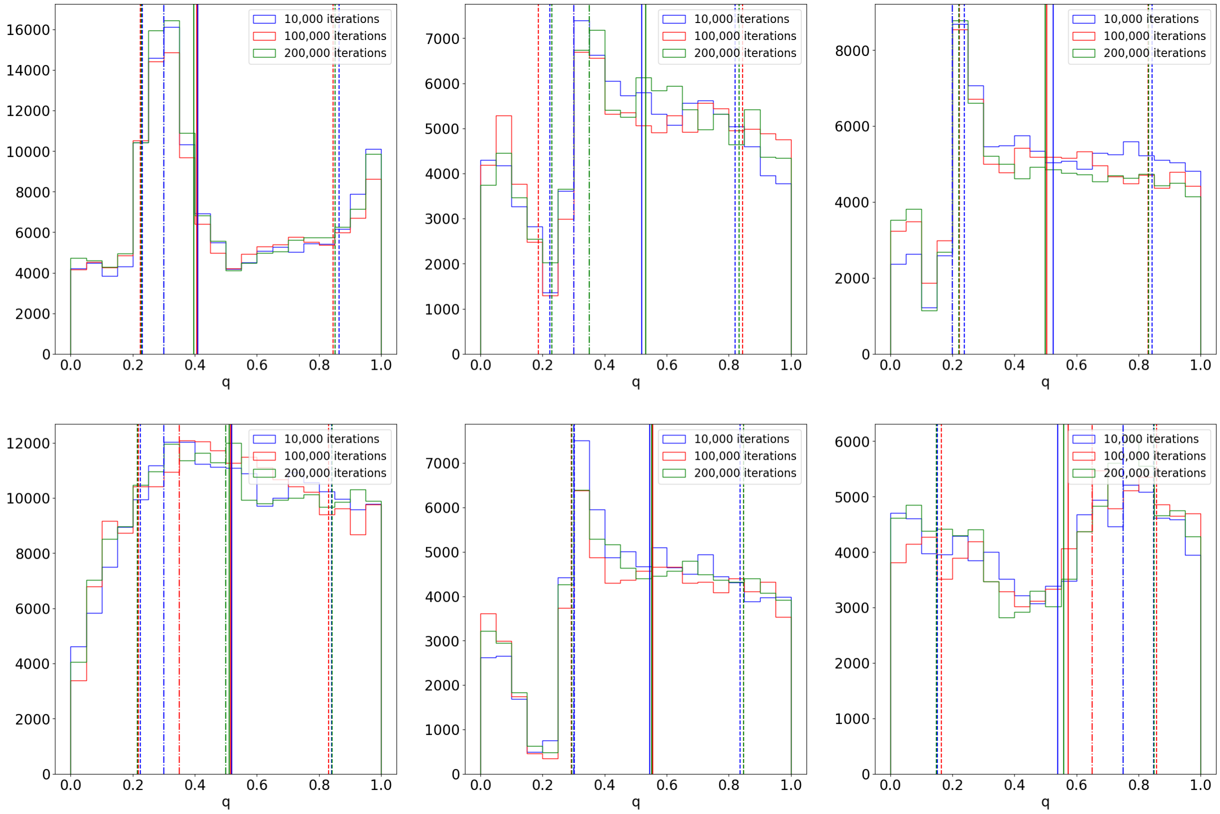


Figure 7.8: Posterior distribution for q for iterations where the SED is fitted as a binary, for the six first stars in association E. For each set of supplementary iterations, the median (solid line), 16th and 84th percentiles (dashed lines) and the peak (dot-dashed lines) of the distribution have been displayed.

the resulting corner plots in Fig. 7.9 to visualise the correlation between MCMC parameters.

A comparison between left and right panels of Fig. 7.9 shows that considering the SED to corresponding to a binary system has no effect on the SED-fitted distance (as it mostly depends on the observed parallax), but the binary model tends to fit lower primary star masses. In the right panel, the smaller values in the posterior distribution of q also tend to correspond to lower values of the primary star mass.

This analysis suggests that limiting the posterior distribution to binary iterations, and selecting the peaks rather than the median value of q, would better sample the

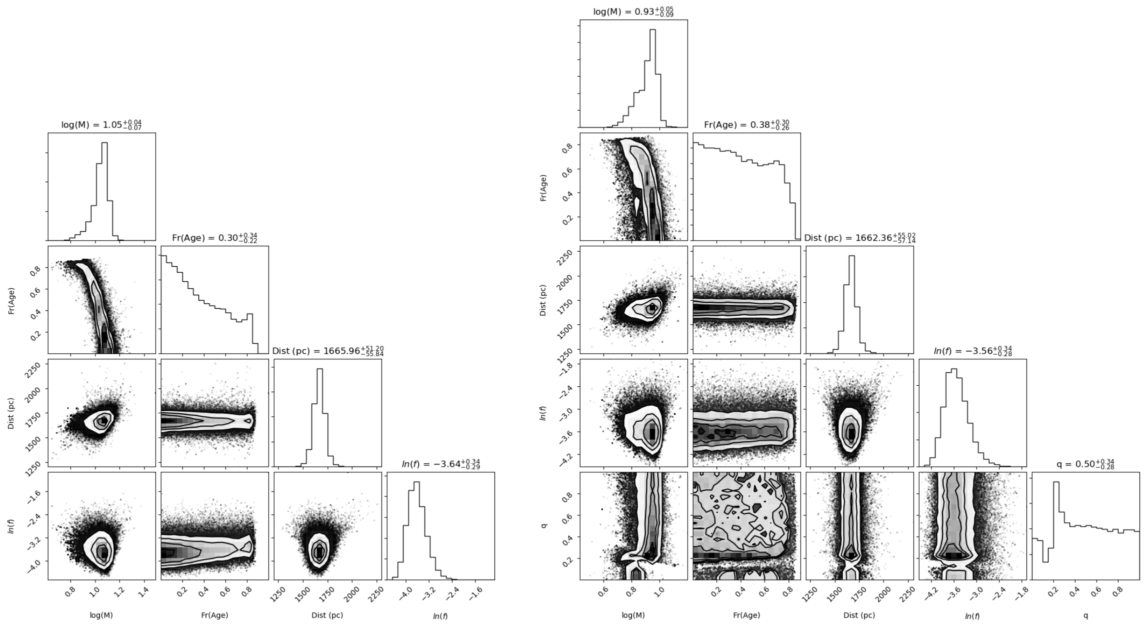


Figure 7.9: Triangle plot for the third star in association E. Left panel: Single star model. Right panel: Binary all star model.

value of q. Constraints on the mass ratio distribution however remain difficult through SED fitting, as the model currently stands. Nevertheless, as emphasized in Section 7.3, it is important to incorporate this model to be more truthful.

Possible future work would be to perform more tests to better understand what the model is doing, before applying it to existing associations once it is finalised. This will allow me to derive more accurate physical parameters for their members, including an estimation of their binary properties. Further improvement could be the replacement of the *Geneva* models from Ekström et al. (2012), which only models the evolution of single stars, by the *BPASS* models (Eldridge et al., 2017), which models the evolution of binary systems.

8 Conclusions and future work

8.1 Conclusions

These last decades have witnessed the arrival of large-scale astronomical surveys at an unprecedented level. This has revolutionized our understanding of a plethora of areas of astronomy, and allowed us to develop more advanced computing techniques. These advancements have provided the foundations for my work.

In this thesis, I utilised photometric and astrometric data to show that most historical OB associations are not coherent structures, and therefore need to be revisited. To that end, I built an SED fitting code that fits an observed SED, based on a *Gaia* parallax and the photometry from several optical and near-IR surveys, with a model SED, computed using stellar atmosphere and evolutionary models and an extinction map. This allowed me to derive a set of physical parameters using Bayesian inference, a maximum-likelihood test and an MCMC process. This code has been continuously improved during my work, with the current version focused on including binarity into the model and thereby also characterizing the properties of the secondary star.

I have applied this SED fitter to identify OB stars across multiple massive star-forming regions, specifically Cygnus, Auriga and Cassiopeia. I used these populations of OB stars to identify new, reliable, kinematically-coherent OB associations using various clustering algorithms. I characterized these OB associations both physically and kinematically, and studied their expansion. Their general properties are similar to those of other OB associations and therefore I have confidence suggesting that they should replace the majority of the historical associations in these regions of study.

I was able to go deeper in Cygnus as I identified two large-scale expansion patterns, which reached their most compact state $\sim$8 Myr ago as calculated using 3D kinematic traceback. I showed through their velocity field that the driver of such dynamics could be feedback, but is more probably due to the inherent turbulence within their primordial molecular cloud. This is in contrast to some recent observational studies that have suggested such large-scale expansion patterns are due to the effects of feedback.

In Auriga I identified an age gradient across many of the newly identified OB associations, whose position coincided with the motion of the Perseus spiral arm within the last ∼20 Myrs, suggesting that OB associations could be used to reconstruct the past positions of local spiral arms.

Finally, I identified tens of thousands of OB stars across the Cassiopeia region, whose kinematics revealed a similar kinematic pattern as in Cygnus, although it can be at least partially attributed to the Galactic rotation curve. From a kinematic traceback there is evidence that the OB stars in this region were more clustered in the past and were at their most compact ∼17 Myr ago, but further work is necessary to obtain a full picture of the region.

The results of this thesis thus offers promising perspectives into using OB associations as probes for better understanding star formation and Galactic evolution.

8.2 Future work

The thesis ends with a short outline of my potential future research plans.

8.2.1 All-sky distribution of OB stars and associations

Throughout this thesis, I have targeted specific massive star-forming regions in order to study the distribution of OB stars and associations in those regions. The issue with such an approach is that it was limited to relatively narrow sightlines. My future ambition is thus to conduct a larger scale and unbiased survey across the entire sky, with the end goal of creating an all-sky census of OB stars and associations that can be used for further analysis.

Several studies have previously mapped out the distribution of OB stars across the local Milky Way. However, they were either limited to O- and early B-type stars (e.g., Chen et al. 2019b; Pantaleoni González et al. 2021) or only identified candidate OB stars without further characterising their masses and ages (e.g. Zari et al. 2021). For this work, I plan to derive physical parameters for the OB stars with my SED fitter. Furthermore, including late B-type stars, that many studies have missed, will allow me to better sample stellar populations, estimate more accurate ages for older systems, and obtain a more complete study of the IMF.

Similarly, a new census of OB associations is particularly welcome. The original identification of OB associations dated back to when photometric (e.g. Humphreys 1978) and astrometric (e.g. de Zeeuw et al. 1999) datasets were of smaller size and lower quality. There have been more recent attempts at mapping out OB associations on a large scale (e.g. Chemel et al. 2022), but this was based on spectroscopically-identified OB stars, and is therefore limited and biased by the available spectroscopic sample. By identifying OB associations with modern tools, I can reconstruct the star formation history of the local Milky Way, because grouping stars that formed together allow me to derive more accurate ages, distances and kinematics than would be possible for individual stars. In this thesis I showed that OB associations are powerful probes to analyse the physical conditions driving the expansion or early evolution of young stellar groups (Chapter 4), and the motion of the Milky Way spiral arms over the past tens of Myr (Chapter 5). Consequently, I can apply this on a larger scale.

I have started to work on the very first steps of this project. An initial, reasonable sample would be to produce a census of OB stars within 1 kpc. To that end, I have prepared a working list of candidate OB stars, whose selection process is outlined below.

1. *Gaia* EDR3 and 2MASS data provide all-sky optical and NIR photometry. My model SEDs predict a maximum K_s of $\sim$14.5 mag for a 9000K star at a distance of 1 kpc, which is just above the magnitude limit for a point source in 2MASS (Cutri et al., 2003). Therefore, I expect every candidate OB star within 1 kpc observed by *Gaia* to also be in 2MASS, and I only selected sources both in *Gaia* and 2MASS.

2. In the *Gaia* archive, I applied the same astrometric and photometric quality

checks as in Chapters 3, 5 and 6, discarding the sources with bad astrometry and non-valid blue photometry. I also performed a liberal cut of $\varpi > 0.9$ mas. Assuming a maximum extinction of $A_V = 10$ mag at 1 kpc, that would translate into $A_{Ks} = 1$ mag, I selected the stars with $M_{Ks} < 2.07$ mag, again based on the absolute magnitude of an A0 star in the K_s band from Pecaut & Mamajek (2013). I used the H and J bands whenever the K_s and H bands had a bad photometric quality flag.

3. I removed stars with a geometric distance larger than 1 kpc in Bailer-Jones et al. (2021), and performed a cut in the NIR colour-colour for the sources with valid photometry in all of the three NIR bands.

This process resulted in a sample of 1,058,259 sources.

The next challenge is to perform a more reliable, extinction-dependent absolute magnitude cut, which requires an estimate of the extinction for each target. So far, I have applied the *Bayestar* extinction map from Green et al. (2019) to constrain stellar extinctions based on the distance of each individual star. Unfortunately this extinction map covers only three quarters of the sky, and misses the high longitudes and low latitudes in Galactic coordinates (see Fig. 8.1).

One solution to overcome this issue would be to combine the *Bayestar* extinction map with other extinction maps. There are two best-suited candidates amongst the list of available ones. The first one is the *Gaia* DR3 extinction map (Delchambre et al., 2023), already introduced in Chapter 2, which will be suited to provide reddening for *Gaia* sources outside the Galactic plane. The second one is the map from Chen et al. (2019a), which provides extinction based on *Gaia*, 2MASS and WISE colours across the entire Galactic plane. Both maps are shown in Fig. 8.2.

Combining these three extinction maps will increase the completeness of my study at the cost of homogeneity[1]. Once this will be implemented, I aim at using them to further trim the list, and obtain a working sample of candidate OB stars of reasonable

---

[1] Other all-sky extinction maps to be released in the near future may also be considered.

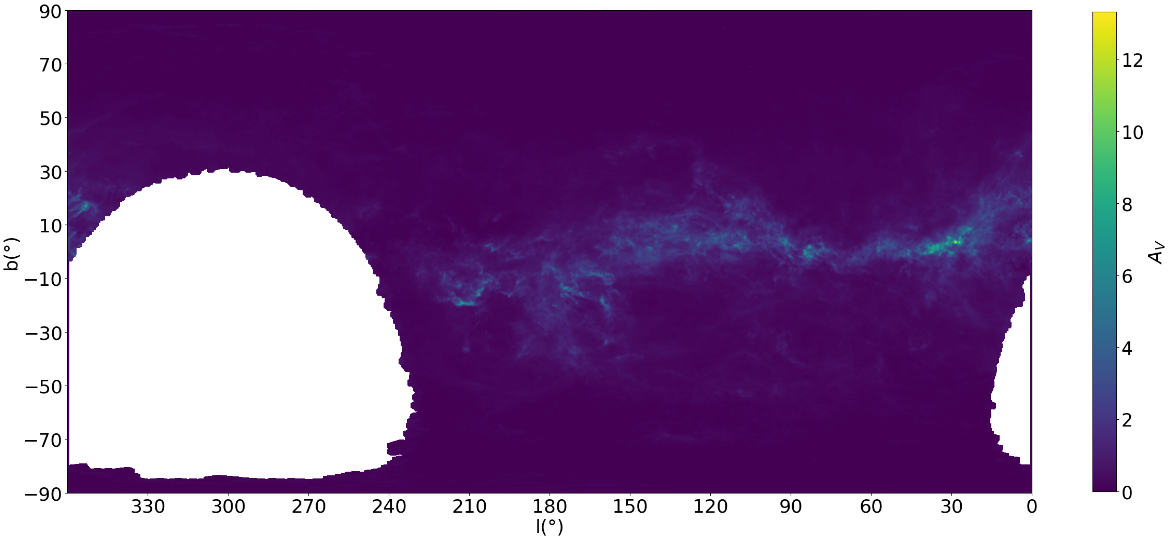


Figure 8.1: All-sky *Bayestar* extinction map from Green et al. (2019), integrated up to 1 kpc, shown in Galactic coordinates. White areas show areas of the sky not covered by the extinction map.

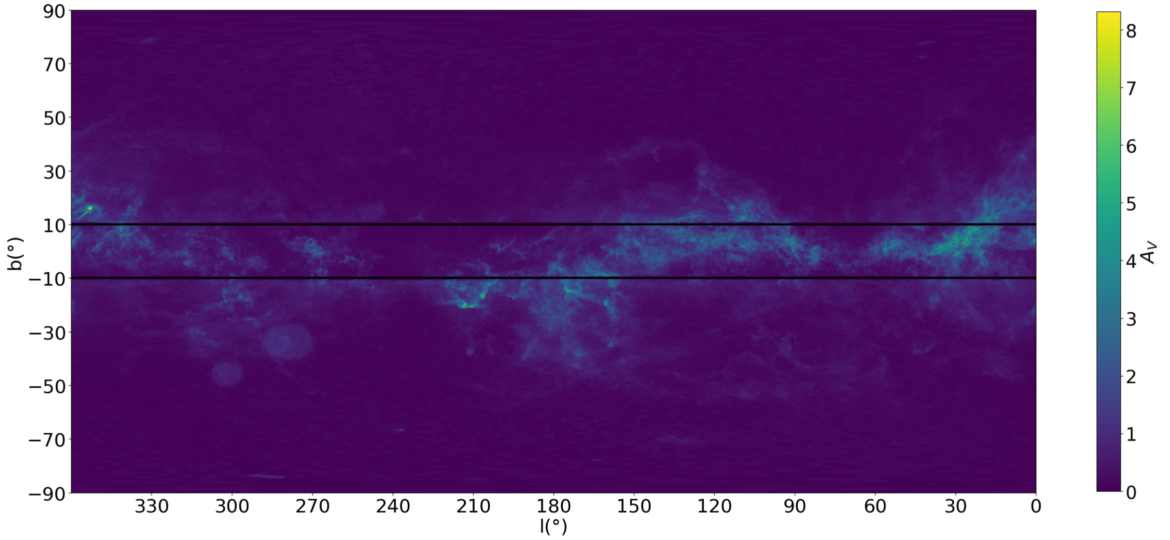


Figure 8.2: All-sky *Gaia* DR3 (Delchambre et al., 2023) and *Chen19* (Chen et al., 2019a) extinction maps, respectively shown outside and within the Galactic plane, as delimited by the black lines.

size (typically a few hundreds of thousands sources). My SED fitter will then estimate their stellar parameters, which I will use to produce a list of OB stars within 1 kpc. This will lay the foundations for the identification and analysis of OB associations within 1 kpc, that will serve as probes to reconstruct the recent star formation history of the local Milky Way.

8.2.2 Including spectroscopy in the SED fitter

8.2.2.1 Spectroscopic surveys

The 2020s will witness the advent of several large-scale spectroscopic surveys that can benefit my future research. This includes the future data releases of APOGEE (García Pérez et al., 2016), which I already used in my thesis, and also 4MOST (de Jong et al., 2019).

But the most important one for my work will be the wide-field, spectroscopic survey, WEAVE, which is expected to start observations in 2023. Its five- to seven-year programme comprises several scientific objectives, including the SCIP survey (P.I. Drew) of about 400,000 OBA stars and YSOs across the Galactic plane with the aim of studying young stars and their surroundings (Jin et al., 2023).

This survey is of particular relevance for my research. In my thesis, the main challenge in analysing the 3D kinematics of OB stars and associations was the lack of precise radial velocities, despite the availability of measurements from APOGEE DR17 (Abdurro'uf et al., 2022). Data from WEAVE will provide more RVs that will allow me to study 3D kinematics with more accuracy and obtain better constraints for the dynamics and kinematic ages of OB associations.

The SCIP survey consists of a low-resolution survey, which will notably target OBA stars alongside their counterparts at a more advanced stage of evolution, encompassing a 1200 deg^2 area across the northern Galactic plane (Jin et al., 2023). This area is shown in Fig. 8.3 and comprises the Cygnus, Auriga and Cassiopeia regions that I have studied in my thesis.

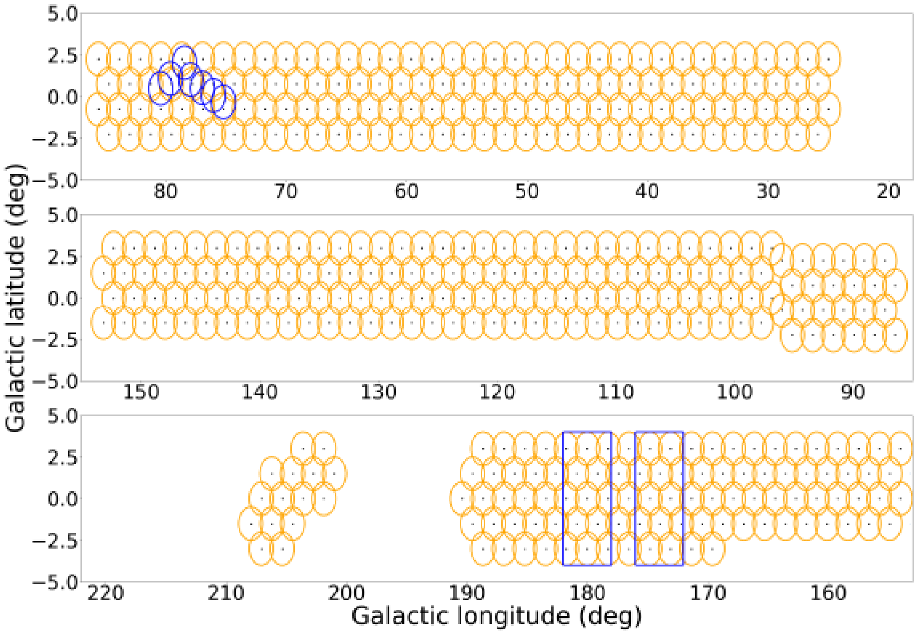


Figure 8.3: Area encompassed by the SCIP survey. The field of view from the low-resolution survey is shown in orange, while the regions covered by the high-resolution survey are represented as blue circles for the Cygnus area and as blue rectangles for the Galactic anticenter. This figure has been taken from Jin et al. (2023).

The SCIP survey will also carry out high-resolution spectroscopic surveys in two regions: the Cygnus region to analyse its prominent OB associations, following up the work from Berlanas et al. (2018a), and the Galactic anticenter to observe B- and A-type stars in order to better understand the dynamics of the Galactic disk (Jin et al., 2023). As displayed in Fig. 8.3, these two areas partially overlap with the Cygnus and Auriga regions I studied in my thesis, and therefore I should obtain high-resolution spectroscopy for some of the OB stars I have identified. Members of my newly-identified OB associations in both regions have already been included in the low-resolution and high-resolution target lists for SCIP observations.

8.2.2.2 Implementation into the SED fitting process

This wealth of spectroscopic data will offer me the opportunity to improve my SED fitter, from relying solely on photometry and astrometry to including spectroscopic information in its process. The potential relevant physical parameters are listed below:

- The spectroscopic temperature of the primary star, can be added as one of the observables to be fitted by my model. This will allow me to derive a tighter constraint on many of the model parameters. By fitting both the temperature of the primary star and the SED of the whole system could help better constrain the properties of both the primary and secondary stars.

- I currently assume a unique surface gravity value ($\log g = 4$) from the stellar atmosphere models. I showed in Chapter 3 that surface gravity influences the shape of the SED, albeit slightly, therefore including the spectroscopic surface gravity would help me model more accurate SEDs.

- I also currently assume a unique rotation rate from the stellar evolutionary models. Obtaining a projected rotational velocity from spectroscopy would allow me to derive more accurate ages for the stars, as the rotational velocity influences the evolution of stars (see e.g. Maeder & Meynet 2000).

- Spectroscopic surveys will provide RV measurements as well. Binarity influences the values of RVs and can be seen through measurements at several epochs, therefore I can exploit this information to detect which SEDs correspond to binary systems.

In the code, I expect the code to be transformed from a SED fitter to a stellar fitter, combining spectroscopy, photometry and astrometry to better constrain the physical properties of stars.

8.2.3 Hierarchical Bayesian model

My SED fitter is currently based on a classical Bayesian model, which models properties of individual stars. The application of a hierarchical Bayesian model could connect stellar populations together based on the prior information that they would have similar properties.

Such an approach could be suited for the analysis of young clusters (as it has been done in e.g. Olivares et al. 2018), whose members were typically born together and would therefore have a similar age (Lada & Lada, 2003). This might however be less applicable to OB associations, whose recent studies showed that their members could be characterized by significant age variations (see e.g. Kounkel et al. 2018; Cantat-Gaudin et al. 2019).

Bibliography

Abdurro'uf et al., 2022, ApJS, 259, 35

Abell G. O., Morrison D., Wolff S. C., 1987, Exploration of the universe. Sky and Telescope, Vol. 74, NO. 4/OCT, P.376

Abt H. A., 1973, ApJS, 26, 365

Abt H. A., 1983, ARA&A, 21, 343

Abt H. A., Gomez A. E., Levy S. G., 1990, ApJS, 74, 551

Adams F. C., Myers P. C., 2001, ApJ, 553, 744

Allen P. R., Koerner D. W., McElwain M. W., Cruz K. L., Reid I. N., 2007, AJ, 133, 971

Ambartsumian V. A., 1947, The evolution of stars and astrophysics. Armenian SSR Academy of Sciences Press

Ambartsumian V. A., 1949, AZh, 26, 3

Anderson L. D., Bania T. M., Jackson J. M., Clemens D. P., Heyer M., Simon R., Shah R. Y., Rathborne J. M., 2009, ApJS, 181, 255

Anderson L. D., Armentrout W. P., Johnstone B. M., Bania T. M., Balser D. S., Wenger T. V., Cunningham V., 2015, ApJS, 221, 26

Andrae R., et al., 2022, arXiv e-prints, p. arXiv:2206.06138

Armstrong J. J., Wright N. J., Jeffries R. D., Jackson R. J., 2020, MNRAS, 494, 4794

Arnold B., Goodwin S. P., Wright N. J., 2020, MNRAS, 495, 3474

Astraatmadja T. L., Bailer-Jones C. A. L., 2016, ApJ, 832, 137

Babusiaux C., et al., 2023, A&A, 674, A32

Bai X.-N., Ye J., Goodman J., Yuan F., 2016, ApJ, 818, 152

Bailer-Jones C. A. L., 2015, PASP, 127, 994

Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147

Ballone A., Mapelli M., Di Carlo U. N., Torniamenti S., Spera M., Rastello S., 2020, MNRAS, 496, 49

Bally J., 2016, ARA&A, 54, 491

Banyard G., Sana H., Mahy L., Bodensteiner J., Villaseñor J. I., Evans C. J., 2022, A&A, 658, A69

Barbon R., Hassan S. M., 1973, A&AS, 10, 1

Barbon R., Hassan S. M., 1996, A&AS, 115, 325

Barentsen G., et al., 2014, MNRAS, 444, 3230

Bartaya R. A., Chargeishvili K. B., Chentsov E. L., Shkhagosheva Z. U., 1994, Bulletin of the Special Astrophysics Observatory, 38, 103

Bastian N., Strader J., 2014, MNRAS, 443, 3594

Basu S., 2000, ApJL, 540, L103

Bate M. R., 2009, MNRAS, 392, 590

Bate M. R., Bonnell I. A., Bromm V., 2003, MNRAS, 339, 577

Battinelli P., 1991, A&A, 244, 69

Baumgardt H., Kroupa P., 2007, MNRAS, 380, 1589

Beck R., 2015, A&A Rv, 24, 4

Benson P. J., Myers P. C., 1989, ApJS, 71, 89

Berlanas S. R., Herrero A., Comerón F., Pasquali A., Bertelli Motta C., Sota A., 2018a, A&A, 612, A50

Berlanas S. R., Herrero A., Comerón F., Simón-Díaz S., Cerviño M., Pasquali A., 2018b, A&A, 620, A56

Berlanas S. R., Wright N. J., Herrero A., Drew J. E., Lennon D. J., 2019, MNRAS, 484, 1838

Bertoldi F., McKee C. F., 1992, ApJ, 395, 140

Binney J., Tremaine S., 1987, Galactic dynamics. Princeton University Press

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

Blaha C., Humphreys R. M., 1989, AJ, 98, 1598

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

Bochkarev N. G., Sitnik T. G., 1985, Ap&SS, 108, 237

Bodenheimer P., Burkert A., Klein R. I., Boss A. P., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. p. 675

Bohlin R. C., 2007, HST Stellar Standards with 1% Accuracy in Absolute Flux. Astronomical Society of the Pacific Conference Series Vol. 364, ASP Conference Series

Bok B. J., 1934, Harvard College Observatory Circular, 384, 1

Bolatto A. D., Leroy A. K., Rosolowsky E., Walter F., Blitz L., 2008, ApJ, 686, 948

Bonazzola S., Heyvaerts J., Falgarone E., Perault M., Puget J. L., 1987, A&A, 172, 293

Bonazzola S., Perault M., Puget J. L., Heyvaerts J., Falgarone E., Panis J. F., 1992, Journal of Fluid Mechanics, 245, 1

Bondi H., 1952, MNRAS, 112, 195

Bonnell I. A., Bate M. R., 2006, MNRAS, 370, 488

Bonnell I. A., Bate M. R., Zinnecker H., 1998, MNRAS, 298, 93

Bonnell I. A., Bate M. R., Clarke C. J., Pringle J. E., 2001, MNRAS, 323, 785

Bonnell I. A., Bate M. R., Vine S. G., 2003, MNRAS, 343, 413

Bose S., Ginsburg I., Loeb A., 2018, ApJL, 859, L13

Bossini D., et al., 2019, A&A, 623, A108

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

Bovy J., 2014, ApJ, 795, 95

Bresolin F., et al., 1998, AJ, 116, 119

Bressert E., et al., 2010, MNRAS, 409, L54

Brown A. G. A., Dekker G., de Zeeuw P. T., 1997, MNRAS, 285, 479

Brunt C. M., Mac Low M.-M., 2004, ApJ, 604, 196

Burrows A., Vartanyan D., 2021, Nature, 589, 29

Camargo D., Bonatto C., Bica E., 2009, A&A, 508, 211

Cantat-Gaudin T., Anders F., 2020, A&A, 633, A99

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

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

Cantat-Gaudin T., et al., 2020, A&A, 640, A1

Carquillat J. M., Prieur J. L., 2007, MNRAS, 380, 1064

Carrera R., et al., 2019, A&A, 623, A80

Casey A. R., 2016, ApJS, 223, 8

Cash W., Charles P., Bowyer S., Walter F., Garmire G., Riegler G., 1980, ApJL, 238, L71

Castro-Ginard A., Jordi C., Luri X., Cantat-Gaudin T., Balaguer-Núñez L., 2019, A&A, 627, A35

Cazzolato F., Pineault S., 2002, in Taylor A. R., Landecker T. L., Willis A. G., eds, Astronomical Society of the Pacific Conference Series Vol. 276, Seeing Through the Dust: The Detection of HI and the Exploration of the ISM in Galaxies. p. 312

Cazzolato F., Pineault S., 2003, AJ, 125, 2050

Cen R., 2021, ApJL, 906, L4

Cesaroni R., Galli D., Lodato G., Walmsley C. M., Zhang Q., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 197 (`arXiv:astro-ph/0603093`)

Chabrier G., 2003, ApJL, 586, L133

Chacón-Tanarro A., et al., 2019, A&A, 623, A118

Chambers K. C., et al., 2016, arXiv e-prints, p. arXiv:1612.05560

Chemel A. A., de Grijs R., Glushkova E. V., Dambis A. K., 2022, MNRAS, 515, 4359

Chen B. Q., et al., 2019a, MNRAS, 483, 4277

Chen B. Q., et al., 2019b, MNRAS, 487, 1400

Chen Y., Li H., Vogelsberger M., 2021, MNRAS, 502, 6157

Chen X., et al., 2022, arXiv e-prints, p. arXiv:2211.00810

Chentsov E. L., 2020, Astrophysical Bulletin, 75, 124

Chevance M., et al., 2020, MNRAS, 493, 2872

Chevance M., Krumholz M. R., McLeod A. F., Ostriker E. C., Rosolowsky E. W., Sternberg A., 2022, arXiv e-prints, p. arXiv:2203.09570

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

Chira R. A., Kainulainen J., Ibáñez-Mejía J. C., Henning T., Mac Low M. M., 2018, A&A, 610, A62

Choi Y. K., Hachisuka K., Reid M. J., Xu Y., Brunthaler A., Menten K. M., Dame T. M., 2014, ApJ, 790, 99

Chojnowski S. D., et al., 2017, AJ, 153, 174

Christopoulou P. E., Goudis C. D., Meaburn J., Dyson J. E., Clayton C. A., 1995, A&A, 295, 509

Churchwell E., et al., 2006, ApJ, 649, 759

Cichowolski S., Romero G. A., Ortega M. E., Cappa C. E., Vasquez J., 2009, MNRAS, 394, 900

Claeys J. S. W., de Mink S. E., Pols O. R., Eldridge J. J., Baes M., 2011, A&A, 528, A131

Clark J. S., Najarro F., Negueruela I., Ritchie B. W., Urbaneja M. A., Howarth I. D., 2012, A&A, 541, A145

Clarke S. D., Whitworth A. P., Duarte-Cabral A., Hubber D. A., 2017, MNRAS, 468, 2489

Coelho P. R. T., 2014, MNRAS, 440, 1027

Comerón F., Torra J., 2001, A&A, 375, 539

Comerón F., et al., 2002, A&A, 389, 874

Comerón F., Pasquali A., Figueras F., Torra J., 2008, A&A, 486, 453

Conrad C., et al., 2017, A&A, 600, A106

Corbelli E., et al., 2017, A&A, 601, A146

Cournoyer-Cloutier C., et al., 2023, MNRAS, 521, 1338

Creevey O. L., et al., 2023, A&A, 674, A26

Crowther P. A., 2007, ARA&A, 45, 177

Crowther P., 2012, Astronomy and Geophysics, 53, 4.30

Cutri R. M., et al., 2003, 2MASS All Sky Catalog of point sources.. NASA/IPAC Infrared Science Archive

Daffern-Powell E. C., Parker R. J., 2020, MNRAS, 493, 4925

Dale J. E., 2015, NewAR, 68, 1

Davidge T. J., 2012, ApJ, 761, 155

De Rosa R. J., et al., 2011, MNRAS, 415, 854

Delchambre L., et al., 2023, A&A, 674, A31

Delfosse X., et al., 2004, in Hilditch R. W., Hensberge H., Pavlovski K., eds, Astronomical Society of the Pacific Conference Series Vol. 318, Spectroscopically and Spatially Resolving the Components of the Close Binary Stars. pp 166–174

Dias W. S., Monteiro H., Caetano T. C., Lépine J. R. D., Assafin M., Oliveira A. F., 2014, A&A, 564, A79

Dias W. S., Monteiro H., Lépine J. R. D., Barros D. A., 2019, MNRAS, 486, 5726

Dias W. S., Monteiro H., Moitinho A., Lépine J. R. D., Carraro G., Paunzen E., Alessi B., Villela L., 2021, MNRAS, 504, 356

Dib S., Schmeja S., Parker R. J., 2018, MNRAS, 473, 849

Dicenzo B., Levesque E. M., 2019, AJ, 157, 167

Dickel H. R., Wendker H. J., 1978, A&A, 66, 289

Dieterich S. B., Henry T. J., Golimowski D. A., Krist J. E., Tanner A. M., 2012, AJ, 144, 64

Dobbs C. L., Pringle J. E., 2013, MNRAS, 432, 653

Dobbs C. L., et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 3 (`arXiv:1312.3223`), doi:10.2458/azu_uapress_9780816531240-ch001

Dobbs C. L., Bending T. J. R., Pettitt A. R., Buckner A. S. M., Bate M. R., 2022, MNRAS, 517, 675

Doherty C. L., Gil-Pons P., Siess L., Lattanzio J. C., 2017, PASA, 34, e056

Dove J. B., Shull J. M., 1994, ApJ, 430, 222

Downes D., Rinehart R., 1966, ApJ, 144, 937

Drew J. E., et al., 2005, MNRAS, 362, 753

Drew J. E., et al., 2014, MNRAS, 440, 2036

Drew J. E., Monguió M., Wright N. J., 2021, MNRAS, 508, 4952

Dubner G., Giacani E., 2015, A&A Rv, 23, 3

Duchêne G., Kraus A., 2013, ARA&A, 51, 269

Duchêne G., Simon T., Eislöffel J., Bouvier J., 2001, A&A, 379, 147

Duquennoy A., Mayor M., 1991, A&A, 248, 485

Edgar R., Clarke C., 2004, MNRAS, 349, 678

Eilers A.-C., Hogg D. W., Rix H.-W., Ness M. K., 2019, ApJ, 871, 120

Ekström S., et al., 2012, A&A, 537, A146

El-Badry K., Rix H.-W., Heintz T. M., 2021, MNRAS, 506, 2269

Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058

Elmegreen B. G., 2008, ApJ, 672, 1006

Elmegreen B. G., 2011, in Charbonnel C., Montmerle T., eds, EAS Publications Series Vol. 51, "EAS Publications Series". pp 19–30 (`arXiv:1101.3109`), doi:10.1051/eas/1151002

Elmegreen B. G., Efremov Y. N., 1997, ApJ, 480, 235

Elmegreen B. G., Hunter D. A., 2010, ApJ, 712, 604

Elmegreen B. G., Lada C. J., 1977, ApJ, 214, 725

Elmegreen B. G., Efremov Y., Pudritz R. E., Zinnecker H., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. p. 179 (`arXiv:astro-ph/9903136`)

Engargiola G., Plambeck R. L., Rosolowsky E., Blitz L., 2003, ApJS, 149, 343

Enoch M. L., et al., 2006, ApJ, 638, 293

Epstein C. R., Pinsonneault M. H., 2014, ApJ, 780, 159

Evans Neal J. I., et al., 2009, ApJS, 181, 321

Fabregat J., Capilla G., 2005, MNRAS, 358, 66

Falgarone E., Puget J. L., Perault M., 1992, A&A, 257, 715

Fall S. M., Krumholz M. R., Matzner C. D., 2010, ApJL, 710, L142

Farrell E., Groh J. H., Meynet G., Eldridge J. J., 2022, MNRAS, 512, 4116

Feast M., Whitelock P., 1997, MNRAS, 291, 683

Fehrenbach C., Burnage R., Figuiere J., 1992, A&AS, 95, 541

Ferrière K. M., 2001, Reviews of Modern Physics, 73, 1031

Fiedler R. A., Mouschovias T. C., 1993, ApJ, 415, 680

Fischera J., Martin P. G., 2012, A&A, 542, A77

Fitzpatrick E. L., 1999, PASP, 111, 63

Fitzpatrick E. L., 2004, Interstellar Extinction in the Milky Way Galaxy. Astrophysics of Dust, ASP Conference Series, Vol. 309, Proceedings of the conference held 26-30 May, 2003 in Estes Park, Colorado. Edited by Adolf N. Witt, Geoffrey C. Clayton and Bruce T. Draine., p. 33

Fitzpatrick E. L., Massa D., Gordon K. D., Bohlin R., Clayton G. C., 2019, ApJ, 886, 108

Fletcher A., Beck R., Shukurov A., Berkhuijsen E. M., Horellou C., 2011, MNRAS, 412, 2396

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Foster T., Brunt C. M., 2015, AJ, 150, 147

Frinchaboy P. M., Majewski S. R., 2008, AJ, 136, 118

Froebrich D., Rowles J., 2010, in Astronomical Society of India Conference Series. p. 1

Fuchs B., Breitschwerdt D., de Avillez M. A., Dettbarn C., Flynn C., 2006, MNRAS, 373, 993

Fujii M. S., Saitoh T. R., Portegies Zwart S. F., 2012, ApJ, 753, 85

GRAVITY Collaboration et al., 2022, A&A, 657, L12

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

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

Gaia Collaboration et al., 2021, A&A, 649, A1

Gaia Collaboration et al., 2023a, A&A, 674, A1

Gaia Collaboration et al., 2023b, A&A, 674, A39

García Pérez A. E., et al., 2016, AJ, 151, 144

Garmany C. D., Stencel R. E., 1992, A&AS, 94, 211

Geen S., Rosdahl J., Blaizot J., Devriendt J., Slyz A., 2015, MNRAS, 448, 3248

Georgelin Y. M., Georgelin Y. P., Roux S., 1973, A&A, 25, 337

Girichidis P., et al., 2020, SSRv, 216, 68

Glover S. C. O., Mac Low M.-M., 2007, ApJ, 659, 1317

Goldreich P., Lynden-Bell D., 1965, MNRAS, 130, 97

Goldsmith P. F., Li D., 2005, ApJ, 622, 938

Gómez G. C., Vázquez-Semadeni E., 2014, ApJ, 791, 124

Gontcharov G. A., 2006, Astronomy Letters, 32, 759

Goodwin S. P., 1997, MNRAS, 284, 785

Goodwin S. P., Kroupa P., 2005, A&A, 439, 565

Goodwin S. P., Whitworth A. P., 2004, A&A, 413, 929

Gorski K. M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, arXiv e-prints, pp astro–ph/9905275

Gouliermis D. A., 2011, PhyS, 84, 048401

Gouliermis D. A., 2018a, PASP, 130, 072001

Gouliermis D. A., 2018b, PASP, 130, 072001

Gouliermis D., Kontizas M., Korakitis R., Morgan D. H., Kontizas E., Dapergolas A., 2000, AJ, 119, 1737

Gouliermis D., Kontizas M., Kontizas E., Korakitis R., 2003, A&A, 405, 111

Gratton R., Bragaglia A., Carretta E., D'Orazi V., Lucatello S., Sollima A., 2019, A&A Rv, 27, 8

Green G. M., et al., 2018, MNRAS, 478, 651

Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, ApJ, 887, 93

Grenier S., et al., 1999, A&AS, 137, 451

Grieves N., et al., 2018, MNRAS, 481, 3244

Groot P. J., et al., 2009, MNRAS, 399, 323

Großschedl J. E., Alves J., Meingast S., Herbst-Kiss G., 2021, A&A, 647, A91

Gutermuth R. A., Megeath S. T., Myers P. C., Allen L. E., Pipher J. L., Fazio G. G., 2009, ApJS, 184, 18

Gyulbudaghian A. L., 2011, Astrophysics, 54, 476

Hancock M., Smith B. J., Giroux M. L., Struck C., 2008, MNRAS, 389, 1470

Hansen C. E., Klein R. I., McKee C. F., Fisher R. T., 2012, ApJ, 747, 22

Hanson M. M., 2003, ApJ, 597, 957

Harries T. J., Hilditch R. W., Hill G., 1998, MNRAS, 295, 386

Harris W. E., 1976, AJ, 81, 1095

Harris W. E., 1991, ARA&A, 29, 543

Hartmann L., 2002, ApJ, 578, 914

Hayashi C., 1961, PASJ, 13, 450

Hayford P., 1932, Lick Observatory Bulletin, 448, 53

Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, ApJ, 591, 288

Heithausen A., Bensch F., Stutzki J., Falgarone E., Panis J. F., 1998, A&A, 331, L65

Heitsch F., Slyz A. D., Devriendt J. E. G., Hartmann L. W., Burkert A., 2006, ApJ, 648, 1052

Henyey L. G., Lelevier R., Levée R. D., 1955, PASP, 67, 154

Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454

Heyer M., Dame T. M., 2015, ARA&A, 53, 583

Heyer M. H., Carpenter J. M., Snell R. L., 2001, ApJ, 551, 852

Higdon J. C., Lingenfelter R. E., 2013, ApJ, 775, 110

Hillenbrand L. A., Hartmann L. W., 1998, ApJ, 492, 540

Hills J. G., 1980, ApJ, 235, 986

Hodge P. W., 1981, Atlas of the Andromeda galaxy. University of Washington Press

Hodge P. W., Wright F. W., 1967, The Large Magellanic Cloud. Washington, Smithsonian Press

Hodge P. W., Wright F. W., 1977, The Small Magellanic Cloud. University of Washington Press

Hodge P. W., Skelton B. P., Ashizawa J., 2002, An Atlas of Local Group Galaxies. Springer Vol. 221, Kluwer Academic Publishers, doi:10.1007/978-0-306-47818-5

Hodgkin S. T., Irwin M. J., Hewett P. C., Warren S. J., 2009, MNRAS, 394, 675

Holgado G., et al., 2018, A&A, 613, A65

Holmberg J., Flynn C., 2004, MNRAS, 352, 440

Hora J. L., et al., 2008, PASP, 120, 1233

Hou L. G., Han J. L., Shi W. B., 2009, A&A, 499, 473

Hoyle F., Lyttleton R. A., 1939, Proceedings of the Cambridge Philosophical Society, 35, 405

Huang W., Gies D. R., 2006, ApJ, 648, 580

Huang W., Gies D. R., McSwain M. V., 2010, ApJ, 722, 605

Humphreys R. M., 1978, ApJS, 38, 309

Humphreys R. M., McElroy D. B., 1984, ApJ, 284, 565

Hunt E. L., Reffert S., 2023, A&A, 673, A114

Hunter D., 1997, PASP, 109, 937

Ivanov G. R., 1987, Ap&SS, 136, 113

Ivezić Ž., Connelly A. J., Vand erPlas J. T., Gray A., 2014, Statistics, Data Mining, and Machine Learning in Astronomy. Princeton University Press, Princeton University, New Jersey

Jackson J. M., Finn S. C., Chambers E. T., Rathborne J. M., Simon R., 2010, ApJL, 719, L185

Jacobson H. R., Cummings J., Deliyannis C. P., Steinhauer A., Sarajedini A., 2002, in American Astronomical Society Meeting Abstracts. p. 124.07

Janes K. A., Phelps R. L., 1994, AJ, 108, 1773

Janka H.-T., 2012, Annual Review of Nuclear and Particle Science, 62, 407

Jeans J. H., 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jeffries R. D., Naylor T., Mayne N. J., Bell C. P. M., Littlefair S. P., 2013, MNRAS, 434, 2438

Jin S., et al., 2023, MNRAS,

Johnson H. L., Morgan W. W., 1954, ApJ, 119, 344

Jones S., et al., 2013, ApJ, 772, 150

Jönsson H., et al., 2020, AJ, 160, 120

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

Jose J., et al., 2008, MNRAS, 384, 1675

Joshi Y. C., Maurya J., John A. A., Panchal A., Joshi S., Kumar B., 2020, MNRAS, 492, 3602

Kalloghlian A. T., 2009, Astrophysics, 52, 157

Kang J.-h., Koo B.-C., Salter C., 2012, AJ, 143, 75

Kaur H., Sharma S., Dewangan L. K., Ojha D. K., Durgapal A., Panwar N., 2020, ApJ, 896, 29

Kawamura A., Onishi T., Yonekura Y., Dobashi K., Mizuno A., Ogawa H., Fukui Y., 1998, ApJS, 117, 387

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

Kerr R. M. P., Rizzuto A. C., Kraus A. L., Offner S. S. R., 2021, ApJ, 917, 23

Kharchenko N. V., Piskunov A. E., Röser S., Schilbach E., Scholz R. D., 2005, A&A, 438, 1163

Kharchenko N. V., Piskunov A. E., Schilbach E., Röser S., Scholz R. D., 2013, A&A, 558, A53

Kim C.-G., Ostriker E. C., 2015, ApJ, 802, 99

Kiminki D. C., Kobulnicky H. A., 2012, ApJ, 751, 4

Kiminki D. C., et al., 2007, ApJ, 664, 1102

Kiminki D. C., et al., 2008, ApJ, 681, 735

Kirsanova M. S., Sobolev A. M., Thomasson M., Wiebe D. S., Johansson L. E. B., Seleznev A. F., 2008, MNRAS, 388, 729

Klassen M., Pudritz R. E., Kuiper R., Peters T., Banerjee R., 2016, ApJ, 823, 28

Knödlseder J., 2000, A&A, 360, 539

Koda J., et al., 2009, ApJL, 700, L132

Kolmogorov A., 1941, Akademiia Nauk SSSR Doklady, 30, 301

Kounkel M., 2020, ApJ, 902, 122

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

Kouwenhoven M. B. N., Brown A. G. A., Zinnecker H., Kaper L., Portegies Zwart S. F., 2005, A&A, 430, 137

Kouwenhoven M. B. N., Brown A. G. A., Portegies Zwart S. F., Kaper L., 2007, A&A, 474, 77

Krasnopolsky R., Königl A., 2002, ApJ, 580, 987

Kroupa P., 2001, MNRAS, 322, 231

Kroupa P., 2002, Science, 295, 82

Kruijssen J. M. D., 2011, in Stellar Clusters & Associations: A RIA Workshop on Gaia. pp 137–141 (`arXiv:1107.2114`)

Kruijssen J. M. D., 2012, MNRAS, 426, 3008

Kruijssen J. M. D., Maschberger T., Moeckel N., Clarke C. J., Bastian N., Bonnell I. A., 2012, MNRAS, 419, 841

Krumholz M. R., 2014, PhR, 539, 49

Krumholz M. R., Matzner C. D., 2009, ApJ, 703, 1352

Krumholz M. R., McKee C. F., Klein R. I., 2005, Nature, 438, 332

Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, ARA&A, 57, 227

Kudritzki R. P., 2002, ApJ, 577, 389

Kuhn M. A., Hillenbrand L. A., Sills A., Feigelson E. D., Getman K. V., 2019, ApJ, 870, 32

Kuiper R., Klahr H., Beuther H., Henning T., 2011, ApJ, 732, 20

Kun M., 2008, in Reipurth B., ed., , Vol. 4, Handbook of Star Forming Regions, Volume I. Handbook of Star Forming Regions, Volume I: The Northern Sky ASP Monograph Publications, Vol. 4. Edited by Bo Reipurth, p.240, p. 240

Kusuno K., Asaki Y., Imai H., Oyama T., 2013, ApJ, 774, 107

Lada C. J., 2006, ApJL, 640, L63

Lada C. J., Lada E. A., 1991, in Janes K., ed., Astronomical Society of the Pacific Conference Series Vol. 13, The Formation and Evolution of Star Clusters. pp 3–22

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

Lada C. J., Margulis M., Dearborn D., 1984, ApJ, 285, 141

Lada C. J., Lombardi M., Alves J. F., 2010, ApJ, 724, 687

Lallement R., Babusiaux C., Vergely J. L., Katz D., Arenou F., Valette B., Hottier C., Capitanio L., 2019, A&A, 625, A135

Langer N., 2012, ARA&A, 50, 107

Langer W. D., van Dishoeck E. F., Bergin E. A., Blake G. A., Tielens A. G. G. M., Velusamy T., Whittet D. C. B., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. p. 29

Larson R. B., 1969, MNRAS, 145, 271

Larson R. B., 1981, MNRAS, 194, 809

Larson R. B., 2003, Reports on Progress in Physics, 66, 1651

Le Duigou J. M., Knödlseder J., 2002, A&A, 392, 869

Ledrew G., 2001, JRASC, 95, 32

Lee C. W., Myers P. C., Tafalla M., 1999, ApJ, 526, 788

Leitherer C., Ekström S., Meynet G., Schaerer D., Agienko K. B., Levesque E. M., 2014, ApJS, 212, 14

Lequeux J., 2005, The Interstellar Medium. The interstellar medium, Translation from the French language edition of: Le Milieu Interstellaire by James Lequeux, EDP Sciences, 2003 Edited by J. Lequeux. Astronomy and astrophysics library, Berlin: Springer, 2005, doi:10.1007/b137959

Levesque E. M., Massey P., Olsen K. A. G., Plez B., Josselin E., Maeder A., Meynet G., 2005, ApJ, 628, 973

Li S., Frank A., Blackman E. G., 2014, MNRAS, 444, 2884

Lim B., Sung H., Kim J. S., Bessell M. S., Park B.-G., 2014, MNRAS, 443, 454

Lim B., et al., 2018, MNRAS, 477, 1993

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

Lindegren L., et al., 2021a, A&A, 649, A2

Lindegren L., et al., 2021b, A&A, 649, A4

Liu Z., Cui W., Liu C., Huang Y., Zhao G., Zhang B., 2019, ApJS, 241, 32

Lozinskaia T. A., Sitnik T. G., Lomovskii A. I., 1986, Ap&SS, 121, 357

Lozinskaya T. A., 1998, Astronomy Letters, 24, 237

Lozinskaya T. A., Sitnik T. G., 1988, Pisma v Astronomicheskii Zhurnal, 14, 240

Lucas P. W., et al., 2008, MNRAS, 391, 136

Maeder A., Meynet G., 2000, ARA&A, 38, 143

Maeder A., Meynet G., 2010, NewAR, 54, 32

Magnani L., Blitz L., Mundy L., 1985, ApJ, 295, 402

Maíz Apellániz J., Pantaleoni González M., Barbá R. H., Simón-Díaz S., Negueruela I., Lennon D. J., Sota A., Trigueros Páez E., 2018, A&A, 616, A149

Maíz Apellániz J., Pantaleoni González M., Barbá R. H., 2021, A&A, 649, A13

Majewski S. R., et al., 2017, AJ, 154, 94

Marco A., Negueruela I., 2016, MNRAS, 459, 880

Marco A., Bernabeu G., Negueruela I., 2001, AJ, 121, 2075

Martin N., 1972, A&A, 17, 253

Martins F., Palacios A., 2013, A&A, 560, A16

Martins F., Schaerer D., Hillier D. J., 2005, A&A, 436, 1049

Maschberger T., 2013, MNRAS, 429, 1725

Mason B. D., Hartkopf W. I., Gies D. R., Henry T. J., Helsel J. W., 2009, AJ, 137, 3358

Massey P., Thompson A. B., 1991, AJ, 101, 1408

Massey P., Johnson K. E., Degioia-Eastwood K., 1995, ApJ, 454, 151

Massey P., DeGioia-Eastwood K., Waterhouse E., 2001, AJ, 121, 1050

Matzner C. D., 2002, ApJ, 566, 302

Matzner C. D., 2007, ApJ, 659, 1394

Mayer P., Drechsel H., Harmanec P., Yang S., Šlechta M., 2013, A&A, 559, A22

McCray R., Snow T. P. J., 1979, ARA&A, 17, 213

McInnes L., Healy J., Astels S., 2017, The Journal of Open Source Software, 2, 205

McKee C. F., 1989, ApJ, 345, 782

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

McKee C. F., Tan J. C., 2003, ApJ, 585, 850

McKee C. F., Williams J. P., 1997, ApJ, 476, 144

Mel'Nik A. M., Dambis A. K., 2009, MNRAS, 400, 518

Mel'Nik A. M., Efremov Y. N., 1995, Astronomy Letters, 21, 10

Mellinger A., 2008, in Reipurth B., ed., , Vol. 4, Handbook of Star Forming Regions, Volume I. ASP Monograph Publications, p. 1

Mel'nik A. M., Dambis A. K., 2017, MNRAS, 472, 3887

Melnik A. M., Dambis A. K., 2020, MNRAS, 493, 2339

Melnik A. M., Dambis A. K., Glushkova E. V., Rautiainen P., 2020, in Elmegreen B. G., Tóth L. V., Güdel M., eds, IAU Symposium Vol. 345, IAU Symposium. pp 39–42, doi:10.1017/S1743921318008335

Miret-Roig N., Galli P. A. B., Olivares J., Bouy H., Alves J., Barrado D., 2022, A&A, 667, A163

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

Mohr-Smith M., et al., 2015, MNRAS, 450, 3855

Monguió M., et al., 2020, A&A, 638, A18

Mooney T. J., Solomon P. M., 1988, ApJL, 334, L51

Moór A., Kiss C., 2003, Commmunications of the Konkoly Observatory Hungary, 103, 149

Morgan W. W., Whitford A. E., Code A. D., 1953, ApJ, 118, 318

Münch L., Morgan W. W., 1953, ApJ, 118, 161

Murray N., 2011, ApJ, 729, 133

Murray N., Quataert E., Thompson T. A., 2010, ApJ, 709, 191

Myers P. C., Lazarian A., 1998, ApJL, 507, L157

Nagayama T., et al., 2015, PASJ, 67, 66

Nakamura F., Li Z.-Y., 2007, ApJ, 662, 395

Narita S., Hayashi C., Miyama S. M., 1984, Progress of Theoretical Physics, 72, 1118

Negueruela I., Marco A., 2003, A&A, 406, 119

Negueruela I., Marco A., Israel G. L., Bernabeu G., 2007, A&A, 471, 485

Nieva M.-F., Przybilla N., 2014, A&A, 566, A7

Nilakshi Sagar R., Pandey A. K., Mohan V., 2002, A&A, 383, 153

Norman C., Silk J., 1980, ApJ, 238, 158

Nürnberger D. E. A., Petr-Gotzens M. G., 2002, A&A, 382, 537

Offner S. S. R., Chaban J., 2017, ApJ, 847, 104

Offner S. S. R., McKee C. F., 2011, ApJ, 736, 53

Offner S. S. R., Clark P. C., Hennebelle P., Bastian N., Bate M. R., Hopkins P. F., Moraux E., Whitworth A. P., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 53 (`arXiv:1312.5326`), doi:10.2458/azu_uapress_9780816531240-ch003

Oka T., Hasegawa T., Sato F., Tsuboi M., Miyazaki A., Sugimoto M., 2001, ApJ, 562, 348

Olivares J., et al., 2018, A&A, 617, A15

Oort J. H., 1954, BAN, 12, 177

Orellana R. B., De Biasi M. S., Paíz L. G., 2021, MNRAS, 502, 6080

Osterbrock D. E., Ferland G. J., 2006, Astrophysics of gaseous nebulae and active galactic nuclei. Astrophysics of gaseous nebulae and active galactic nuclei, 2nd. ed. by D.E. Osterbrock and G.J. Ferland. Sausalito, CA: University Science Books

Ostriker J., 1964, ApJ, 140, 1056

Paladini R., Burigana C., Davies R. D., Maino D., Bersanelli M., Cappellini B., Platania P., Smoot G., 2003, A&A, 397, 213

Pandey A. K., Bhatt B. C., Mahra H. S., Sagar R., 1989, MNRAS, 236, 263

Pandey A. K., Upadhyay K., Ogura K., Sagar R., Mohan V., Mito H., Bhatt H. C., Bhatt B. C., 2005, MNRAS, 358, 1290

Pandey A. K., Sharma S., Upadhyay K., Ogura K., Sandhu T. S., Mito H., Sagar R., 2007, PASJ, 59, 547

Pandey A. K., et al., 2013, ApJ, 764, 172

Pandey A. K., Sharma S., Kobayashi N., Sarugaku Y., Ogura K., 2020, MNRAS, 492, 2446

Pantaleoni González M., Maíz Apellániz J., Barbá R. H., Reed B. C., 2021, MNRAS, 504, 2968

Parker R. J., Wright N. J., Goodwin S. P., Meyer M. R., 2014, MNRAS, 438, 620

Parker R. J., Goodwin S. P., Wright N. J., Meyer M. R., Quanz S. P., 2016, MNRAS, 459, L119

Payne C. H., 1925, PhD thesis, RADCLIFFE COLLEGE.

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

Pecaut M. J., Mamajek E. E., 2016, MNRAS, 461, 794

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

Pigulski A., Kopacki G., Kołaczkowski Z., 2001, A&A, 376, 144

Piskunov A. E., Belikov A. N., Kharchenko N. V., Sagar R., Subramaniam A., 2004, MNRAS, 349, 1449

Piskunov A. E., Schilbach E., Kharchenko N. V., Röser S., Scholz R. D., 2008, A&A, 477, 165

Plume R., Jaffe D. T., Evans Neal J. I., Martín-Pintado J., Gómez-González J., 1997, ApJ, 476, 730

Poelarends A. J. T., Herwig F., Langer N., Heger A., 2008, ApJ, 675, 614

Polychroni D., et al., 2013, ApJL, 777, L33

Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, ARA&A, 48, 431

Preibisch T., Balega Y., Hofmann K.-H., Weigelt G., Zinnecker H., 1999, NewA, 4, 531

Preibisch T., Weigelt G., Zinnecker H., 2001, in Zinnecker H., Mathieu R., eds, Proceedings of IAU Symp. 200 Vol. 200, The Formation of Binary Stars. p. 69 (`arXiv:astro-ph/0008014`), doi:10.48550/arXiv.astro-ph/0008014

Prialnik D., 2000, An Introduction to the Theory of Stellar Structure and Evolution. Cambridge University Press

Quintana A. L., Wright N. J., 2021, MNRAS, 508, 2370

Quintana A. L., Wright N. J., 2022, MNRAS, 515, 687

Quintana A. L., Wright N. J., Jeffries R. D., 2023, MNRAS, 522, 3124

Raghavan D., et al., 2010, ApJS, 190, 1

Rauch T., Deetjen J. L., 2003, Handling of Atomic Data. Stellar Atmosphere Modeling, ASP Conference Proceedings, Vol. 288. in Tuebingen, Germany. Editors: Ivan Hubeny, Dimitri Mihalas, and Klaus Werner. San Francisco: Astronomical Society of the Pacific, p. 103

Reid I. N., Gizis J. E., 1997, AJ, 113, 2246

Reid M. J., et al., 2019, ApJ, 885, 131

Reipurth B., Bally J., 2001, ARA&A, 39, 403

Reipurth B., Schneider N., 2008, Star Formation and Young Clusters in Cygnus. The Northern Sky ASP Monograph Publications, University of Hawaii, p. 36

Riello M., et al., 2021, A&A, 649, A3

Roberts William W. J., 1972, ApJ, 173, 259

Robitaille J. F., Scaife A. M. M., Carretti E., Haverkorn M., Crocker R. M., Kesteven M. J., Poppi S., Staveley-Smith L., 2018, A&A, 617, A101

Roman-Lopes A., Roman-Lopes G. F., 2019, MNRAS, 484, 5578

Ruffert M., Arnett D., 1994, ApJ, 427, 351

Ruiz-Lapuente P., et al., 2019, ApJ, 870, 135

Russeil D., 2003, A&A, 397, 133

Rygl K. L. J., et al., 2012, in Proceedings of the 11th European VLBI Network Symposium & Users Meeting. 9-12 October. p. 40

Sahu M. S., 1992, PhD thesis, University of Groningen, Kapteyn Astronomical Institute

Saigo K., Hanawa T., 1998, ApJ, 493, 342

Sale S. E., et al., 2014, MNRAS, 443, 2907

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

Sana H., 2017, in Eldridge J. J., Bray J. C., McClelland L. A. S., Xiao L., eds, Proceedings of the International Astronomical Union, IAU Symposium Vol. 329, The Lives and Death-Throes of Massive Stars. pp 110–117 (`arXiv:1703.01608`), doi:10.1017/S1743921317003209

Sana H., Evans C. J., 2011, in Neiner C., Wade G., Meynet G., Peters G., eds, Proceedings of the International Astronomical Union, IAU Symposium Vol. 272, Active OB Stars: Structure, Evolution, Mass Loss, and Critical Limits. pp 474–485 (`arXiv:1009.4197`), doi:10.1017/S1743921311011124

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

Sander A., Hamann W. R., Todt H., 2012, A&A, 540, A144

Sanders D. B., Scoville N. Z., Solomon P. M., 1985, ApJ, 289, 373

Santos-Silva T., et al., 2021, MNRAS, 508, 1033

Schinnerer E., et al., 2019, ApJ, 887, 49

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schneider N., Bontemps S., Simon R., Jakob H., Motte F., Miller M., Kramer C., Stutzki J., 2006, A&A, 458, 855

Schneider N., Simon R., Bontemps S., Comerón F., Motte F., 2007, A&A, 474, 873

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

Scoville N. Z., Hersh K., 1979, ApJ, 229, 578

Scoville N. Z., Wilson C. D., 2004, in Lamers H. J. G. L. M., Smith L. J., Nota A., eds, Astronomical Society of the Pacific Conference Series Vol. 322, The Formation and Evolution of Massive Young Star Clusters. p. 245

Seifried D., Walch S., 2015, MNRAS, 452, 2410

Sharma S., Pandey A. K., Ojha D. K., Chen W. P., Ghosh S. K., Bhatt B. C., Maheswar G., Sagar R., 2007, MNRAS, 380, 1141

Shatsky N., Tokovinin A., 2002, A&A, 382, 92

Shu F. H., Allen A., Shang H., Ostriker E. C., Li Z.-Y., 1999, in Lada C. J., Kylafis N. D., eds, NATO Advanced Study Institute (ASI) Series C Vol. 540, The Origin of Stars and Planetary Systems. p. 193

Shull J. M., Saken J. M., 1995, ApJ, 444, 663

Sitnik T. G., Egorov O. V., Lozinskaya T. A., Moiseev A. V., Tatarnikov A. M., Vozyakova O. V., Wiebe D. S., 2019, MNRAS, 486, 2449

Sitnik T. G., Rastorguev A. S., Tatarnikova A. A., Tatarnikov A. M., Egorov O. V., Tatarnikov A. A., 2020, MNRAS, 498, 5437

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

Smith M., 2016, PhD thesis, University of Hertfordshire

Solomon P. M., Rivolo A. R., Barrett J., Yahil A., 1987, ApJ, 319, 730

Sota A., Maíz Apellániz J., Walborn N. R., Alfaro E. J., Barbá R. H., Morrell N. I., Gamen R. C., Arias J. I., 2011, ApJS, 193, 24

Sparke L. S., Gallagher John S. I., 2000, Galaxies in the universe : an introduction. Cambridge University Press

Spitzer Lyman J., 1958, ApJ, 127, 17

Spitzer L., 1978, Physical processes in the interstellar medium. A Wiley-Interscience Publication, New York, doi:10.1002/9783527617722

Spitzer L., 1987, Dynamical evolution of globular clusters. Princeton University Press

Squicciarini V., Gratton R., Bonavita M., Mesa D., 2021, MNRAS, 507, 1381

Stahler S. W., Shu F. H., Taam R. E., 1980, ApJ, 241, 637

Stahler S. W., Korycansky D. G., Brothers M. J., Touma J., 1994, ApJ, 431, 341

Stahler S. W., Palla F., Ho P. T. P., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. pp 327–352

Stark A. A., Blitz L., 1978, ApJL, 225, L15

Sternberg A., Le Petit F., Roueff E., Le Bourlot J., 2014, ApJ, 790, 10

Straižys V., Lazauskaitė R., 2009, Baltic Astronomy, 18, 19

Straižys V., Drew J. E., Laugalys V., 2010, Baltic Astronomy, 19, 169

Stroe A., Sobral D., Paulino-Afonso A., Alegre L., Calhau J., Santos S., van Weeren R., 2017, MNRAS, 465, 2916

Suad L. A., Cichowolski S., Noriega-Crespo A., Arnal E. M., Testori J. C., Flagey N., 2016, A&A, 585, A154

Subramaniam A., Sagar R., 1999, AJ, 117, 937

Tan J. C., Krumholz M. R., McKee C. F., 2006, ApJL, 641, L121

Tang X. D., et al., 2021, A&A, 655, A12

Tapia M., Costero R., Echevarria J., Roth M., 1991, MNRAS, 253, 649

Tauris T. M., van den Heuvel E. P. J., 2006, in , Vol. 39, Compact stellar X-ray sources. Cambridge University Press, pp 623–665

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., Maciejewski G., 2010, MNRAS, 402, 2369

Tovmassian H. M., Hovhannessian R. K., Epremian R. A., Huguenin D., 1994, MNRAS, 266, 337

Trundle C., Dufton P. L., Hunter I., Evans C. J., Lennon D. J., Smartt S. J., Ryans R. S. I., 2007, A&A, 471, 625

Turner D. G., Rosvick J. M., Balam D. D., Henden A. A., Majaess D. J., Lane D. J., 2011, PASP, 123, 1249

Urquhart J. S., et al., 2018, MNRAS, 473, 1059

Usenko I. A., Kovtyukh V. V., Klochkova V. G., Panchuk V. E., Yermakov S. V., 2001, A&A, 367, 831

Uyanıker B., Fürst E., Reich W., Aschenbach B., Wielebinski R., 2001, A&A, 371, 675

Vallée J. P., 2014, AJ, 148, 5

Vázquez-Semadeni E., González-Samaniego A., Colín P., 2017, MNRAS, 467, 1313

Vázquez R. A., May J., Carraro G., Bronfman L., Moitinho A., Baume G., 2008, ApJ, 672, 930

Velasco S., García M., Negueruela I., 2013, in Guirado J. C., Lara L. M., Quilis V., Gorgas J., eds, Highlights of Spanish Astrophysics VII. pp 675–675

Verbeek K., et al., 2012, MNRAS, 420, 1115

Vidotto A. A., et al., 2014, MNRAS, 441, 2361

Vink J. S., de Koter A., Lamers H. J. G. L. M., 2000, A&A, 362, 295

Vink J. S., de Koter A., Lamers H. J. G. L. M., 2001, A&A, 369, 574

Vink J., Patnaude D. J., Castro D., 2022, ApJ, 929, 57

Walborn N. R., 1971, ApJL, 164, L67

Walch S., Naab T., 2015, MNRAS, 451, 2757

Wang P., Li Z.-Y., Abel T., Nakamura F., 2010, ApJ, 709, 27

Ward J. L., Kruijssen J. M. D., 2018, MNRAS, 475, 5659

Ward-Thompson D., André P., Crutcher R., Johnstone D., Onishi T., Wilson C., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 33 (`arXiv:astro-ph/0603474`)

Werk J. K., Putman M. E., Meurer G. R., Oey M. S., Ryan-Weber E. V., Kennicutt R. C. J., Freeman K. C., 2008, ApJ, 678, 888

Werner K., Dreizler S., 1999, Journal of Computational and Applied Mathematics, 109, 65

Werner K., Deetjen J. L., Dreizler S., Nagel T., Rauch T., Schuh S. L., 2003, Model Photospheres with Accelerated Lambda Iteration. Stellar Atmosphere Modeling, ASP Conference Proceedings, Vol. 288. Editors: Ivan Hubeny, Dimitri Mihalas, and Klaus Werner. San Francisco: Astronomical Society of the Pacific, p. 31

Whitworth A., 1979, MNRAS, 186, 59

Wilson C. D., 1991, AJ, 101, 1663

Woosley S. E., Bloom J. S., 2006, ARA&A, 44, 507

Wright N. J., 2020, NewAR, 90, 101549

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

Wright N. J., Parker R. J., Goodwin S. P., Drake J. J., 2014, MNRAS, 438, 639

Wright N. J., Drew J. E., Mohr-Smith M., 2015, MNRAS, 449, 741

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

Wright N. J., et al., 2019, MNRAS, 486, 2477

Wright N. J., Kounkel M., Zari E., Goodwin S., Jeffries R. D., 2023, in Inutsuka S., Aikawa Y., Muto T., Tomida K., Tamura M., eds, Astronomical Society of the Pacific Conference Series Vol. 534, Astronomical Society of the Pacific Conference Series. p. 129

Wu Z.-Y., Zhou X., Ma J., Du C.-H., 2009, MNRAS, 399, 2146

Yoon S.-J., Lee Y.-W., 2002, Science, 297, 578

Zari E., Rix H. W., Frankel N., Xiang M., Poggio E., Drimmel R., Tkachenko A., 2021, A&A, 650, A112

Zhang Q., 2005, in Cesaroni R., Felli M., Churchwell E., Walmsley M., eds, Vol. 227, Massive Star Birth: A Crossroads of Astrophysics. pp 135–144, doi:10.1017/S174392130500445X

Zhong J., Chen L., Wu D., Li L., Bai L., Hou J., 2020, A&A, 640, A127

Zinnecker H., 1982, Annals of the New York Academy of Sciences, 395, 226

Zinnecker H., Yorke H. W., 2007, ARA&A, 45, 481

de Geus E. J., 1991, in Janes K., ed., Astronomical Society of the Pacific Conference Series Vol. 13, The Formation and Evolution of Star Clusters. pp 40–54

de Jong R. S., et al., 2019, The Messenger, 175, 3

de Rossi M. E., Tissera P. B., Pedrosa S. E., 2010, A&A, 519, A89

de Zeeuw P. T., Hoogerwerf R., de Bruijne J. H. J., Brown A. G. A., Blaauw A., 1999, AJ, 117, 354 de la Fuente Marcos R., de la Fuente Marcos C., 2009, NewA, 14, 180

van der Hucht K. A., 2001, NewAR, 45, 135

A List of publications

A.1 Peer-reviewed journal papers

Quintana, Alexis L., Wright, Nicholas J. and Jeffries, Robin D., 2023, *Mapping the distribution of OB stars and associations in Auriga*, MNRAS, 522, 3124.

Quintana, Alexis L. & Wright, Nicholas J., 2022, *Large-scale expansion of OB stars in Cygnus*, MNRAS, 515, 687.

Quintana, Alexis L. & Wright, Nicholas J., 2021, *Revisiting the Cygnus OB associations*, MNRAS, 508, 2370.