The clustering environment of Herbig Ae/Be stars: A Gaia view

Alice Gabriela Yumiry Pérez Blanco

School of Physics and Astronomy

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

September 2023

'Un astrónomo necesita más imaginación poética que sentido común, porque la magnífica complejidad del universo no puede medirse ni explicarse, solo puede intuirse'

— *El cuaderno de Maya,* Isabel Allende.

Declaration

I hereby declare that the contents of this dissertation are my own. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. My right to be identified as author of this work has been asserted by me in accordance with the Copyright, Designs and Patents Act 1988.

Alice Gabriela Yumiry Pérez Blanco
2023

Preface

The work conducted for this PhD thesis project led, to greater or lesser extents, to the following publications:

Non-refereed publications

1. **Pérez Blanco, A.**, Oudmaijer, R. D., Pérez-Martínez, R., and Baines, D. (2020). *Clustering properties of Herbig Ae/Be stars - CEREAL*. Contributions to the XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society, id. 176.
2. **Pérez Blanco, A.**, Oudmaijer, R. D., Pérez-Martínez, R., and Baines, D. (2019). *Clustering properties of Herbig Ae/Be stars - CEREAL* 53rd ESLAB Symposium: The Gaia Universe, id. 19.
3. **Pérez Blanco, A.**, Oudmaijer, R. D., Pérez-Martínez, R., and Baines, D. (2019). *Clustering properties of Herbig Ae/Be stars - CEREAL* Proceedings of the XIII Scientific Meeting of the Spanish Astronomical Society, p. 427-427.
4. **Pérez Blanco, A.**, Oudmaijer, R. D., and Baines, D. (2017). *The clustering properties of intermediate mass young stars* Memorie della Societa Astronomica Italiana, v.88, p.838.

Regarding unpublished work, appropriate credit has been given in the text where reference has been made to the work of others.

Acknowledgements

Doing a PhD was a wild adventure, which started when I made the decision to do a PhD abroad. This process has involved several steps including preparing the application, preparing for the interview (which almost happened at 6am, Venezuelan time) and waiting for the answer (which was a few very stressful days). My adventure then began on March 11th 2016 when I received the email with my PhD offer. Beginning that day, I started the frenzy of paperwork needed to finish my Master's degree, get the correct visa to move to the UK, find a house and the so many other things that took a lot of time and effort.

Saying goodbye to my family was one of the hardest things I ever had to do; they have been my strongest emotional support all my life. My family always occupy a big part of my heart and my mind; they are the first people with whom I share the happiest and also the most difficult moments in my life. I have learnt that it does not matter if we are in different parts of the world, they are and always will be present in my life. *Los amo muchisimo y los llevo en mi corazón a cada momento*.

Without the STARRY project (my Marie Curie fellowship), Patricia Grant and Professor René Oudmaijer I would not be here. Firstly, I want to give a large thank you to *Trish* for being there for all of us, at a moments notice, to support and guide us with all the "bureaucracy" that was required on the project. I have no words to express my gratitude towards René, for his patience, teaching and advice during all of this process. Sincerely, thank you for all your help, professionally and personally, during this time. *Bedankt voor alles*.

The STARRY project, split my PhD adventure into two wonderful places. The first place was *Leeds*. This city welcomed me with both its open arms and its typical English weather, cold, grey and rainy! On the first day, this city gave me two wonderful friends, Bethany Ashcroft and Robert Pomohaci. We have

had a great connection since day one, when they hosted me without previously knowing me. I am still grateful for all your support and the fantastic moments that we have shared and will continue to share.

I am grateful for the warm welcome I got from the university staff, who helped me with all my requirements. I am grateful for the professors and staff from the School of Physics and Astronomy and especially to all the PhD students of the Astro Group, all those that started in my same year and to the ones who came along later. Alice B., Miguel, Abigail, Willice, Tonye, Mark, Rosie, Marcin, Eve, Rob S; *Thank you for all the great times we had during these years.* Particularly to Chumpon for all our lunchtimes together; James for being the best joint-desk buddy and *for all the grapes!* and I am very grateful to Evgenia Koumpia and Ivayla Kalcheva for their friendship and support during the good and bad times. *Ladies, you do not know how happy I am to know and have you in my life. Love you both.* Another person from the University that give me all her support and friendship, from day one, was Katherine Taylor from the Language Centre of the University. Thank you Kate for everything that you have taught me, both professionally and personally, and for all our delightful conversation with coffee and teacakes.

The second place was *Madrid*. It was 18 exciting months, full of wonderful people and unforgettable moments. I would like to give thanks to Deborah Baines and Ricardo Pérez-Martínez for all their support and their words of encouragement during my time at the European Space Agency (ESA). A big thanks to all Ingeniería de Sistemas para la Defensa de España (ISDEFE) staff for their teachings; and, to the staff of the Centro de Astrobilogia (CAB) who made many wonderful moments for me during the time I spent in their facilities. Especially to Beatriz "Bea" Gonzalez for all the morning conversations before starting work and to Dr Benajmin Montesinos for all the interesting conversations about astronomy and life in general from Madrid to ESA and vice versa. In addition, I want to thank Dr Hector Canoves and Carles Cantero, for all the knowledge they provided about the operation of clustering algorithms and the help they provided me to compare my results with the algorithms in the literature. Finally, I cannot leave aside the ESA

trainees who made my life in Madrid more enjoyable, with all the outings to eat, dance, board game nights, traveling or just walks in town. Special thanks to Amy Joyce, Lionel Métrailler, Christos Ntinos, Francesca Husanu and Alfredo Escalante, for all the wonderful times together.

Thanks to the Model and Observation of Disk Evolution in Latin America (MODELA) group for their interesting conversations about astronomy. Especially to Jesús Hernández, for his advice and words of encouragement throughout my academic training and for always reminding me that *Astronomy is a science of patience.*

Last and certainly not least, I cannot leave out the person who has been by my side during the last months giving me his support and help, at all times. Thomas Speak, I am grateful for all the help and love you have given me and continue to give me.

Gracias.

Abstract

It is a well-established result that many stars do not form in isolation; young stars are usually found to be members of clusters. But little is known or understood about the origin of these clusters. In particular, evidence that pre-main sequence stars of intermediate (2-10 $M_{\odot}$) and higher masses are found in clusters has been found in several studies, at both optical and infrared wavelengths. Additionally, there has been an increased interest in the study of intermediate-mass stars, such as the Herbig Ae/Be stars, in the past ten years. These stars represent the most massive objects to experience an optically visible pre-main sequence phase, bridging the gap between low- and high-mass stars.

Studies in the nineties into the occurrence of young stellar clusters around Herbig Ae/Be stars concluded, based on near-infrared imaging data, that there was a difference in clustering properties between low and high mass stars.

Following the ideas included in those studies, the work in this thesis investigated the presence of clusters around 269 known Herbig Ae/Be stars, with the detailed astrometric data offered by the second Gaia data release with a novel clustering detection algorithm (CEREAL), which was developed for this task during this project. CEREAL is available from https://github.com/yumiry/CEREAL, along with several examples of its use for cluster classification.

Prior to the work included in this thesis, only 15 Herbig Ae/Be stars have been found to be in a cluster but, with CEREAL, it was possible to find clusters around 76 (28%) Herbig Ae/Be stars, which represents a significant improvement in the number of clusters found around Herbig Ae/Be stars.

In parallel to this, the results obtained with CEREAL were compared with other density-based clustering algorithms, where it was demonstrated that CEREAL

was a simple and powerful, algorithm that could find clusters, of any size, from any sample.

In addition, the work included in this thesis has proved that clusters around the Herbig Ae/Be does not just appear for the sub spectral type range from B7 and earlier; this work has shown it is possible to find clusters around all the sub-spectral types of B stars.

This work then concluded through a comparison between the clusters found with CEREAL, around the Herbig Ae/Be stars, with Monolithic collapse and Competitive Accretion models from the literature which the aimed to evaluate whether the clusters around the massive stars appear to follow any of these models. The result of this comparison is that neither of these models appeared to fully describe the observed clustering behaviour of the Herbig Ae/Be stars observed by CEREAL; although, elements of the behaviours expected by both models were observed, which raises the possibility that a combination of both models may be required to fully explain massive star formation.

Contents

List of Figures

List of Tables

Abbreviations

CEREAL	ClustER detEction ALgorithm
CMD	Colour - Magnitude Diagram
CTTSs	Classical T Tauri stars
DCAs	Density-base Clustering Algorithms
DEC	Declination
Gaia DR1	Gaia Data Release 1
Gaia DR2	Gaia Data Release 2
Gaia eDR3	Gaia Early Data Release 3
GCs	Globular clusters
G magnitude	Photometry mean G magnitude
GMC	Giant Molecular Cloud
HAeBe star	Herbig Ae/Be star
HRD	Hertzsprung - Russell Diagram
IMF	Initial Mass Function
IMTTS	Intermediate Mass T Tauri Stars
ISM	Interstellar Medium
MAYBE or *maybe*	Possible stars found to be part of a cluster by CEREAL
NO or *no*	Stars not found in cluster by CEREAL
OCs	Open clusters
para or ϖ	Parallax
pmdec or μ_δ	Proper motion in DEC
pmra or μ_α	Proper motion in RA
PMS	Pre-main sequence stars

RA	Right Ascension
RC	Radius of the cluster or Clusters Radius
RUWE	Re-normalized Unit Weight Error
SED	Stellar energy distributions
SIMBAD	Set of Identifications Measurements and Bibliography for Astronomical Data (Wenger *et al.*, 2000)
TGAS	Tycho-Gaia astrometric solution
TTS	T Tauri Stars
WTTSs	Weak-line T Tauri stars
year	yr
YES or ***yes***	Stars found in cluster by `CEREAL`
YMCs	Young massive clusters
ZAMS	Zero-age main sequence

Chapter 1

Introduction

'El destino se lo forja uno a golpes y trabajo, yo haré con mi existencia lo que me de la gana... siempre que salgo vivo y pueda volver a casa'

— *El plan infinito,* Isabel Allende.

1.1 Overview

The pioneering works of Hartmann (1904) and Trumpler (1930) showed that the space between the stars is not empty, but consists of a low-density interstellar medium (ISM), this is composed of 99 % gas and 1% dust. The ISM is very diverse and shows a wide range of densities (from 10^{-3} to 10^{3} particles per cm^3) and temperatures (from 10 K to 10^{6} K, McKee & Ostriker, 1977; Schulz, 2005). The ISM can be generally classified into three phases: a hot phase (made up of coronal gas and shocked gas from supernova explosions), a warm phase (primarily atomic and ionized hydrogen) and a cold phase (containing both molecular and atomic hydrogen, along with dust; McKee & Ostriker, 1977; Schulz, 2005); molecular clouds can be found within the cold regions and these represent the densest parts of the ISM. Molecular clouds are complexes of interstellar material with temperatures between 10-50 K with typical densities $> 10^{3}$ particles per cm^3 (Schulz, 2005).

A molecular cloud collapses when the gravitational force exceeds the combined forces, of the thermal pressure of the gas, the turbulent movements of the material and the magnetic fields, that all act against this collapse (Hartmann, 2009). As the collapse progresses, the cloud decreases in size and becomes more dense, dividing into smaller parts (fragmentation). The force of the gravitational attraction compacts the material in the centre in each fragment on time scales of a few thousand years. These fragments or dense cores are the protostars (which is the final stage before the formation of the star. Prialnik, 2000, See figure 1.1). This process of fragmentation implies that the new stars usually appear in clusters, which are a set of stars whose formation appears to have been related (Bennett *et al.*, 2011; Lada & Lada, 2003; Ward-Thompson & Whitworth, 2011). These clusters have been found to possess as few as 10 members or as many as 10^7 (Ward-Thompson & Whitworth, 2011). An advantage of studying clusters is that the stars share a common origin and the cluster members therefore have the same general properties, such as distance, chemical composition, kinematic properties (proper motion, parallax and radial velocity) and age.

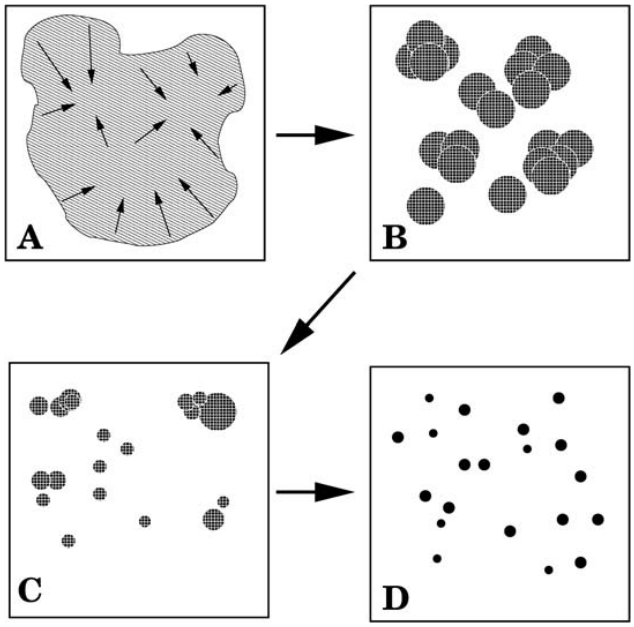
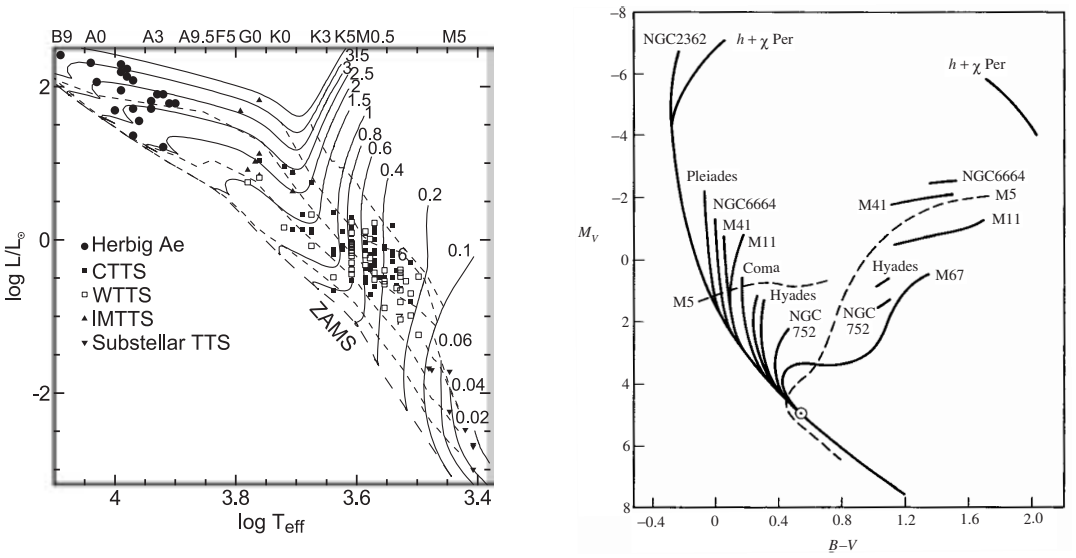


Figure 1.1: A schematic of the fragmentation process. In (A) parts of the cloud fractionalize into cores, which then begin to collapse (B) into protostars (C). The latter evolve (D) towards the main sequence. Figure taking from Schulz (2005).

The Hertzsprung-Russell diagram (HRD), or its observational equivalent, a Colour-Magnitude diagram (CMD), are an indispensable tool in the study of stars formed in clusters and their stellar evolution (Lada & Lada, 2003). Figure 1.2 shows an example of how a variety of young stellar objects in different mass ranges can be identified in a HRD, together with theoretical evolutionary tracks which indicate these objects are young and still approaching the main sequence. Figure 1.2 also shows a representation of several clusters of different ages (Garcia, 2011; Kutner, 2003)

There has been an increased interest in the study of intermediate-mass stars in the past ten years, these are stars with a mass range between 2-10 $M_{\odot}$ (Waters & Waelkens, 1998, references therein). Herbig Ae/Be stars (HAeBe stars) are pre-main sequence (PMS) stars of intermediate-mass between 2-10 $M_{\odot}$. These represent the most massive objects known to experience an optically visible PMS phase, they can therefore be con-



(a) (b)

Figure 1.2: Hertzsprung-Russell diagram. *Left:* Location of young stellar objects in HRD. Shown are sub-stellar T Tauri stars, classical T Tauri stars (CTTSs), weak-line T Tauri stars (WTTSs), Herbig Ae stars and Intermediate-mass T Tauri stars (IMTTSs). Zero-age main sequence (ZAMS, long-dashed line), evolutionary tracks (solid lines, labeled by the corresponding stellar mass in solar masses), and isochrones (dashed lines, corresponding to 0.3, 1, 3, 10, and 30 Myr from top to bottom). Figure taken from Garcia (2011). *Right:* Schematic HRD for various galactic clusters. Figure taken from Kutner (2003)

sidered to bridge the transition between low- and high-mass stars (Gomez *et al.*, 1993; Mendigutía *et al.*, 2012; Testi *et al.*, 1997).

It is a well-established result that many stars do not form in isolation; young stars are usually found to be members of clusters (Lada *et al.*, 1993; Zinnecker *et al.*, 1993). But little is known or understood about the origin of these clusters. In recent years, the evidence that the clustering effect begins, at intermediate mass level has been found in several studies at optical and infrared wavelengths.

This chapter will provide an overview of star formation relevant to the work that was carried out in this study; which investigated the presence of clusters around intermediate mass stars, using the data available of the Gaia mission. This overview will start with the theory of star formation, which is followed by a discussion about clusters and finally a general picture about the theory of cluster formation. In addition, this chapter will provide information about the Gaia mission with which the cluster analysis around the HAeBe stars will be performed and then it will provide an outline of what is included in the thesis to follow.

1.2 Star formation

Star formation both regulates the evolution of galaxies, and is responsible for generating the initial conditions that lead to the formation of planetary systems. The knowledge on star formation has primarily focused on the formation of low mass stars (stars up to only a few solar masses). This is as these stars dominate the total stellar mass of galaxies; however, massive star formation is less well studied. Massive star ($8 M_{\odot}$) formation is key to fully understanding the evolution of galaxies, as massive stars have high impacts on their galaxies due to their winds, outflows and supernovae which affect the chemistry, morphology and nature of the molecular clouds, and the galactic superwinds (Frost *et al.*, 2021; Louvet, 2018, references therein). Understanding massive star formation is also important for generating a more general theory of star formation, that can describe the formation of stars of all masses in different formation environments (Tan *et al.*, 2014).

1.2.1 Low mass star formation

Star formation originates within giant molecular clouds (GMCs). GMCs are structures supported against gravitational collapse by the combination of magnetic fields and the thermal and turbulent pressure of the cloud (Hartmann, 2009).

These clouds have typical sizes on the order of hundreds of parsecs, typically contain a total mass between $10^4 - 10^6\ M_\odot$, and have typical temperatures around 10K. The main component from which GMCs are formed is molecular hydrogen (H_2; Carruthers, 1970) , and one definition of GMCs requires that these structures contain greater than $10^5\ M_\odot$ of H_2 (Solomon *et al.*, 1980).

The magnetic fields within these structures prevent the charged particles from collapsing; however, as the neutral molecules within the GMCs are not affected by this field, these particles separate from the ionised particles. As the proportion of particles that are ionized within GMCs is small this separation results in a loss of both magnetic and turbulent support within the cloud. As this support is lost the neutral particles begin to collapse into a pre-stellar core with the only force opposing this collapse being the thermal pressure. During this collapse the behaviour of the cloud can be described using virial theorem: $2K + U = 0$, where K is the kinetic energy and U is the potential energy; for gravitationally bound systems at equilibrium, it can be shown that the kinetic energy is one-half of the potential energy. This theorem describe the overall dynamic behaviour of a large assembly of bodies, rather than the precise behaviour of any individual body belonging to the assembly (Carroll & Ostlie, 2007; Harwit, 2006; Prialnik, 2000).

The GMC can be treated as a structure that is at equilibrium until any perturbation arises. Such perturbations can then cause the GMC to either collapse under its own gravity or begin to expand due to the gas pressure. There are many different processes that can cause perturbations to the cloud. These can include the explosive shocks of a supernova, stellar winds from nearby large (OB) stars or the effect of passing through the spiral arms of the galaxy (Smith & Brooks, 2008). As mentioned earlier collapse of these clouds can be described using virial theorem, using this equation it is then possible to calculate the minimum mass at which the cloud can then collapse. This minimum mass is known as the Jeans mass (M_J, Jeans, 1902) and is given by the following equation:

$$M_J = (\frac{5k_BT}{G\mu m_H})^{3/2}(\frac{3}{4\pi\rho})^{1/2} \tag{1.1}$$

In the equation 1.1 k_B is the Boltzmann constant, T is the temperature, G is the gravitational constant, μ is the mean molecular weight, m_H is the atomic mass of hydrogen, and ρ is the density of the cloud. If the mass is greater than M_J then cloud will collapse, this collapse then occurs under gravitational free fall. As the collapse proceeds the density increases which then itself causes M_J to decrease further. Smaller areas of the cloud can then additionally begin to collapse, the collapse of these regions is referred to as fragmentation.

Collapse of the GMCs then continue in their densest regions, this leads to the formation of clumps and cores within these areas. The densest and largest clumps will then form the basis for the formation of stellar clusters; the cores will instead lead to the formation of individual stellar systems (Williams *et al.*, 2000). When the cores first become unstable and begin gravitational collapse gravitational potential energy is lost, this release of potential energy leads to the luminosity of these objects. The release of energy from the core leads to a maintenance of the temperature of the collapsing core, as such this initial collapse can be considered to be isothermal.

As the collapse of the core continues the centre of the core becomes increasingly opaque to the energy it releases. This increasing opacity stops the collapsing core from effectively releasing the energy it is generating through gravitational collapse; this then leads to a rapid increase in the temperature of the core until it approaches hydrostatic equilibrium. Once the temperature of the core has elevated sufficiently ($>$ 2000K) molecular hydrogen (H_2) will begin to dissociate. Dissociation of molecular hydrogen is an endothermic process and this removes energy that would otherwise support the cloud against further gravitational collapse. Due to this removal of energy a second phase of collapse occurs which leads to the formation of a protostar from the prestellar material. This protostar does not contain all the material from the preceding core and the remaining material is referred to as the protostars envelope. The protostars will then accrete mass from its surrounding envelope. The effect of centrifugal forces on this envelope lead to the formation of an accretion disk, as the centrifugal forces are stranger than gravity at the equator. The formation of this accretion disk also helps to conserve angular momentum during the collapse. The temperature within the protostar will continue to rise due to shocks from accreted material falling into the core. Once the core has contracted significantly, it will be hot enough (10^6 K) to start nuclear burning of deuterium.

After a significant amount of further accretion has occurred, the protostar ceases to be convective and radiative processes begin to dominate the energy transfer within it. Eventually after further accretion and contraction, the internal densities and temperatures of the protostar will be so high that fusing hydrogen becomes the main source of the object's luminosity. Once hydrogen fusion takes over the object will stop contracting and it can be considered to have reached the main sequence.

The different phases of the pre-stellar evolution, that are observed for low-mass stars, are typically divided into sub classes (Class 0-III) depending on the shape of their stellar energy distributions (SED, Andre *et al.*, 1993; Lada *et al.*, 1993). A schematic that represents the evolutionary sequence, from the molecular cloud to a protoplanetary disk, of young low-mass stellar objects is shown as figure 1.3, this figure also presents the typical stellar energy distributions for each of the four sub classes.

The youngest class of pre-stellar objects are Class 0; the peak emission of these objects is observed in the submilimetre range with these emissions being dominated by the black-body radiation from the dust. This phase of the evolution of a pre-stellar object is relatively short and represents a rapid accretion phase and is typically believed to last around 10^4 years. Class I objects predominantly emit in the far-infrared. This phase represents the main accretion phase in the evolution for a pre-stellar object and typically this lasts around 10^5 years. Class II objects have an emission spectra with a dominant peak in the the near-infrared and this occurs when the disk around the object is present; the Class II phase lasts around 10^6 years. The final phase, Class III, has a spectra similar to that of a main sequence star; although, these objects still have remnant of circumstellar dust around them that has yet to be fully dispersed by the stellar winds.

1.2.2 Massive Star formation

Although, massive stars ($M_* > 8M_\odot$) only represent a small portion of the total stellar population ($\sim$1%), these objects have a dominant role in controlling the evolution and structure of their galaxies. This effect is due to the fact that massive stars input so much energy and momentum into the interstellar medium; this is through their large stellar winds, outflows, radiation, gravitational interactions and supernovae affecting many galactic phenomena such as the nature of the molecular clouds and the chemistry of the ISM along with the morphology of the galactic super-winds (Frost *et al.*, 2021; Leitherer,

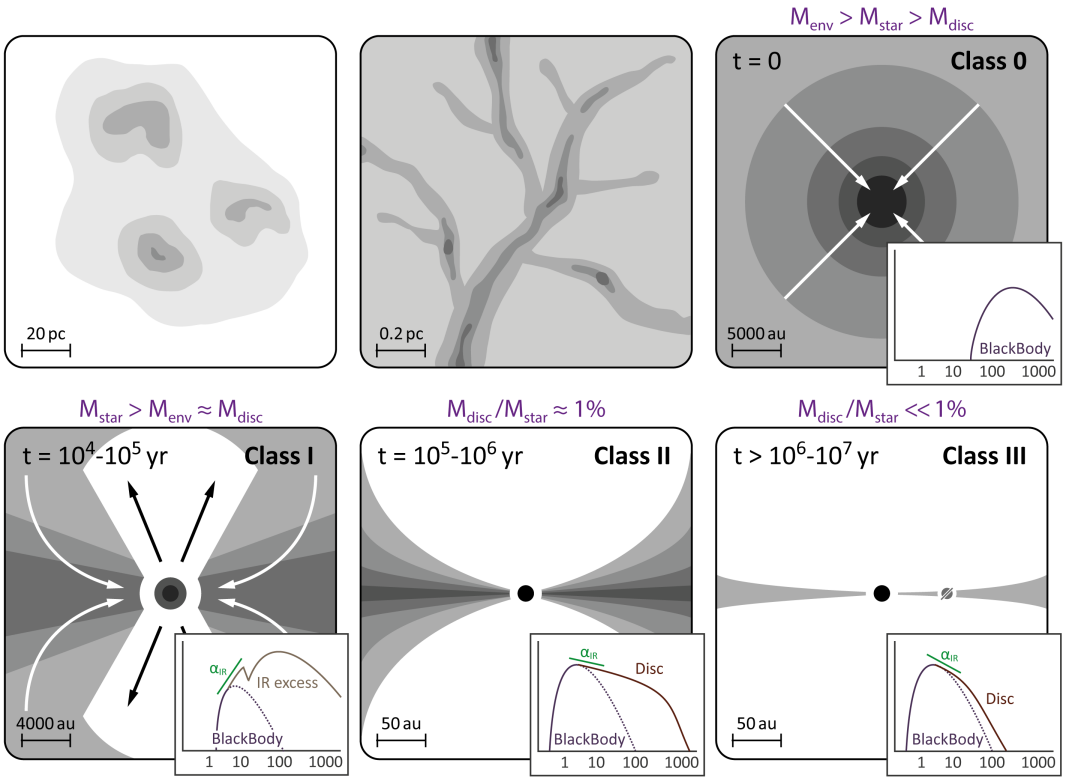


Figure 1.3: Illustration of the low mass star formation from molecular cloud to planetary disc. The figure shows the sizes, evolutionary class and typical formation times after cloud collapse. The figure also shows the SED of each class adapted from Jonkheid (2006). Image credit: Evans (2018).

1994). However, despite their importance to the understanding of galactic evolution these objects are not as well understood as low mass stars, in particular there are still significant uncertainties surrounding their formation (Louvet, 2018). There are several reasons that lead to this and one of these, that massive stars tend to have very high extinctions due to dust making observation of the early phases of their formation very difficult. This is as massive stars are typically found to form within massive dense cores ($\sim$100 $M_\odot$ within $\sim$0.1 pc) that are themselves found within hyper massive and hyper dense clouds (Louvet, 2018). Additionally, massive stars evolve much more quickly than low mass stars (see, e.g., Shu *et al.*, 1987) and unlike low mass stars there is no well established observational evolutionary sequence for high-mass star formation (Louvet, 2018).

Another factor that has inhibited their study is simply down to their scarcity coupled

with their fast evolution meaning that there are very few examples. Generally, massive stars are not seen to form in isolation and the proximity of other high mass stars along with their high gravitational interactions, powerful stellar winds, outflows and radiation significantly affecting the local environment of the young or forming massive star. That massive stars possess very high luminosities allowing them to be studied at larger distances than low mass stars; although, very bright objects can mask the presence of less luminous companions in their neighbourhood (Zinnecker & Yorke, 2007, references therein).

Massive star formation is therefore a very important area of study. Massive star formation as with low mass stars, begins through their contractions during the collapse of a molecular cloud; although these massive protostars have very high luminosities. A massive protostars can be considered to be a large object $\geq$ 8 $M_{\odot}$ in which hydrogen fusion has not yet begun to occur. As has been stated earlier the lifetime of these objects in this phase are very short and they exist only briefly between the presence of an acreting intermediate mass protostar and the presence of a massive star that is still accreting from the cloud. It should be noted that these objects are not observationally distinguishable from massive stars that have begun the process of hydrogen fusion (Zinnecker & Yorke, 2007, references therein). As these massive protostars contracts its temperature also increases and these objects are seen to take a horizontal path on a Hertzsprung- Russell diagram, following a Henyey track of increasing temperature this can be seen in figure 1.4.

As has been stated earlier the formation of massive stars happens on a much shorter timescale than that of low mass stars. To better understand the time it takes for a massive star to form the Kelvin-Helmholtz timescale can be used; this describes the time it takes for a star to convert though contraction its gravitational potential energy into luminosity. The Kelvin-Helmholtz timescale represents a reasonable proxy for the formation timescale of a star. This timescale is shorter for massive stars, than the free-fall time of the molecular clouds from which they form; this means that the massive protostars form as deeply embedded objects in their molecular clouds and the entire process of massive star formation can be obscured by the surrounding material.

Although they are obscured, a hydrostatic core will still form in the centre of the collapsing cloud and this will still grow through accretion as was described for low mass star formation earlier. But, due to their size massive star formation differs from that of low mass star formation with hydrogen fusion beginning whilst accretion is still occurring during

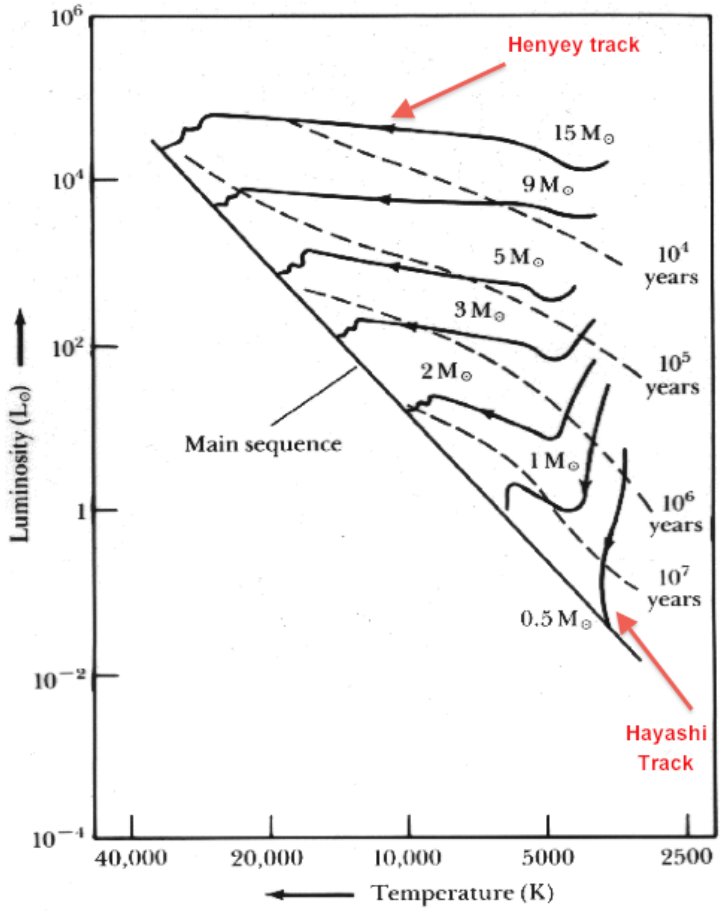


Figure 1.4: Hertzsprung-Russell diagram for stars with different masses, where the Hayashi and Henyey tracks shows how the stars evolve into the main sequence. The evolution tracks are also show. Image credit: Frost (2019).

the formation of massive stars. This occurs when the cores reach sufficient temperature for hydrogen fusion to begin, with the larger denser cores reaching sufficient temperature for hydrogen burning whilst accretion is still occurring.

As a result the massive stars still continue to grow after they have reached the main sequence as they continue to gain mass (they can no longer be considered massive proto-stars as they are burning hydrogen) which is markedly different to low mass stars which stop accretion of mass and growth once they reach the main sequence. Additional information of massive star formation that covers both the theory of their formation and observation can be found in the reviews by Zinnecker & Yorke (2007, references therein)

and Motte *et al.* (2018, references therein)

The study of massive star formation is of great importance to the understanding of galactic evolution and as has been described above the studying of young massive stellar objects is complex as unlike low mass stars which lose their envelope early in their evolution, massive stars remain deeply embedded during their formation. This is in addition to these objects being scarce within the universe. One area of work in understanding the formation of massive stars is through the study of large intermediate mass stars as proxies for massive stars. One such type of intermediate mass stellar object are Herbig Ae/Be stars.

1.3 Herbig Ae/Be stars

Herbig Ae/Be stars (HAeBe stars) are optically visible intermediate pre-main sequence stars whose masses range from about 2 to 10$M_{\odot}$. These objects were originally identified by Herbig (1960). During his work, he was searching for young pre-main sequence stars which are the higher mass analogues to CTTs. To find these stars, the objects had to meet the following criteria: spectral type A or earlier with emission lines; lies in an obscured region; and the star illuminates fairly bright nebulosity in its immediate vicinity. The first criteria is what defines these stars, while the other two are used to reduced the contamination of the sample by stars which are not of PMS nature. The criteria for HAeBe stars have been relaxed in more recent surveys attempting to find more candidate HAeBe stars (de Wit *et al.*, 2014; Hartmann, 2009; Hernández *et al.*, 2004, 2005; The *et al.*, 1994; Vieira *et al.*, 2003). These studies have considered other indicators associated with young stars, in order to find more targets. For example, include stars with spectral type F (F5 or earlier), the location of the HAeBe in region less obscured and the presence of infrared excess and UV.

An important issue relating to the study of the intermediate mass stars is that there is not a large number of those objects identified which makes the statistics non-robust. This has led to an increase in interest in the study of intermediate-mass stars, with mass range between 2-10 $M_{\odot}$ (Waters & Waelkens, 1998, references therein). The lower limit ($2M_{\odot}$) corresponds to the mass at which stars are radiatively stable when they begin their contraction. The upper limit (10 $M_{\odot}$) corresponds to the mass above which stars start burning hydrogen before they emerge from their contracting envelope. Therefore,

intermediate-mass stars spend a relatively long time in the protostar contraction phase (Waters & Waelkens, 1998, references therein). Stars like the HAeBe stars play an important role in understanding massive star formation, because they bridge the gap between low-mass stars whose formation is relatively well understood, and high-mass stars whose formation is still unclear (Vioque *et al.*, 2018, 2020; Wichittanakom *et al.*, 2020). Recently, Vioque *et al.* (2020, see figure 1.5) obtained a new catalogue of 8470 new pre-main sequence stars where at least 1361 sources are potentially new Herbig Ae/Be candidates according to their position in the Hertzsprung-Russell diagram.

Studies like Hillenbrand *et al.* (1995); Testi *et al.* (1997, 1998, 1999, 2000); Wang & Looney (2007, see figure 1.6) on the infrared; the work carried out in this thesis, represents the first steps in bringing together the study of the formation of intermediate mass stars with the formation of high-mass stars.

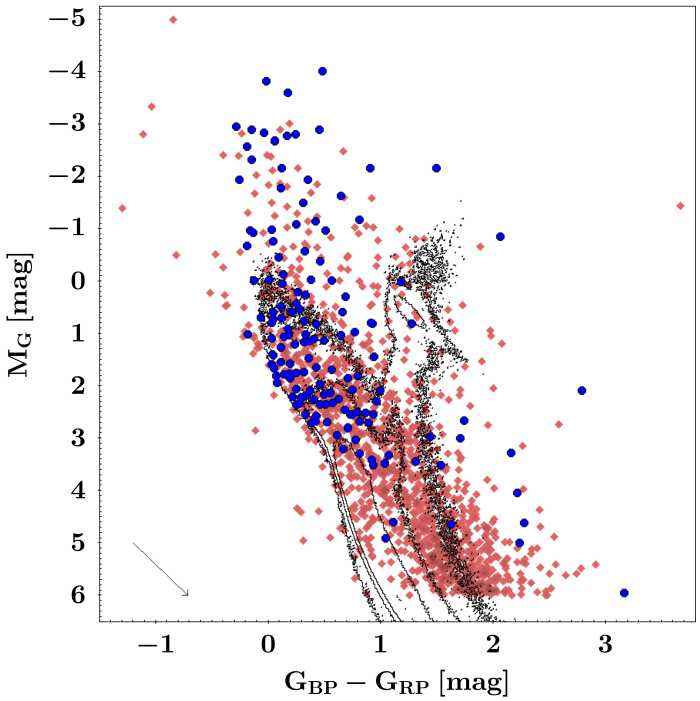


Figure 1.5: Colour magnitude diagram of known Herbig Ae/Be stars (blue dots) and new candidates to Herbig Ae/Be stars (red diamonds), corrected from interstellar extinction. The Black contours trace Gaia sources within 500 pc with good astrometric solution. The extinction vector corresponding to $A_G = 1$ is also shown. Figure 1.5 is taking from Vioque *et al.* (2020).

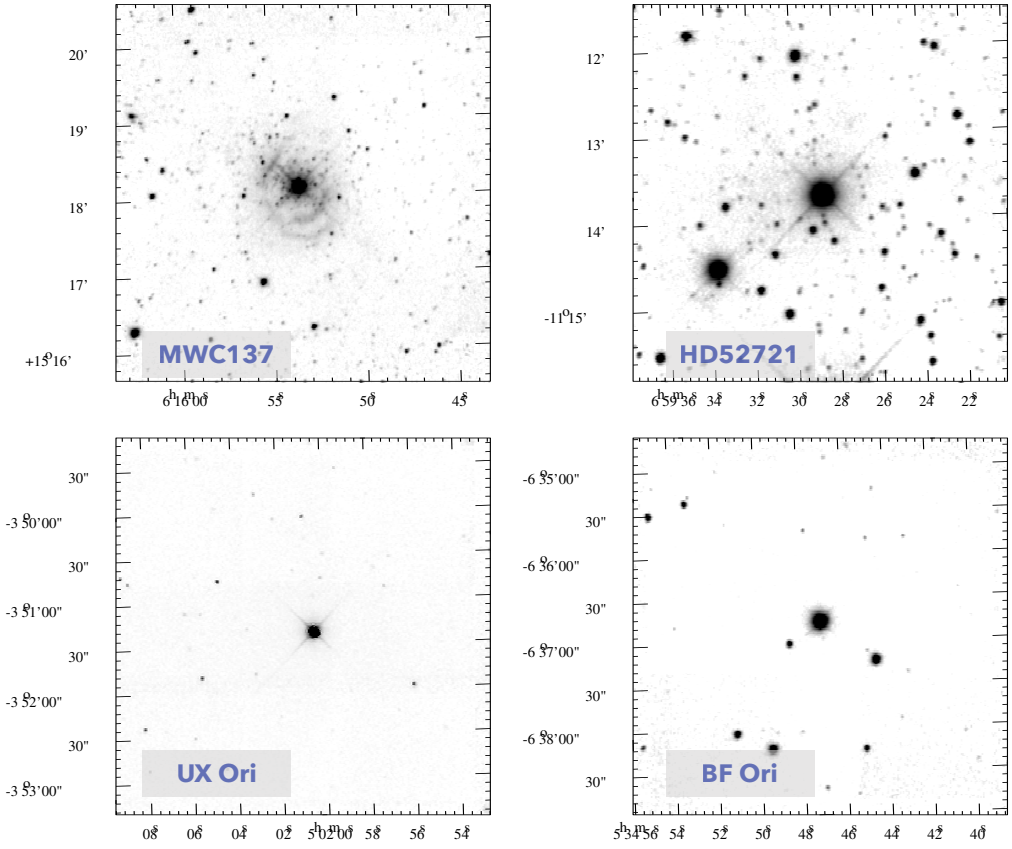


Figure 1.6: Examples of Herbig Ae/Be stars in rich and poor clusters. The figure represent the K-band image, where the axes are the right ascension and declination for the 1950 epoch, of four Herbig Ae/Be stars which appear to be surrounded by a large number of low mass companies (top panel) or single (bottom panel); Figure 1.6 is taking from Testi *et al.* (1997, references therein).

1.4 Clusters

Prior to a detailed evaluation of the presence of stellar clusters it is important to precisely define what a cluster is. There have been several reviews that have each provided their definition of what a cluster is and how to differentiate it from other multiple populations (Krumholz *et al.*, 2019, references therein). For example, one definition, taken from Portegies Zwart *et al.* (2010), is that a star cluster is a set of stars that are gravitationally bound to one another; while Lada *et al.* (1993) defines a cluster as a collection of stars with a mass density large enough ($\geq 1\ M_{\odot}\ \mathrm{pc}^{-3}$) to resist tidal disruption in Solar neigh-

borhood conditions. With this in mind, a general definition for a stellar cluster can be, the one offer by Krumholz *et al.* (2019, references therein) that a cluster is a group of at least 12 stars with a mean density that is several times that of the background density, such that the overdensity is then still statistically significant. In addition, it is also an important characteristic of a stellar cluster that the stars contained within them need to have been physically associated with each other during their formation (Krumholz *et al.*, 2019; Lada *et al.*, 1993; Trumpler, 1930, references therein).

Clusters are born in molecular clouds, specifically in high density concentrations (clumps) within giant molecular clouds with typical masses up to 10^6 $M_\odot$ (Krumholz *et al.*, 2019, references therein). Figure 1.7 shows a variety of clusters which have a wide range of ages from $\sim$1 Myr to $>$ 10 Gyr, and with masses ranging from $\sim$ 10^2 - 10^6 $M_\odot$ (Krumholz *et al.*, 2019, references therein). Star clusters can be characterized as gravitationally bound or unbound; however, young clusters can be a mixture of both bound and unbound. Bound clusters can be categorized as either open clusters (OCs) or globular clusters (GCs); the most massive and young OCs are sometimes referred to as young massive clusters (YMCs, Krumholz *et al.*, 2019, references therein). Unbound clusters can also be described as associations, these associations can then be sub-divided into OB associations (Wright, 2020), which are regions which contain an excess density of spectral type O and B stars; and, T associations, formed by T Tauri stars (Herbig, 1962; Janes, 2001). Although, the stars in these unbound associations will disperse after a short period of time as the associations lose their identities, these stars will still have almost identical velocities; as these stars will still retain very similar movements whilst spreading out through space (Janes, 2001).

1.4.1 Massive star Clusters

The process of star formation from GMCs has been described earlier in this description the composition of these GMCs has been described. Where these GMCs are observed they contain both dense gas and evidence of the initiation of star formation. This leads to the reasonable conclusion that dense cores form very rapidly after the dense GMCs form from the less dense and diffuse ISM, star formation itself proceeds after the formation of these dense cores. In more detail, the GMCs are gravitationally bound systems that form from the ISM, these systems have motions within them that contain very turbulent

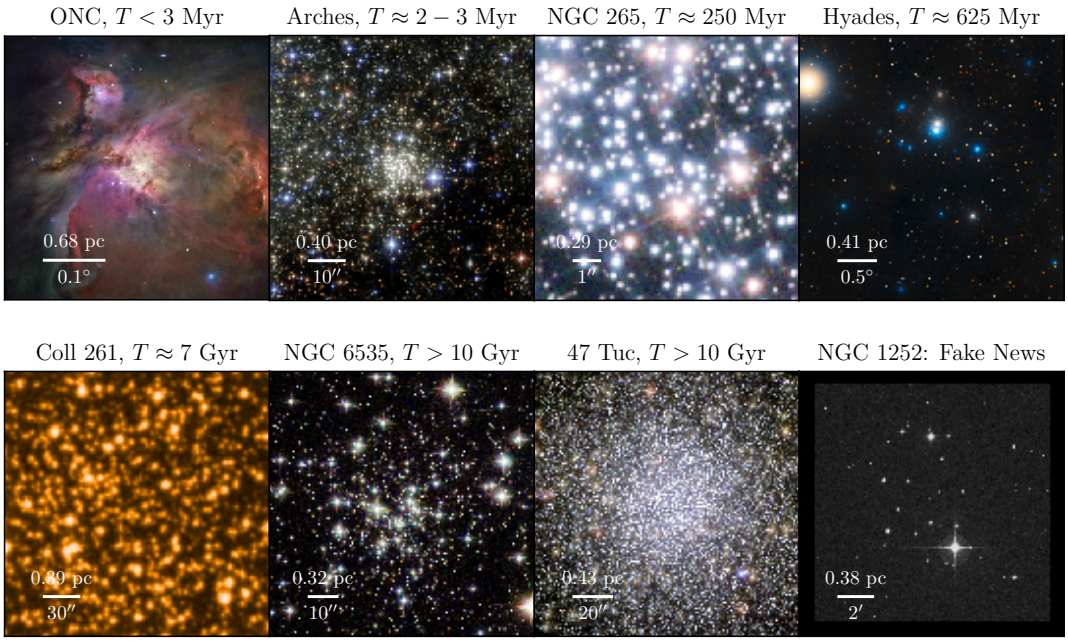


Figure 1.7: Collection of stellar clusters. This figure shows that, in general, clusters can have different masses, size, and density scales. The cluster represent in the figure are: Orion Nebular Cluster (ONC), Arches cluster, NGC265, Hyades, Collinder 261, NGC6535, 47 Tuc and NGC1252. NGC1252 was previously classified as a cluster but now is know as an asterism. Figure taking from Krumholz *et al.* (2019, references therein).

flows, and are stabilized against collapse by their turbulent pressure. Several numerical simulations of GMCs have been carried out that show that supersonic turbulent flow conditions produce collisions and shocks that dissipate energy within the clouds. These collisions can then lead to the formation of dense cores which are then gravitationally unstable and decoupled from the turbulent flows of the GMCs. These decoupled dense cores within the GMC represent the potential sites for cluster formation (Lada *et al.*, 1993).

YMCs are dense aggregates of young stars and are sometimes referred to as young globular clusters; although, it is actually not known that these systems will evolve into clusters like the GCs that are currently observed. YMCs have traditionally been defined as those that have masses $> 10^4\ M_\odot$ and ages < 100 Myr; there are several examples of YMCs in the Milky Way Galaxy and within the Local Group, but are predominantly observed in starbursts and interacting galaxies (see figure 1.8, Bastian, 2016).

Figure 1.8: Colour composition image of Young Massive clusters Cep OB3b (4-5 Myr), part of the Cep OB3 molecular cloud complex. The image was created using three colour images from Spitzer (3.6(blue), 5.8 (green), and 24 μm (red)). Figure taking from Krause *et al.* (2020).

Although, a minimum mass and age limit was given above for YMCs there is no strict definition of what is a YMC and these limits are not based on a physical properties of the clusters themselves. This type of clusters can be considered to be one of the fundamental building blocks of galaxies.

As a result, a significant amount of recent research has focused on YMCs; of this a large amount of work has focused on the question of whether at least some of the YMCs that have been associated with the most violent starburst events in the local Universe will evolve into equivalents of the GCs that are common place within our local universe and which are observed in almost all of the local galaxies (de Grijs, 2010). YMCs tend to be formed in some of the more massive GMCs. The largest and most massive GMCs are often (and its reasonable to expect them to be) found within gas rich regions for instance in starbursts or forming galaxies (Krumholz *et al.*, 2019, references therein).

The study of YMCs is of additional importance as the clustering properties of massive

stars are still not fully understood. It is known that the majority of massive stars that are observed are part of, or are a 'runaway' that has been ejected from a star cluster; however, as it is more likely that massive stars will be found in clusters it is still unclear if this is because massive stars form only or predominantly in clusters (Krumholz & Bonnell, 2007, references therein).

There is still an important question around the formation of YMCs and how they get to the sizes and shapes that we observe them to have. To attempt to solve this question there have been several studies which have been both observational and modelling works which have studied GMCs and these works have concluded that massive clusters are formed along filamentary substructures within the clouds. These structures are small scale with respect to the GMC on the order of $\leq$0.3 pc (Banerjee & Kroupa, 2017a). Additional information of massive clusters formation can be found in the review of Motte *et al.* (2018).

1.5 Theoretical models

As has been briefly discussed earlier work on studying massive star formation and the formation of young massive clusters has not been merely observational; modelling studies have also been carried out to investigate the origins and properties of these systems. In particular, models of cluster formation have shown that the formation of gravitationally unstable clumps within the turbulent flows of a GMC, can be followed by these clumps fragmenting hierarchically into smaller systems (clusters) that eventually merge to form large stellar systems (Bonnell *et al.*, 2003; Krumholz & Bonnell, 2007, references therein). These results have been supported by observations, where the fragmentation of filamentary structures within infrared dark clouds has been shown to be linked to the formation of stellar clusters.

Within the computational studies of the dynamics of star formation numerical simulations have shown the importance of the multiple factors, including the turbulence and collapse of the cloud, along with the following fragmentation, accretion and other interactions (Bonnell *et al.*, 2003). These simulations are very useful in their ability to highlight the importance of certain physical processes within star formation; however, these models are reliant on highly idealised behaviours and conditions and as such are not necessarily fully representative of real (and often very non ideal) systems (Smilgys & Bonnell, 2017).

The models for massive star formation have to differ from those of low mass star formation; this is in particular due to the fact that the intense radiative pressures associated with the massive stars mean that no star over 40 $M_{\odot}$ can form via spherical accretion (Kahn, 1974). In addition, high mass stars can not sustain a magnetic field during formation, as energy transfer through the envelopes of massive stars is radiative as opposed to convective as is the case for low mass stars. Two main theoretical approaches have been taken to describe massive star formation, and these approaches are referred to monolithic collapse and competitive accretion (Bonnell *et al.*, 2003).

1.5.1 Monolithic collapse or Core Accretion model

The core accretion model (which is also referred to as monolithic collapse) is a general model of star formation, that attempts to explain the formation of both low and high mass stars. This model predicts that stars form via a fragmentation process, whereby the molecular cloud breaks down sequentially into smaller and then even smaller parts, due to factors related to their turbulence and magnetic fields along with their gravitational forces during the collapse of the cloud. This fragmented collapse process continues until the formation of parts that no longer undergo any further fragmentation prior to their collapse to stellar densities, these parts are referred to as cores (see figure 1.9, Krumholz & Bonnell, 2007, references therein).

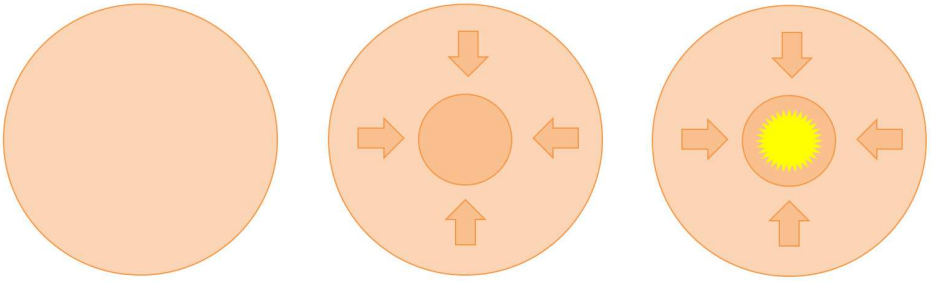


Figure 1.9: Representation of the Monolithic collapse or core accretion scenario. The fragmentation of the core is halted, allowing the formation of a single massive (or a few) objects. According to this model, the massive stars are born in isolation or accompanied by other massive stars, rather than with a low-mass stellar cluster inside the core. Figure taking from Rivilla *et al.* (2013).

This model can be seen as an adapted version of the model of low mass star formation which has been developed to account for observed differences between the formation of low mass stars and massive stars. Within this model a single massive core leads to a single massive star and the initial mass of the clump within which this massive core is located is a significant factor in the final stellar mass of the star formed through this model. This means that the mass function that describes the central pre-stellar core should follow the initial mass function of a stellar object, this effect has been observationally seen by André *et al.* (2010). This has the caveat that in the case of close multiple stellar systems these protostars may compete with each other over the material they accrete.

Within this general set of models is the turbulent core model (McKee & Tan, 2003), which treats the clumps as structures at quasi-equilibrium formed as a result of the fragmentation. Within this model the quasi-equilibrium state of these clumps is maintained through kinetic energy released from outflows and accretion shocks from the pre-stellar cores. These pre stellar cores themselves have a density much greater than the clumps within which they are located and they have pressures much higher than both the average of the ISM and the cloud. Within this model the cores can be treated as supersonically turbulent as in fact there is no requirement within this model for them to become subsonically turbulent prior to the initiation of star formation.

This model produces a resultant timescale for the formation of massive stars on the order of 10^5 yrs, this timescale is only weakly dependent on the stellar mass ($t \propto M_*^{1/4}$) but, is very sensitive to the surface density of the initial clumps ($t \propto \Sigma_{cl}^{-3/4}$). The accretion rate from the clump to the core is of the same magnitude as the rate at which material is processed into the star itself, and the accretion rates are expected to grow with time. The accretion rate is expected to peak at $10^{-3}\ M_\odot\ yr_{-1}$.

The core accretion model for massive star formation will lead to an initial slow increase in stellar density over the first 4×10^5 years, this will then be followed by a gradual loss in stellar density at larger ages. This model predicts that massive stars will form with a single massive core that then leads to the formation of a single massive star, that can then be surrounded by many low mass stars (Bonnell *et al.*, 2003).

1.5.2 Competitive Accretion model or Hierarchical collapse

A separate approach to modelling massive star formation is the competitive accretion model (also referred to as hierarchical collapse, Bonnell & Bate, 2006; Bonnell *et al.*, 1998). In this model it is predicted that the fragmentation of the cloud into cores is a relatively inefficient processes and that it therefore leads to the formation of many cores that are still surrounded by a large amount of material that they can accrete. This material continues to be affected by gravitational effects and this leads to a distribution of gas within the cloud that is more centrally dense. This leads to the cores that are nearer to the centre of the cloud accreting more than those in the outer areas. In addition, as there are multiple cores competing with each other to accrete material, it is the earliest and largest of the cores that are formed within the cloud that will grow to become the largest stars in this model. Although the material that results in the formation of an individual massive star can come from different parts of the cloud (see figure 1.10).

The model assumes that the parental molecular cloud fragments into several condensations with masses around the Jeans mass and relies on the fact that nearly all massive stars are formed in stellar clusters with low-mass stars (Rivilla *et al.*, 2013, references therein).

One important difference in the evolution of massive stars by this model to that of the core accretion model is that the initial mass of the clump will not be representative of the final mass of the star once it reaches the main sequence. The accretion of matter into the cores is governed by the Bondi-Hoyle law where the mass within the stellar

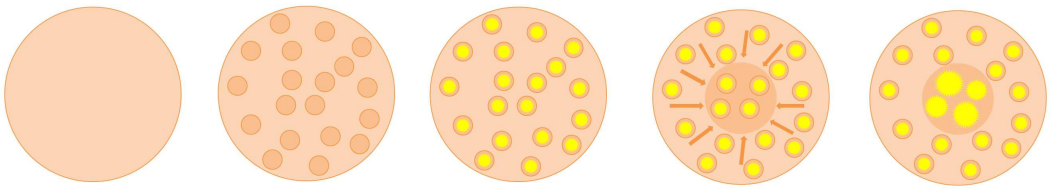


Figure 1.10: Representation of the Hierarchical or Competitive accretion scenario. The core fragments into sub-cores producing a cluster of low mass stars; where a fraction of the gas is incorporated into the low mass stars and the rest is distributed throughout the core. This scenario produces a whole cluster of stars with different masses. Figure taking from Rivilla *et al.* (2013).

mass within the core being greater than the gravitational potential; this effect leads to the early accretion process being relatively slow and limited to the material that is close to the core. However, as the cores then increase in mass the area over which their mass dominates also expands and they can accrete material from further away. As the cores can accrete material from further away the rate of accretion increases leading to faster increases in mass and the ability to accrete from even further away in a form of runaway accretion.

In the competitive accretion model it is the location and time at which a core forms within a clump that determines the final mass of the star that the core will form, not the initial mass of the core. One prediction of this model is that the disks formed in this model will be smaller than those expected from the core accretion model which predicts much larger disks will form. This is because the large number of cores present in this models interactions inhibit the formation of large disks. One important assumption of this model is that in the initial conditions, the cloud is strongly gravitationally bound, whereas observations appear to show that these clouds are primarily turbulently supported. Also, the assumption of Bondi-Hoyle accretion rates may not accurately represent the true accretion rates due to the presence of high relative velocities between the cores and the turbulent gas of the cloud. In addition, it has been suggested that the objects formed within this model are less resilient to the stellar radiative feedback (Krumholz *et al.*, 2005).

However, in their subsequent work Bonnell & Bate (2006) find that even with initial conditions with high turbulent energy that is of a similar strength to the gravitational energy their results are still similar. They also stipulate that Larson's law predicts that stars and neighbouring gas have similar velocities. In addition, Bonnell & Bate (2006) claim that in their critique of the competitive accretion Krumholz *et al.* (2005) did not account for the local properties of the massive star forming cluster but instead used more global parameters of the cloud. Bonnell & Bate (2006) argue that the initial mass function formed through their competitive accretion model agrees well with those that are observed around massive stars, and that this model accounts both for the mass segregation that is observed in young clusters and the binary properties of stars that have been observed.

The most massive stars formed through this competitive accretion model will form at the centre of the clumps and they will be formed with little motion compared to the surrounding gas.

1.6 Observational data

The purpose of this thesis is to perform an assessment of the clustering properties of known Herbig Ae/Be stars. For this analysis, I use the data from the Gaia mission as the main source of information of this thesis. In addition, I used the recent compilation of the stellar parameters of the Herbig Ae/Be stars made by Vioque *et al.* (2018).

The main aim of the Gaia satellite is to measure the three-dimensional spatial and velocity distribution of stars and to determine their astrophysical properties, improving the results of the Hipparcos mission (ESA, 1997; Gaia Collaboration *et al.*, 2016b; Perryman *et al.*, 1997). Launched at the end of 2013, it provides highly accurate positions, parallaxes and proper motions for $\sim$ 1 billion of sources brighter than 20.7 magnitudes in the photometric G band (Michalik *et al.*, 2015). The astrometry data is complemented by multi-colour photometry, measured for all sources observed by Gaia, and radial velocities which are collected for stars brighter than G $\approx$ 17 (Gaia Collaboration *et al.*, 2016a).

The first data release, after 14 months of the mission, contains the astrometric dataset $\sim$ 2 million sources in common between Hipparcos and Tycho-2 catalogues (TGAS; ESA, 1997; Gaia Collaboration *et al.*, 2016a; Høg *et al.*, 2000). The second release, 22 months of observations made by the mission (Gaia Collaboration *et al.*, 2018b, see figure 1.11), include a five parameter astrometric solution for over 1 billion sources. This also includes multi band photometry (in three bands for up to G $\leq$21 mag) and a large radial velocity survey at the bright end (G $\lesssim$ 13; Gaia Collaboration *et al.*, 2018b). Recently, the European Space Agency released a subset of the future Gaia data release 3 (Gaia eDR3) planned to be available in 2022. Gaia eDR3, 34 months of observations made by the mission, contains provisional astrometric and photometric data for $\sim$ 1.8 billion sources in the magnitude range G = 3 to 21 (Lindegren *et al.*, 2020, 2021). The median uncertainty in parallax and annual proper motion is 0.02-0.03 mas at magnitude G =9 to 15, and around 0.5 mas at G = 20 (Lindegren *et al.*, 2020, 2021). In this thesis, an analysis with Gaia eDR3 was not performed, since this data was available after the analysis made in this study.

The Gaia mission represents an innovation in astronomy, since new accurate data has become available which can be used to identify and analyse a large number of low-, intermediate- and high-mass stars (Wright, 2020) and, for example, to trace stars that are now part of the field back to their birth places (Krumholz *et al.*, 2019).

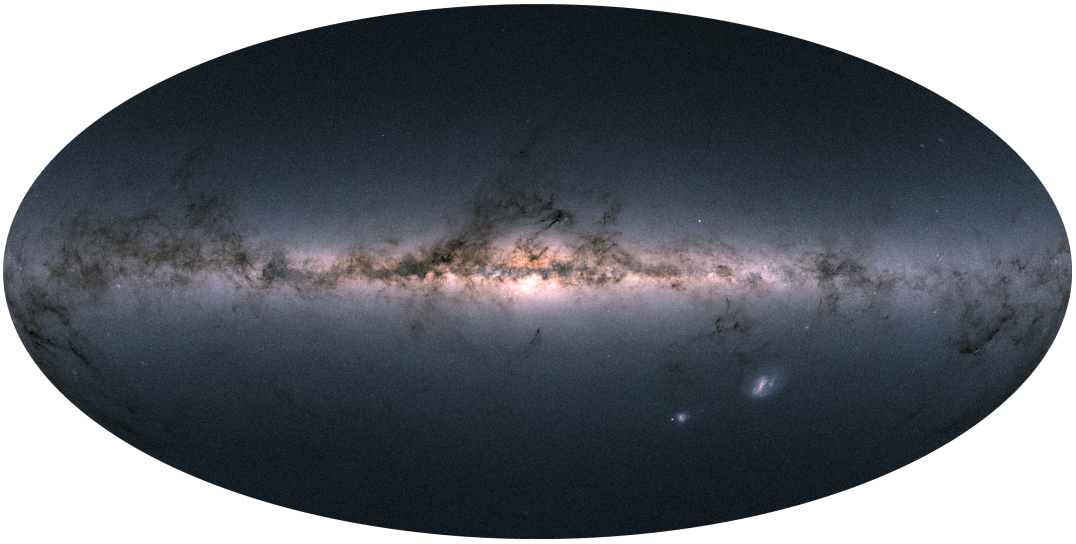


Figure 1.11: Map of the total flux measured in the G, G_{BP} and G_{RP} bands for the Gaia DR2 sources. The colours represents the flux measured in the G_{BP} (blue), G (green) and G_{RP} (red). Figure taking from Gaia Collaboration *et al.* (2018b).

1.7 Thesis outline

The absence of a statistically meaningful sample of intermediate mass stars in clusters is what motivated this project. Adapting the ideas proposed by Testi *et al.* (1997, 1998, 1999), this thesis will describe work carried out to detect and analyse the presence of stellar clusters around known Herbig Ae/Be stars, with the astrometric data that became available through the first two data releases from the *Gaia* satellite. In order to use the astrometric data to find the relationships between the Herbig Ae/Be stars and their possible low mass companies a new clustering algorithm was developed, `CEREAL`.

The next part of this thesis will describe `CEREAL` and will then evaluate its ability to find low mass companions around stars in known clusters, using the data available from the first Gaia data release (Gaia DR1, Chapter 2). The following step was then to perform the reverse, analysing the presence of clusters for a given Herbig Ae/Be star and identify its low mass companions (Chapter 3). This was only possible using the data available from Gaia DR2, where `CEREAL` was used to evaluate the presence of clusters around a sample of 269 known HAeBe stars. The effectiveness of `CEREAL` was then assessed by comparing it to the combined results of 3 density-based clustering algorithms which are

a subset of machine learning algorithms for the study of overdensities on large data sets. The results of this analysis were then used to produce a catalogue of objects found to be in clusters by both techniques.

This thesis will then describe a more detailed examination of the clusters that were found (Chapter 4). This assessment will focus on the impact of spectral type (and sub spectral type) on the likelihood of a Herbig Ae/Be star to be found to be in a cluster by CEREAL. In addition, an examination to determine if spectral type was related to the likelihood of the Herbig Ae/Be star to be in the centre of the cluster that it was associated with.

The results found by CEREAL for the clusters found around B type stars were then used to examine if there were any observable trends in the stellar density of these clusters with age, and if these results were in agreement with the models of massive star formation (Chapter 5).

The final chapter (Chapter 6) summarizes the results of the previous ones and suggests areas of interest for future work that were raised during the work that is described in this thesis.

Chapter 2

ClustER detEction ALgorithm: CEREAL

'Every adventure requires a first step'

— The Cheshire Cat, Alice's Adventures in Wonderland.

The *C*lust*ER* det*E*ction *AL*gorithm (CEREAL) was created with the aim of detecting and analyzing the presence of clusters around intermediate mass stars using data from Gaia. Specifically, CEREAL searches for clusters around HAeBe stars, where it is expected that any companions surrounding these stars will both share similar astrometric information and be located nearby to the HAeBe stars.

Gaia data had been presented to the scientific community through three releases. The first, Gaia Data Release 1 (Gaia DR1), was based on the data collected during the first 14 months of the mission lifetime (Gaia Collaboration *et al.*, 2016a). This release contains astrometry, G-band photometry, and a modest number of variable star light curves. The astrometric dataset, which contains the positions, parallaxes, and proper motions; is formed by >1 billion stars, with $\sim$ 2 million sources in common between Hipparcos and Tycho-2 catalogues (ESA, 1997; Gaia Collaboration *et al.*, 2016a; Høg *et al.*, 2000). This dataset represents the realization of the Tycho-Gaia astrometric solution (TGAS; Gaia Collaboration *et al.*, 2016a), where the typical uncertainty is about 0.3 mas for the positions and parallax, and about 1 mas yr^{-1} for the proper motions (Gaia Collaboration *et al.*, 2016a). CEREAL uses TGAS as a training set to evaluate the functionality of the code.

The second data set was the Gaia Data Release 2 (Gaia DR2), which became available in April 2018 and was based on the data collected during the first 22 months of the mission lifetime (Gaia Collaboration *et al.*, 2018b). Gaia DR2 represented a major advancement with respect to the work presented in Gaia DR1 (or TGAS; Gaia Collaboration *et al.*, 2016a), via the inclusion of a high-precision parallax and proper motion catalogue for over 1 billion sources, supplemented by precise and homogeneous multi band all-sky photometry and a large radial velocity survey at the bright end ($G \lesssim 13$; Gaia Collaboration *et al.*, 2018b). However, it is also important to note that Gaia DR2 does not incorporate any astrometric information from Hipparcos and Tycho-2 (ESA, 1997; Høg *et al.*, 2000; Lindegren *et al.*, 2018b; Perryman *et al.*, 1997), which made it independent from these catalogues. All the sources in Gaia DR2 are treated as single stars and thus representable by five astrometric parameters (Lindegren *et al.*, 2018b). The typical median uncertainty for sources with five-parameter astrometric solutions, in parallax and position, at the reference epoch J2015.5, is about 0.04 mas for bright ($G < 14$ mag) sources, 0.1 mas at $G = 17$ mag, and 0.7 mas at $G = 20$ mag. From the proper motion components the corresponding uncertainties are 0.05, 0.2, and 1.2 mas yr^{-1}, respectively (Lindegren *et al.*, 2018b).

The third data set represent a subset of the future Gaia data release 3 (Gaia eDR3) which became available in 2022. Gaia eDR3 contains astrometric and photometric data for $\sim$ 1.8 billion sources in the magnitude range $G = 3$ to 21 (Lindegren *et al.*, 2020, 2021), with uncertainties of 0.02−0.03 mas at magnitude $G = 9$ to 14 and around 0.5 mas at $G = 20$, for the parallax and uncertainties of 0.02−0.03 mas yr^{-1} for G<15 and 0.5 mas yr^{-1} at $G = 20$ for the proper motion (Lindegren *et al.*, 2020, 2021). In this thesis, an analysis with Gaia eDR3 was not performed, since this data was available after the analysis made in this study.

During the period of this work, `CEREAL` was a code under continued development via multiple updates. The first version implemented worked with Gaia DR1 (TGAS); later versions worked with Gaia DR2 data[1].

This chapter is organised as follow: section 2.1 will describe how the data was gathered from the Gaia archive and then pre-processed before `CEREAL` evaluated the presence of clusters in the sample. Section 2.2 describes an example of the analysis made by `CEREAL` for a group of known open clusters, chosen from the literature.

[1]The current version of `CEREAL` is available at https://github.com/yumiry/CEREAL

2.1 How does the algorithm work?

2.1.1 PRE-CEREAL

The data for each target was gathered directly from the Gaia archive using the astroquery Gaia package (Segovia, 2016) and a search radius angle; which was defined as $angle = (2.5/Distance) * (180/\Pi)$, where the value 2.5 refers to the radius of an open cluster for which the typical diameter was 5 pc (Janes, 2001). The distance was obtained from Vioque *et al.* (2018).

Vioque *et al.* (2018) calculated the distance following the methodology of Luri *et al.* (2018) on how to manage the parallax from Gaia, and estimate the distance by applying a simple exponentially decreasing prior (Vioque *et al.*, 2018, references therein). For those stars without a distance reported, the $angle$ was set equal to one degree; using one degree as the default value allows the selection of all the possible clusters members around the HAeBe star.

In addition, a selection was applied to "clean" the sample from the Gaia DR2 catalogue. The method used is called *re-normalized unit weight error* (RUWE). It was previously defined in Lindegren *et al.* (2018a). The targets were selected with RUWE$<$ 1.40, which improved the selection of data and provides a smaller, but more reliable, sample. RUWE has been demonstrated previously to be a reliable indicator of the astrometric solution.

After gathering the data for each target in the Gaia archive, a final file was created with the essential information for the objects, such as position, parallax, proper motion, magnitude, etc; that would be used by CEREAL to begin the analysis of the data.

2.1.2 CEREAL

CEREAL begins the inspection of the data, for each of the stars to be analyzed, by displaying different plots to represent the data from Gaia; as shown in figure 2.1, where selections over the parallax and proper motions, within the input data, were made to find possible clusters around objects.

In figure 2.1 the top panel presents the histogram distributions of the parallax (ϖ; mas) and proper motion in both RA (μ_{α^*}; mas/yr) and DEC (μ_δ; mas/yr); and the density profile distribution for the given star. The density profiles were constructed by measuring the local density of the source in an annular radius from the position of the target, as

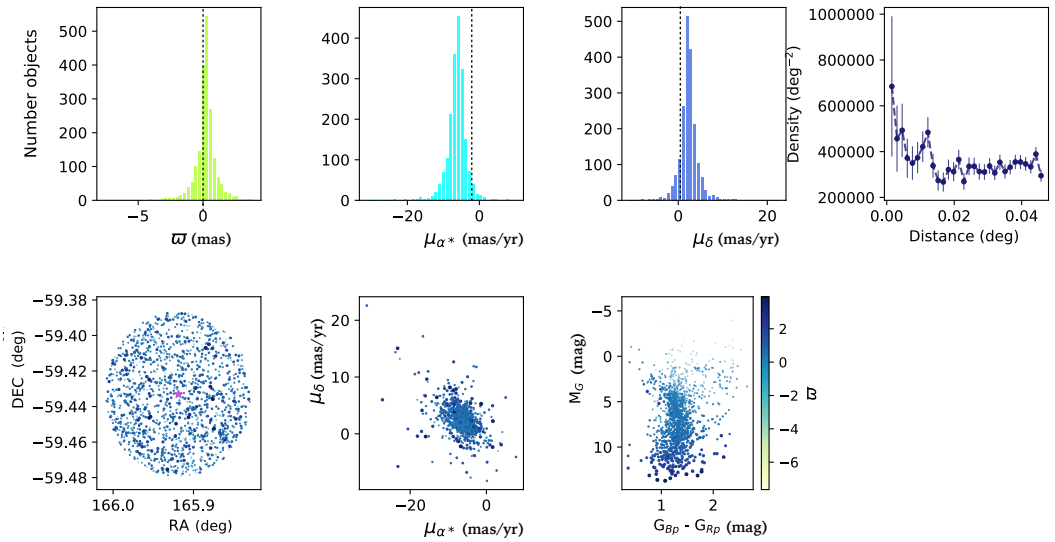


Figure 2.1: Example of CEREAL plots. This shows the input data from Gaia DR2 selected with a search radius angle (describe in § 2.1.1) of 0.046 degrees around the Herbig Ae/Be star HD96042. The top panel presents the histograms for the parallax, proper motion in RA and proper motion in DEC and the density profile distribution. The black dashed vertical line shows the known value of the parallax and proper motion of the Herbig Ae/Be star. The bottom panel presents the spatial distribution, the proper motion distribution and the colour-magnitude diagram (CMD). The parallax values are presented via the colour of the objects in the lower graphs, where the coloured bar to the right provides the scale for this. The purple stars represent the Herbig Ae/Be star position in each bottom figure, which might be hidden behind the small coloured dots.

was done by Testi *et al.* (1997). The bottom panel, presents the spatial distribution, the proper motion distribution and the Colour-Magnitude diagram (CMD).

The assumption made is that for a cluster the low mass companions around a HAeBe star will share similar properties, like their parallax and proper motion, with the rest of the cluster members. Therefore, it is expected that the presence of a cluster can be observed by finding a distinctive broad peak around the known value of the parameter; where, the known values were the literature value of each parameter for the HAeBe star. Figure 2.2 gives a representative example of peaks in parameters around the known value from the HAeBe star when a cluster was present.

The selection process is started by setting the limits around the parallax and proper motion using knowledge of the known values for each parameter to set the ranges. These

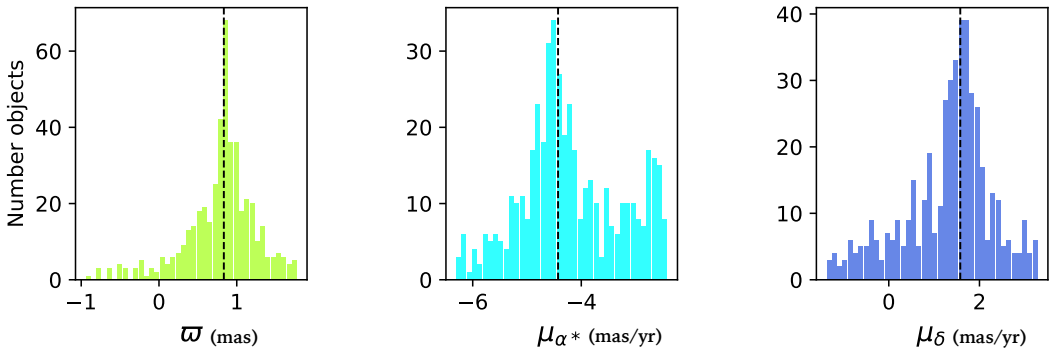


Figure 2.2: Example of peaks in parameters around the known value from the HAeBe star when a cluster was present. The panels show the histograms for the parallax, proper motion in RA and proper motion in DEC. The black dashed vertical line shows the known value of the parallax at 0.83 mas and proper motion in RA and DEC at -4.42 mas/yr and 1.57 mas/yr, respectively, of the Herbig Ae/Be star HUCMa.

ranges are interactively chosen by the user of the code using their judgment. The procedure starts by analysing the parallax, followed by the proper motion in RA, and finally, the proper motion in DEC. The evaluation of the presence of clusters around an object is then continued until the user of the code is satisfied with the results.

The presence of clusters can be inferred when representative peaks are visible in the histograms of the parallax and proper motions, like in figure 2.2. Additionally, the presence of a cluster can be inferred, for example, when a group of stars (the possible low mass companions of the Herbig Ae/Be star) are located nearby the Herbig Ae/Be star and share similar parallax to it; or, by using the density profiles (Testi *et al.*, 1997).

At the end of the evaluation of the target, the user can make decisions on whether the target is in a cluster or not, before starting the analysis of the next object. For this, a system was created to classify the targets whereby: if the star is considered to be in a cluster the flag is set equal to 1, which means *yes*; if the flag is equal to 2, this means *maybe*, the star could potentially be in a cluster; and if the flag is equal to 0, this means *no*, the star is not in a cluster. The code created a file for each star with all the data from the selections.

In addition, the mean value of each parameter was also reported with its respective associated error. The error was calculated using the weighted mean equation (Wall & Jenkins, 2003). The weighted mean equation is defined as $\bar{X}_w = \sum_{j=1}^{n} \frac{w_j * X_j}{\sum_{j=1}^{n} w_j}$, where the

weights are given by $w_j = \frac{1}{\sigma_j^2}$ and σ_j^2 is the reciprocal of the sample variances. The best estimation of the variance of the weighted mean equation is $\sigma_w^2 = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} \longrightarrow \sigma_w^2 = \frac{1}{w_j}$ (Wall & Jenkins, 2003).

A schematic representation of the step-by step process on how `CEREAL` finds clusters around the objects of interest using Gaia data is shown in the appendix A.

2.2 Studying known clusters using Gaia DR1 data

It was not possible to evaluate the presence of clusters around HAeBe stars using Gaia DR1, because Gaia DR1 only contained an astrometric dataset for ∼2 million stars (Gaia Collaboration *et al.*, 2016a) with which it was not possible to find lower mass companions around the HAeBe stars; Gaia DR1 did not have a sufficiently complete catalogue that covered the possible clusters members around the HAeBe stars, in particular this catalogue is incomplete at the bright end and has an ill-defined faint magnitude limit and these values depended on the position of the objects (Gaia Collaboration *et al.*, 2016a). This would be possible with the future data release Gaia DR2 which offered an independent astrometric dataset from Gaia DR1 for 1 billion sources (Gaia Collaboration *et al.*, 2018b).

To evaluate the ability of `CEREAL` to asses the presence of clusters, a sample of known clusters was studied with data taken from Gaia DR1 (specifically TGAS, Gaia Collaboration *et al.*, 2016a).

The sample of known open clusters was chosen randomly from the literature (See table 2.1; Gaia Collaboration *et al.*, 2017) for which their positions, parallaxes and proper motions were known, as well as the total number of members found by previous studies (Gaia Collaboration *et al.*, 2017).

Table 2.1: Parameters of known clusters

Cluster	RA	DEC	ϖ	μ_{α^*}	μ_δ	nMemb
	deg	deg	mas	mas yr^{-1}	mas yr^{-1}	
NGC6475	268.530	-34.849	3.57 ± 0.02	3.10 ± 0.06	-5.32 ± 0.04	78
Blanco1	0.855	-30.079	4.34 ± 0.11	18.20 ± 0.12	2.66 ± 0.11	43
IC2602	159.809	-64.496	6.74 ± 0.05	-17.67 ± 0.09	11.06 ± 0.13	66
α Per	52.069	49.060	5.91 ± 0.03	23.06 ± 0.06	-25.36 ± 0.07	116

Data taken from table 6 in Gaia Collaboration *et al.* (2017).

A step-by-step manual selection process was performed on the TGAS data for the selection of known cluster members. To begin the evaluation with `CEREAL`, data was selected from TGAS in a circular area with a radius of $\sim$3 degrees around the central position of the cluster. This large radius was used to ensure that all the possible members of the cluster were chosen. However, this introduced a contamination of the sample through the selection of foreground and background stars.

During the selection process with `CEREAL` it was expected that a representative peak around the known value of each parameter would be found. Those known value were obtained from Gaia Collaboration *et al.* (2017) and are presented in table 2.1.

The cluster NGC6475 is a well known cluster and as such it makes a good example case to test the code. NGC6475 has an estimated age of 200 to 260 Myr (van Leeuwen, 2009, references therein) and a distance of 280 pc (Gaia Collaboration *et al.*, 2017). van Leeuwen (2009) determined the cluster members through their proper motions and radial velocities.

Figures 2.3, 2.4 and 2.5 show the progression of the selection process over the parallax and proper motion used to find the cluster candidates for NGC6475 with `CEREAL`. In these examples, it can be seen that as each selection was made, the number of objects was reduced, whilst those objects that remained were distributed as a distinctive peak about the known value. It should be noted, that in some cases when the number of objects was reduced there was not an easily distinguishable peak, a cluster could still be found using the proper motion distribution or the density profile.

In addition, these figures show a characteristic peak at zero in both distributions of parallax and proper motions. This appears in all the input samples because regardless of how the data is selected, there are more objects at large distances and as a consequence, more objects with smaller values for the parallax and proper motions.

The mean values found by `CEREAL` for each parameter of the cluster NGC6475 were 3.60 ± 0.04 mas for the parallax, 3.26 ± 0.01 mas yr^{-1} for the proper motion in RA and -5.37 ± 0.01 mas yr^{-1} for the proper motion in DEC. These results were in good agreement with those reported by Gaia Collaboration *et al.* (2017, see table 2.1), where the values found were 3.57 ± 0.02 mas for the parallax, 3.10 ± 0.06 mas yr^{-1} for the proper motion in RA and -5.32 ± 0.04 mas yr^{-1} for the proper motion in DEC.

`CEREAL` found 84 stars in the cluster NGC6475 compared with 78 stars reported by Gaia Collaboration *et al.* (2017), of these, 67 stars were found in common in both

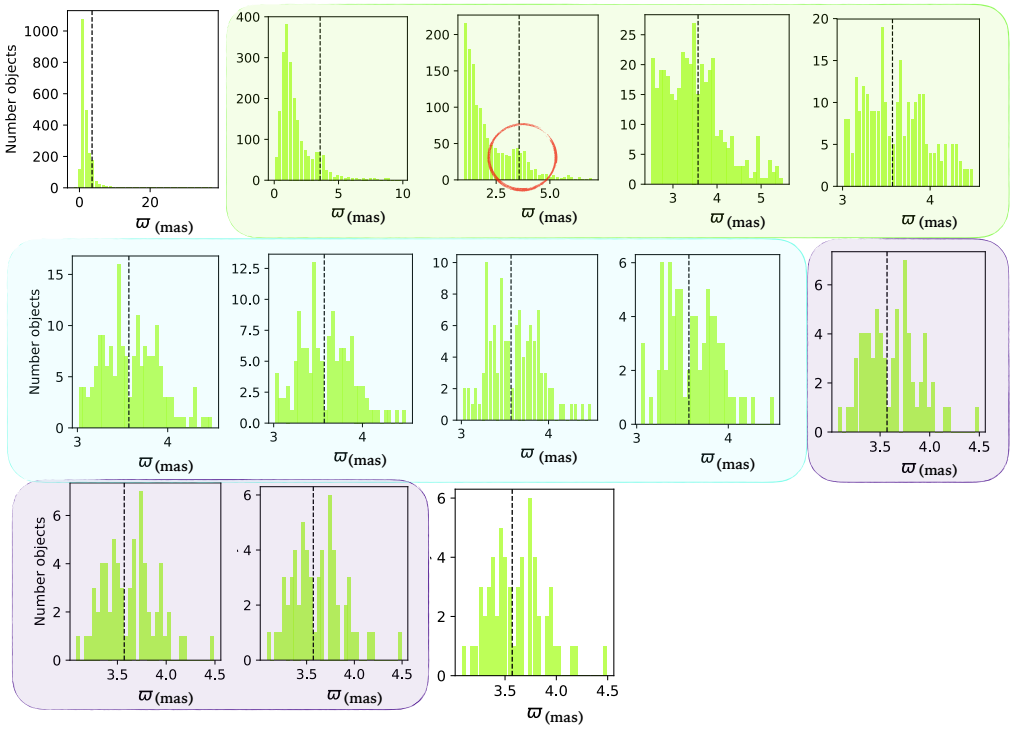


Figure 2.3: Parallax selection steps for the cluster NGC6475 with the TGAS data. The black dashed vertical line shows the known value for the parallax; and the red circle shows the first hint of the peak around the known value. Each coloured box represents how the parallax was affected during the different stages in the selection process made by CEREAL. The coloured boxes represent the selection made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple).

catalogues. Similar results were obtained for the other clusters analysed. CEREAL found 116 stars for the cluster αPer, 42 stars for Blanco1, 66 stars for IC2602; compared with the 116, 43 and 66 stars respectively found by Gaia Collaboration *et al.* (2017). Again, most of these stars were found to be in common in both catalogues, 88 stars for the cluster αPer, 34 stars for the cluster Blanco1 and 53 stars for the cluster IC2602.

Although the results are similar for all the clusters, there are some differences. The differences between those reported by Gaia Collaboration *et al.* (2017), and those found here by CEREAL might be due to the selection process used for each study. The stars were selected with CEREAL by choosing ranges around the known values for each parameter. Gaia Collaboration *et al.* (2017) estimated the cluster membership based on all the

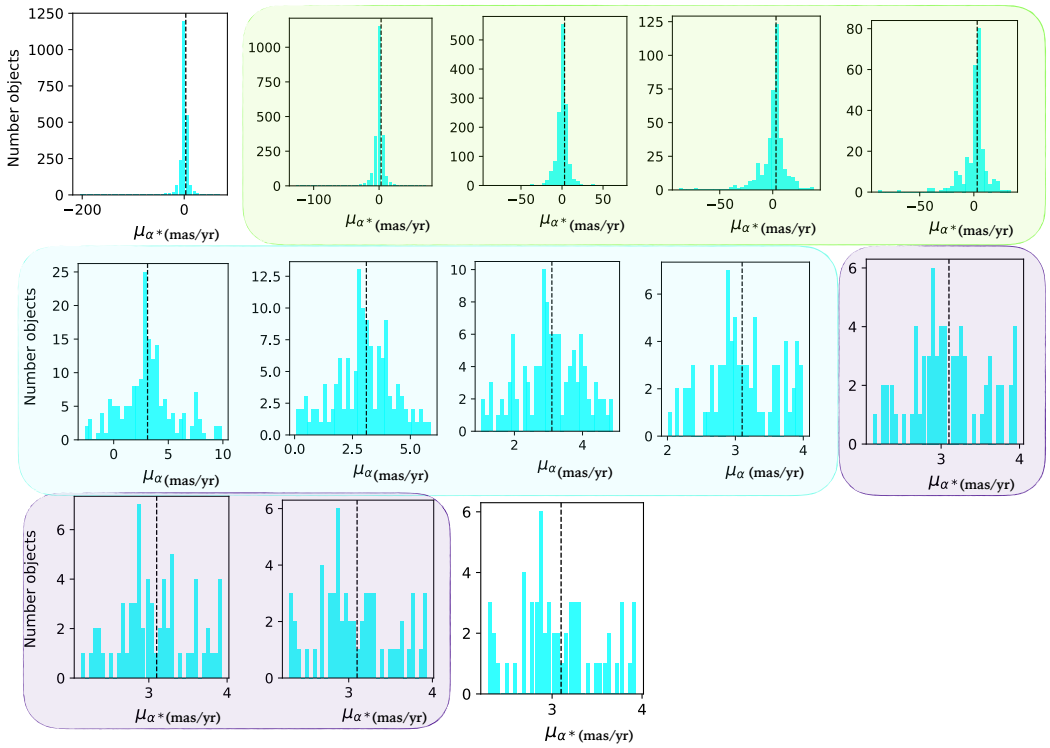


Figure 2.4: Proper motion in RA selection steps for the cluster NGC6475 with the TGAS data. The black dashed vertical line shows the known value for the proper motion in RA; and each coloured box represents how the proper motion in RA is affected during the different stages in the selection process made by CEREAL. The coloured boxes represent the selection made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple).

coincidences within a volume around the position of the distance (which is calculated as $\frac{1}{\varpi}$) of the cluster centre; where the main uncertainties came from the parallax measured. For more details see section 3 and appendix C from Gaia Collaboration *et al.* (2017).

Table 2.2 compares the results obtained by CEREAL with those obtained by Gaia Collaboration *et al.* (2017) and van Leeuwen (2009, using Hipparcos mission data). The difference between the results of the code and the literature for each parameter were very small and this is shown in figure 2.6. For example, for the cluster NGC6475 the difference between the value found by CEREAL and those reported by Gaia Collaboration *et al.* (2017) and van Leeuwen (2009) for the parallax were 0.03 ± 0.04 mas and 0.10 ± 0.14 mas, respectively; for the proper motion in RA these differences were 0.16 ± 0.06 mas/yr

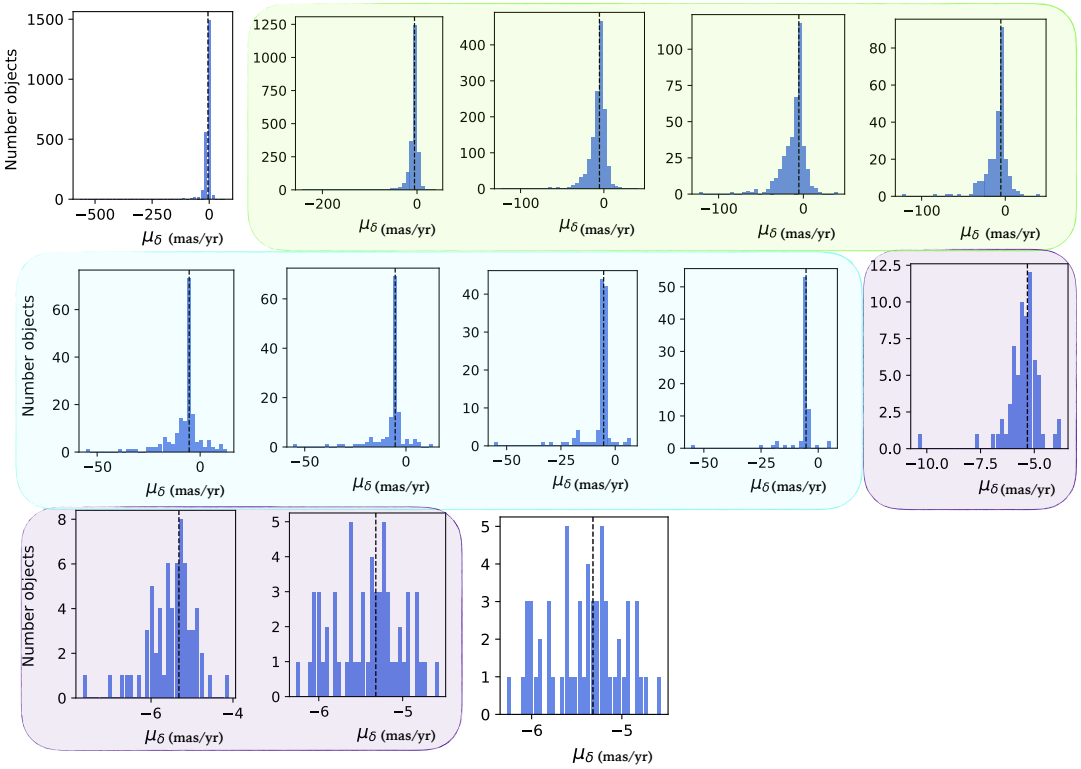


Figure 2.5: Proper motion in DEC selection steps for the cluster NGC6475 with the TGAS data. The black dashed vertical line shows the known value for the proper motion in DEC; and each coloured box represents how the proper motion in DEC is affected during the different stages in the selection process made by `CEREAL`. The coloured boxes represent the selection made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple).

and 1.20 ± 0.17 mas/yr, respectively; for the proper motion in DEC these differences were 0.05 ± 0.04 mas/yr and 0.39 ± 0.10 mas/yr, respectively.

The agreement obtained from the analysis above with the literature Gaia Collaboration *et al.* (2017); van Leeuwen (2009), showed that `CEREAL` could be an excellent tool for the selection of candidate members of a cluster, while it could also provide a good estimation of their parallax and proper motion.

Once Gaia DR2 was released, additional analysis was performed for the same known clusters but using the newer Gaia release. With the improvement of the survey, `CEREAL`

Table 2.2: Astrometric parameters of the known clusters

Cluster	ϖ mas	μ_{α^*} mas yr^{-1}	μ_δ mas yr^{-1}	Reference
NGC6475	3.57 ± 0.02	3.10 ± 0.06	-5.32 ± 0.04	Gaia Collaboration *et al.* (2017)
	3.70 ± 0.14	2.06 ± 0.17	-4.98 ± 0.10	van Leeuwen (2009)
	3.61 ± 0.04	3.08 ± 0.06	-5.42 ± 0.05	CEREAL†
Blanco1	4.34 ± 0.11	18.20 ± 0.12	2.66 ± 0.11	Gaia Collaboration *et al.* (2017)
	4.83 ± 0.27	20.11 ± 0.35	2.43 ± 0.25	van Leeuwen (2009)
	4.19 ± 0.06	18.62 ± 0.15	2.68 ± 0.15	CEREAL†
IC2602	6.74 ± 0.05	-17.67 ± 0.09	11.06 ± 0.04	Gaia Collaboration *et al.* (2017)
	6.73 ± 0.09	-17.02 ± 0.24	11.15 ± 0.23	van Leeuwen (2009)
	6.75 ± 0.05	-17.69 ± 0.15	10.77 ± 0.13	CEREAL†
α Per	5.91 ± 0.03	23.06 ± 0.06	-25.36 ± 0.07	Gaia Collaboration *et al.* (2017)
	5.80 ± 0.09	22.73 ± 0.17	-26.51 ± 0.17	van Leeuwen (2009)
	5.74 ± 0.04	23.10 ± 0.11	-25.25 ± 0.09	CEREAL†

† Errors were calculated using the standard error equation: SE $= \frac{StandardDeviation}{\sqrt{N}}$; where N is the total number of objects in the sample.

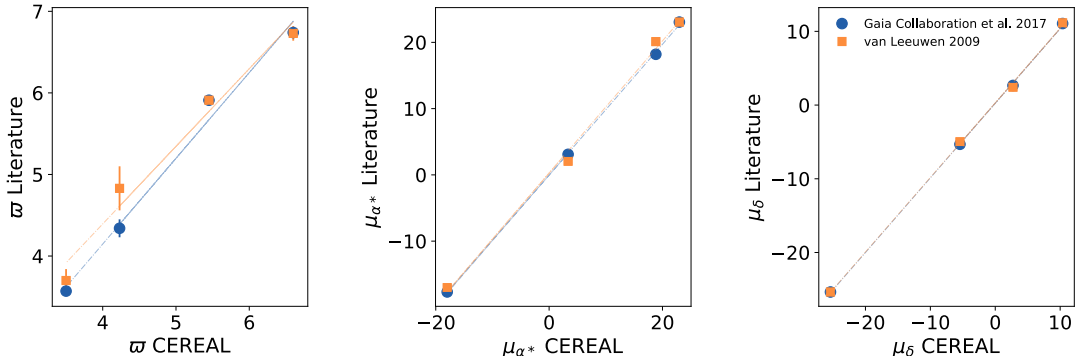


Figure 2.6: Comparison of the astrometric parameters from Gaia Collaboration *et al.* (2017), van Leeuwen (2009) and those obtained with CEREAL. The panel on the left shows the parallax, the panel in the middle the proper motion in RA and the panel on the right the proper motion in DEC. The straight line in the plots represents the least squares best fit of a line to the data points. These values are represented in Table 2.2.

was, again, able to find a large number of companions around the known clusters (230 stars around NGC6475, 198 stars around Blanco1, 157 stars around IC2602 and 302 stars around α Per) compared with the companions found with Gaia DR1 for the same known clusters (84 stars around NGC6475, 42 stars around Blanco1, 66 stars around IC2602 and 116 stars around α Per); however, these were not as large as the ones reported by Gaia Collaboration *et al.* (2018a, 952 stars around NGC6475, 493 stars around Blanco1, 490 stars around IC2602 and 745 stars around α Per). These differences may again be due to the specific selection process used by each method.

It was clear, that with the update of the survey, that `CEREAL` would not be restricted only to improving the results for the analysis of known clusters, but would have the ability to assess the presence of clusters around new targets.

2.2.1 Quantitative analysis

A detailed quantitative analysis of the over-densities in the samples was performed on the results obtained with `CEREAL` for the Gaia DR1 and Gaia DR2 datasets. This analysis was carried out on the known clusters and a subset of the Herbig Ae/Be stars that will be studied further in the following chapters.

3σ method

This analysis is focused on the determination of the radius at which the density of the stars is significantly higher than the density of the background (3σ). The density profiles obtained by `CEREAL` in the previous section are used as a starting point for this analysis. The analysis was only carried out using the stellar density profiles and did not include any of the astrometric parameters (parallax and proper motions) provided by Gaia. As a consequence the results determined in this section will differ from those presented in section §2.2.

The first step in the assignment of the radius of the over-density, was to evaluate the background density of each density profile of the stars. The background region was assigned as the area where the density appeared to be constant; the over-density (possible cluster) was expected to be located within the region where the density deviates from this constant value. Figure 2.7 shows an example of the possible location of the background and the over-density (possible cluster) for the known cluster αPer.

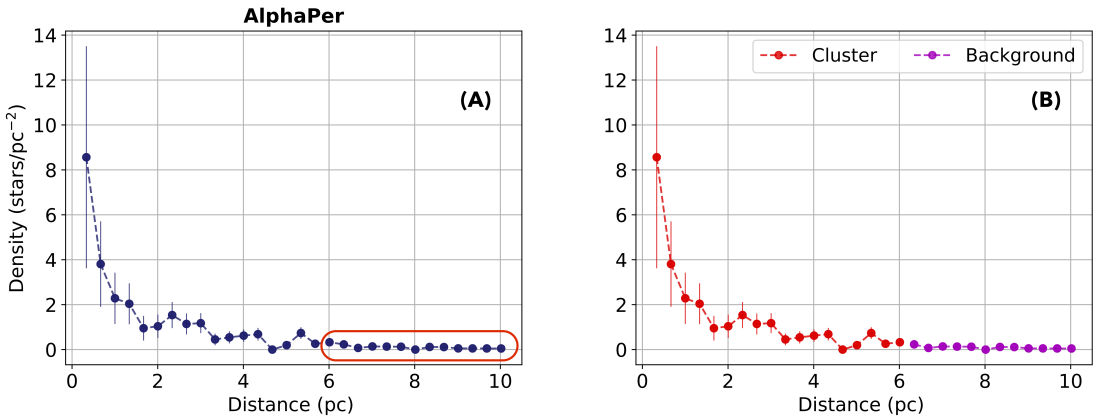


Figure 2.7: Background selection with Gaia DR1 data for the known cluster αPer. Panel (A) shows the density profile obtained by `CEREAL` previously (dark blue filled circles) where the possible location of background is shown in the red box. This area, which starts at $\sim$ 6pc, was used in the first evaluation of the background. Panel (B) is the same density profiles where the anticipated location of the cluster (light red filled circles) and background (purple filled circles) are shown.

The background density was then calculated by adding all the stars located within the selection region (e.g. each point inside the red box in figure 2.7(A)) and dividing these by the area.

A comparison of the density profiles and the background was then carried out to evaluate where the over-density (possible cluster) rose 3σ above the background. This led to the observation of a limitation to this approach (relying only on density profiles to assign the presence of a cluster) as the uncertainties, in the densities within each ring area, was generally too large to assign the presence of an over-density even for known clusters. This is because the stellar densities within these ring areas are based on only a small number of observed objects and therefore, despite the uncertainty in the background density being generally very low, it was not possible to use the density profiles of the form shown in figure 2.7(B) for the statistical analysis of the location of the cluster radius.

To be able to find the value for the radius of each cluster in the sample, it was necessary to find a different way to compare the density of the background with the density of the cluster. To do this it is important to increase the number of objects (and therefore improve the statistical error) within the regions that were to be compared to the background.

To achieve this, a similar process to the one used to evaluate the density of the background was applied. The area of the density profile where the possible cluster is located (red full points in figure 2.7(B)) is used to generate new subgroups (cluster apertures) that are going to be compared with the background density. These new subgroups were generated by sequentially moving point by point outwards from the centre and summing all the objects together and dividing by the new area within which they were found. These new larger sub groups contained more objects than the original ring areas and as such had smaller statistical errors associated with their densities. The new stellar densities found for these subgroups could then be compared to the background densities, to evaluate if any of these sub groups showed a statistically significant difference (3σ) from the background value.

Table 2.3 presents the results obtained for all the subgroups generated for the known cluster αPer. The table shows the subgroup (cluster apertures) created for the cluster areas in the density profile (red full points in figure 2.7(B)), the distance associated to each subgroup, the number of stars found in these subgroups and the density they represent.

The results displayed in table 2.3, for αPer, show that as the subgroups increase in size there are several subgroups that contain densities greater than the background at the 3σ level. All the sub groups with distances between 1.0 and 5.7 pc show a stellar density significantly greater than the background density. This result shows that there is a large over density present in this profile, with approximately 70 possible cluster members observed.

The outcome of this analysis did not provide a clear value that could be used to define an exact radius for each cluster by using these subgroups. Due to multiple sub groups being found which meet the condition of being statistically significantly higher than the background density (3σ level).

The same behaviour repeated for the other known clusters in the sample. Figure 2.8 and table 2.4 shows the same analysis made previously for NGC6475. For this known cluster it was also found that as the subgroups increase in size there are several subgroups that contain densities greater than the background at the 3σ level. The stellar density is significantly greater than the background density between the distance of 1.49 and 5.97 pc, with 42 possible cluster members observed.

Additional tests were carried out to evaluate if the background region influenced the results obtained from this analysis. The background density was recalculated by adding a

Table 2.3: Cluster apertures and densities for the known cluster αPer with Gaia DR1 data

Area	Cluster aperture pc^2	Distance pc	Number of Stars	Density stars/pc^{-2}
(1)	(2)	(3)	(4)	(5)
Background	201.81	6.01	18	0.09 ± 0.02
Cluster	0.35	0.33	3	8.56 ± 4.94
	1.4	0.67	7	4.99 ± 1.89
	3.15	1.0	11	3.49 ± 1.05
	5.61	1.34	16	2.85 ± 0.71
	8.76	1.67	19	2.17 ± 0.5
	12.61	2.0	23	1.82 ± 0.38
	17.17	2.34	30	1.75 ± 0.32
	22.42	2.67	36	1.61 ± 0.27
	28.38	3.01	43	1.52 ± 0.23
	35.04	3.34	46	1.31 ± 0.19
	42.39	3.67	50	1.18 ± 0.17
	50.45	4.01	55	1.09 ± 0.15
	59.21	4.34	61	1.03 ± 0.13
	68.67	4.68	61	0.89 ± 0.11
	78.83	5.01	63	0.8 ± 0.1
	89.69	5.34	71	0.79 ±0.09
	101.26	5.68	74	0.73 ± 0.08

Columns (3) represent the distance from the centre of the cluster.

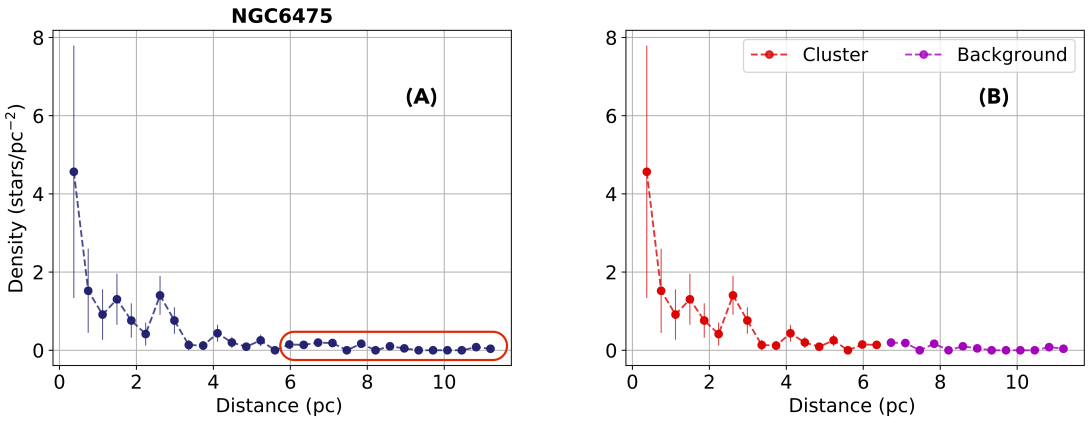


Figure 2.8: Background selection with Gaia DR1 data for the known cluster NGC6475. See description of figure 2.7.

Table 2.4: Cluster apertures and densities for the known cluster NGC6475 with Gaia DR1 data

Area	Cluster aperture pc^2	Distance pc	Number of Stars	Density stars/pc^{-2}
(1)	(2)	(3)	(4)	(5)
Background	267.66	6.35	15	0.06± 0.01
Cluster	0.44	0.37	2	4.51± 3.23
	1.75	0.75	4	2.23± 1.14
	3.94	1.12	6	1.47± 0.62
	7.01	1.49	10	1.37± 0.45
	10.95	1.87	13	1.13± 0.33
	15.77	2.24	15	0.9± 0.25
	21.47	2.61	23	1.02± 0.22
	28.04	2.99	28	0.94± 0.19
	35.48	3.36	29	0.76± 0.15
	43.81	3.73	30	0.63± 0.13
	53.01	4.11	34	0.59± 0.11
	63.08	4.48	36	0.51± 0.1
	74.03	4.85	37	0.44± 0.08
	85.86	5.23	40	0.41± 0.07
	98.57	5.6	40	0.35± 0.06
	112.15	5.97	42	0.32± 0.06

Columns (3) represent the distance from the centre of the cluster.

reduced selection of stars located within the selection region where the density appeared constant (e.g. points inside the red show in figure 2.7(A), from 7 to 10 pc and from 8 to 10pc) and dividing these by the area. As before, the area of the density profile where the possible cluster is located was used to generate new subgroups that were then compared with the background density.

Following the same procedure described previously, it was again found that multiple subgroups (cluster apertures) meet the condition of being statistically significantly higher than the background density (3σ level). This means that within the work carried out here, the size of the selected background region does not have any influence on the evaluation of the cluster radius.

Appendix B shows that for the other known clusters in the sample the same pattern of multiple subgroups which meet the condition of being statistically significantly higher than the background density (3σ level) was also found. Additionally, appendix B shows an example of the same analysis carried out in this section applied to the HAeBe stars.

Rolling window method

Following this assessment and the observation that by increasing the areas to include more objects it was possible to generate areas that had densities that were statistically significantly over the background level another method was considered. In this new test the decision was made to pass rolling 'windows' across the previous density profiles with the window containing a variable number of the previous ring areas. For the points in the density profiles that fell within the window, the ring areas and number of objects found were added together and this was then used to calculate the density and error in the density of this larger region. The window was then moved iteratively away from the central star across the original density profiles, adding the next ring area into the window and removing the ring area closest to the central star from the window. For example, in this analysis a window size of 1 corresponds to the unchanged density profiles and a window size of 3 represents the combination of the adjacent ring areas in the original unchanged density profiles. Therefore, the first point in the new profiles generated with a window size of 3 contains the sum of the first 3 ring areas in the original density profiles. The second point contains the sum of the second third and fourth ring areas in the original density profiles and this was then iteratively continued outwards until the final point in the new profile generated with a window size of three containing the sum of the last 3 ring areas in the original density profiles. Windows of size 1 to 10 were passed over the data and used to generate new density profiles; the window size refers to the number of the original ring areas included within the window.

An example of how the new density profiles for the known clusters αPer and NGC6475 generated with different window sizes appear is shown in figure 2.9. These new density profiles were then compared to the background density to evaluate where the first point that rose 3 sigma over the background was located; the location of all the points to rise 3 sigma and 2 sigma over the background are shown as a red full circle and open black squares for each of the window sizes in figure 2.9. The location of the points that rise 2 and 3 sigma over the background were tabulated for each of the know clusters (and Herbig Ae/Be stars) for window sizes 1 to 10; this data is provided in appendix B. This was done to see if once windows of sufficiently large size were used the location of the furthest point at 3σ from the cluster centre was well defined or if the location depended strongly on the window size chosen.

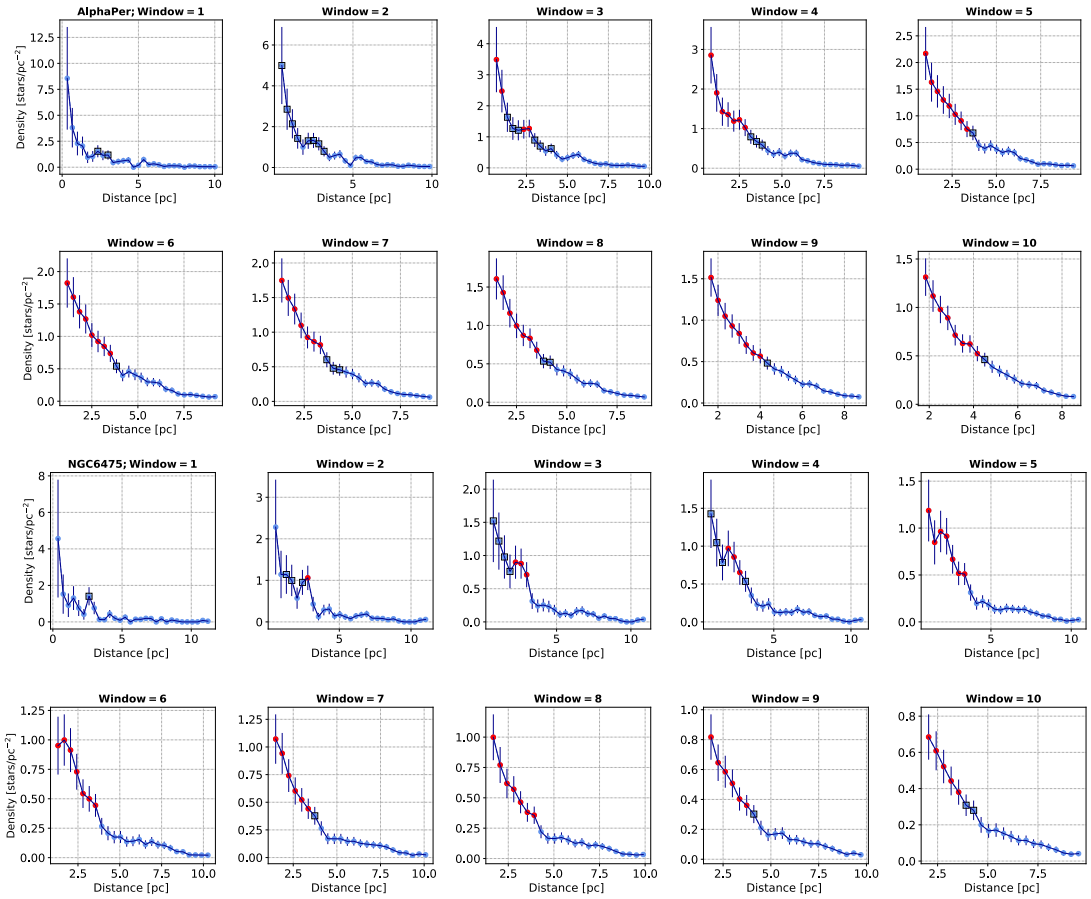


Figure 2.9: Density profile created with different window of sizes for the known cluster αPer (top panel) and NGC6475 (bottom panel). Each panel show the density profile created with the different window of sizes from 1 to 10. The red full points and open black squares show the location of the all the points that rise 3 and 2 sigma over the background, respectively.

For the known clusters evaluated here (and the sample of HAeBe stars evaluated in appendix B) the location of the points that rise over the background were not found to be consistent for all the windows sizes. The changes to the assigned cluster radius from the individual window sizes was different for different clusters. For some of the clusters (e.g. αPer) the cluster radius assigned gradually increased with the size of the window, for these clusters the change in the cluster radius assigned was generally not significantly different once the error in these radii were included (see figure 2.10). A different dependence of the assigned cluster radius on the window size is present for other clusters such as IC2602,

where there is a step change in the assigned cluster radius. These step changes are caused where several groups of density that alone did not rise above the 3σ threshold with smaller windows are then combined within a larger window. For some clusters this step change in assigned cluster radius is statistically different (differ beyond combined errors) from the radius assigned with smaller windows (e.g. for the Herbig Ae/Be star BD+30549, see figure B.6 in appendix B).

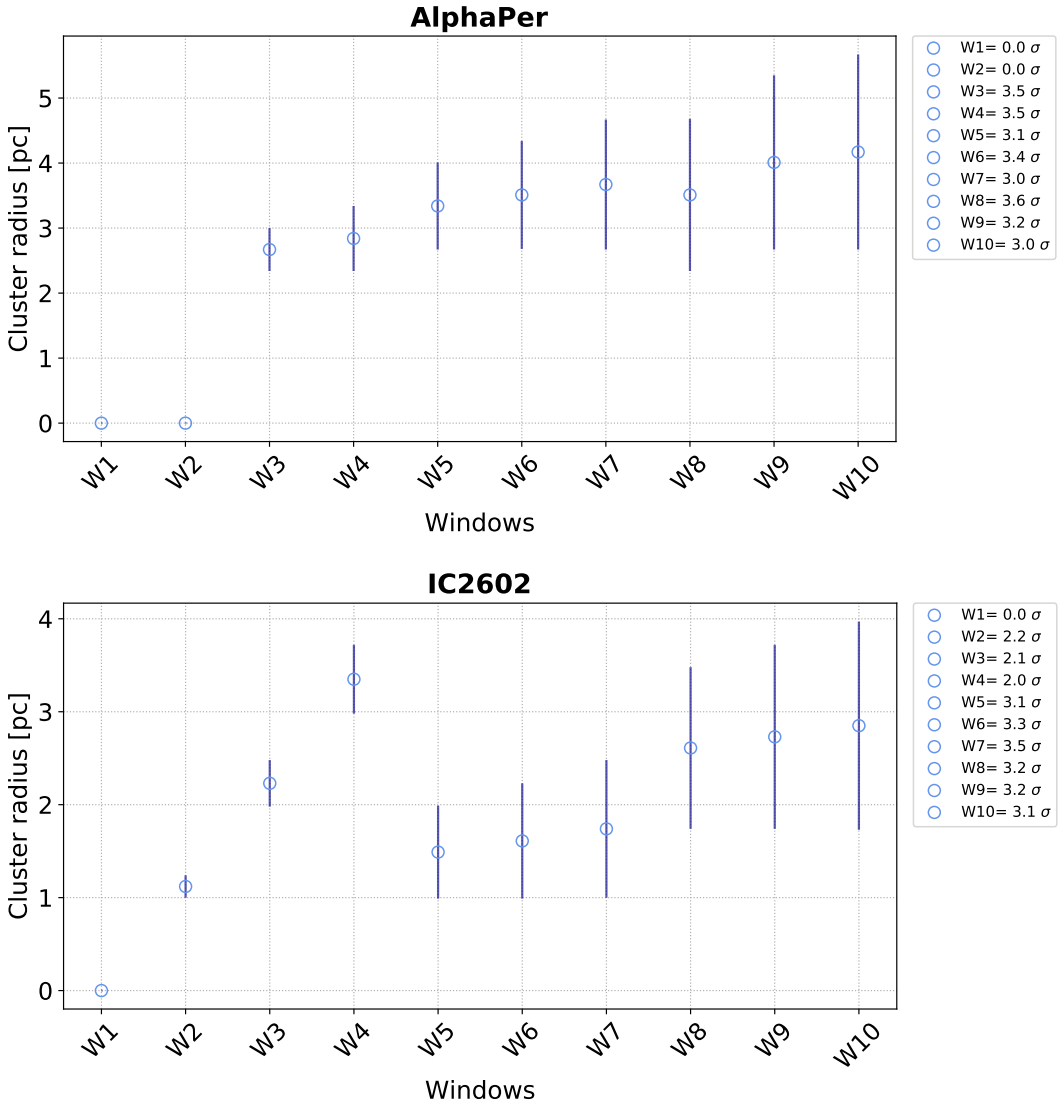


Figure 2.10: Cluster radius assigned for the different window sizes for the known cluster αPer (top panel) and IC2602 (bottom panel). Each panel show the different window sizes from 1 to 10 versus the cluster radius find for each window size.

Where the changes in the assigned cluster radius with the window size fell within their combined errors, the cluster radius found from the smallest window size to rise 3 sigma above the background density was used in the further analysis. For the clusters where a step change in the assigned cluster radius larger than the combined errors was observed the cluster radius was assigned from the smallest window size following the change in the radius to a new consistent position. For example the radius was assigned using a window size of 6 for the Herbig Ae/Be star BD+30549 as the assigned radius for windows of size 6-10 were consistent within their errors where the radius found with windows of size 4 and 5 differed, see figure B.6 in appendix B. For clusters where no densities rose 3σ above the background with window sizes 1-10, then the first point to meet the 2σ threshold was considered instead.

The results of this rolling window analysis are summarised in table 2.5 for the known clusters. Additionally, table B.5 in appendix B shows the full results obtained for all the window sizes for each known cluster (and HAeBe stars) evaluated using this method. The cluster radius found in this section were then used in the further analysis of the clusters (such as in chapter §5) that have been detected by CEREAL.

Additionally, the results obtained by CEREAL for the quantitative analysis are compared to the analysis in table 2.2 from section 2.2. This comparison is shown in table 2.6 where CEREAL and other studies found similar values for the kinematic parameters of the known clusters for the TGAS data. These results demonstrated that CEREAL is a good tool to analyse the presence of clusters and find their possible members.

Table 2.5: Summary of the cluster radius found using the windows method for Known clusters

Object	Cluster Radius pc	Sigma Level	Window size
AlphaPer	2.67 ± 0.33	3.47	3
Blanco1	0.97 ± 0.19	3.23	2
IC2602	1.49 ± 0.50	3.14	5
NGC6475	2.80 ± 0.19	3.13	2

Table 2.6: Astrometric parameters of the known clusters, adding quantitative analysis results.

Cluster	ϖ mas	μ_{α^*} mas yr^{-1}	μ_δ mas yr^{-1}	Reference
α Per	5.91 ± 0.03	23.06 ± 0.06	-25.36 ± 0.07	Gaia Collaboration *et al.* (2017)
	5.80 ± 0.09	22.73 ± 0.17	-26.51 ± 0.17	van Leeuwen (2009)
	5.74 ± 0.04	23.10 ± 0.11	-25.25 ± 0.09	CEREAL†
	5.74 ± 0.05	22.93 ± 0.13	-25.39 ± 0.10	CEREAL: Over 3σ - TGAS†
Blanco1	4.34 ± 0.11	18.20 ± 0.12	2.66 ± 0.11	Gaia Collaboration *et al.* (2017)
	4.83 ± 0.27	20.11 ± 0.35	2.43 ± 0.25	van Leeuwen (2009)
	4.19 ± 0.06	18.62 ± 0.15	2.68 ± 0.15	CEREAL†
	4.18 ± 0.07	18.68 ± 0.17	2.77 ± 0.15	CEREAL: Over 3σ - TGAS†
IC2602	6.74 ± 0.05	-17.67 ± 0.09	11.06 ± 0.04	Gaia Collaboration *et al.* (2017)
	6.73 ± 0.09	-17.02 ± 0.24	11.15 ± 0.23	van Leeuwen (2009)
	6.75 ± 0.05	-17.69 ± 0.15	10.77 ± 0.13	CEREAL†
	6.75 ± 0.06	-17.44 ± 0.13	11.03 ± 0.11	CEREAL: Over 3σ - TGAS†
NGC6475	3.57 ± 0.02	3.10 ± 0.06	-5.32 ± 0.04	Gaia Collaboration *et al.* (2017)
	3.70 ± 0.14	2.06 ± 0.17	-4.98 ± 0.10	van Leeuwen (2009)
	3.61 ± 0.04	3.08 ± 0.06	-5.42 ± 0.05	CEREAL†
	3.59 ± 0.04	3.16 ± 0.06	-5.43 ± 0.07	CEREAL: Over 3σ - TGAS†

† Errors were calculated using the standard error equation: SE = $\frac{StandardDeviation}{\sqrt{N}}$; where N is the total number of objects in the sample.

2.3 Final remarks

CEREAL is a simple, but powerful, algorithm which takes large data sets as an input and then retrieves objects which share similar astrometric parameters; like parallax and proper motion, with the known target.

With CEREAL the user has total control over the selection process of the sample, while with other algorithms the user needs to try different input parameters several times, to obtain a result. Moreover, the simplicity of CEREAL allows it to evaluate data without the need for making a pre-selection of the sample or the use of sophisticated techniques, that other algorithms which also search for clusters use.

This section also shows that there is not a simple method to estimate the value of

the cluster radius (§2.2.1). To assign a value for the cluster radius, in this study a rolling window method was used to evaluate the densities around each object in the sample.

In the future, an improved statistical assignment of the cluster radius could be achieved by combining all the data provided by Gaia, such as parallax and proper motions (see §2.2), to make a decision about how large the cluster might be. Since CEREAL has been designed as a visual program for the detection of clusters around stars using user defined selections across multiple parameters it does not provide the ability to carry out simultaneous statistical testing of multiple parameters; instead simultaneous visual assessment of the data across these multiple parameters is carried out by the user.

In the following chapter, CEREAL will use Gaia DR2 data to analyse a sample of known Herbig Ae/Be stars (mention in §2.2.1) to assess the presence of clusters around them. In addition, the ability of CEREAL at finding clusters will be compared with other algorithms which also evaluate the presence of overdensities in large samples, like Gaia DR2.

Chapter 3

Presence of clusters around Herbig Ae/Be stars

Let everything happen to you: beauty and terror.
Just keep going. No feeling is final.

— *The book of Hours,* Rainer Maria Rilke.

The results of the analysis performed on the selected known clusters in the previous chapter (see §2.2) are encouraging. These results demonstrated that CEREAL could be a useful tool for assessing the presence of clusters using Gaia data.

In this chapter, CEREAL will be used to investigate the presence of clusters around a sample of known HAeBe stars using Gaia DR2 data. The results of this evaluation will be used to create a catalogue, containing all the possible clusters discovered around the HAeBe stars. Additionally, the accuracy of CEREAL will be tested by comparing the clusters found using this method with those found by other algorithms, using the same sample of HAeBe stars. These algortihms are all based on machine learning codes.

Finally, this chapter addresses the question of whether CEREAL was able to find the same clusters found around the HAeBe stars by Testi *et al.* (1997, 1998, 1999).

3.1 Herbig Ae/Be star sample

A compilation of the known HAeBe stars from the literature was made. The sample which contains 269 objects was obtained, by a cross-match at 2 arcsec, from the Chen *et al.* (2016) and Vioque *et al.* (2018) catalogues.

For this sample, 234 stars were found in common in both catalogues and have Gaia DR2 data. The remainder was found in only one of the catalogues, 18 of these stars in the Vioque *et al.* (2018) catalogue and therefore had Gaia DR2 data. The remaining 17 stars were found only in the Chen *et al.* (2016) catalogue and do not all have Gaia DR2 data. The differences between the catalogues may be due to miss-classification of the objects, due to their spectral type, type of star, or the name given to the object.

From the Chen *et al.* (2016) catalogue, 14 stars do not have parallax and proper motions reported in Gaia DR2. For these objects, a search in the SIMBAD astronomical database (Wenger *et al.*, 2000, hereafter SIMBAD) was performed at 2 arcsec radius to verify the availability of their parallax and proper motions. The result obtained from the search in SIMBAD was that 7 stars (GSC5987-1399, HD203024, HD37411, HKOri, TCrA, V375Lac, V633Cas) have reported values for their proper motion in both RA and DEC by Høg *et al.* (2000); Zacharias *et al.* (2003, 2012). The remaining 7 stars (GSC4805-1306, Hen3-1475, MWC778, PDS394, PDS406, RMon, V376Cas) do not have parallax or proper motion values reported in SIMBAD.

In addition, of the 269 stars in the sample, 98 stars have been classified by The *et al.* (1994) as true HAeBe stars, or candidate members of the class.

The combination of the Chen *et al.* (2016) and Vioque *et al.* (2018) catalogues constitutes a new sample to study the clustering properties of the HAeBe stars. The tables with the main parameters for the 269 HAeBe stars is given in appendix C.

3.1.1 Spectral type

The spectral types were obtained from Chen *et al.* (2016) and Vioque *et al.* (2018). The sample of the 269 HAeBe stars cover the spectral type range between $O7$ to $G9$ which includes 7 (3%) O stars, 131 (49%) B stars, 98 (36%) A stars, 28 (10%) F stars and 5 (2%) G stars.

By definition HAeBe stars have spectral type B and A, and indeed, these are the dominant spectral types in this sample. For this analysis some early types, like O stars;

and, late types, like F and G have also been included. These last types, are also known as Intermediate Mass T Tauri stars (IMTTS, Calvet *et al.*, 2004), which are not well studied, when compared to T Tauri stars. The IMTTS are a subset of T Tauri stars with spectral types ranging from late F to early K and represent the link between the HAeBe stars and the Classical T Tauri Stars (Calvet *et al.*, 2004; Pérez-Blanco *et al.*, 2018). The advantage of including a wide range of spectral types within the sample is that it provides a more representative sample of intermediate mass stars than a sample of only HAeBe stars.

Unlike the spectral types, which were taken from Chen *et al.* (2016) and Vioque *et al.* (2018), the luminosity classes were taken from SIMBAD (Wenger *et al.*, 2000), Fairlamb *et al.* (2015) and Wichittanakom *et al.* (2020). These parameters are given in appendix C.

In the following sections, the properties of the sample will be described in more detail. The analysis focused on the stars with spectral types B and A and how their spectral type was related to their other parameters, such as parallax or G magnitude.

3.2 Detecting the presence of clusters around the Herbig Ae/Be stars using CEREAL

The same procedures applied in section 2.2 to the known clusters using CEREAL to analyse Gaia DR1 data, were applied to the sample of 269 Herbig Ae/Be stars using Gaia DR2 data.

The following sections show examples of the classification system created to define whether a target was found to be in a cluster or not (see §2.1). The system assigned a flag of *yes*, *maybe* or *no* if the star was considered to be in a cluster, potentially in a cluster or was not in a cluster, respectively.

3.2.1 *Yes*: an example where CEREAL assigned a candidate to be in a cluster

The analysis for each target with CEREAL starts when the code displays an image which contains the data obtained from the Gaia survey.

An example of the result obtained from the first iteration of CEREAL, for an object, is shown in figure 3.1. In this case, it displays the information for the HAeBe star MWC137, which is located in the cluster SH 2-666. Mehner *et al.* (2016) estimates that this cluster is at least 3 Myr old.

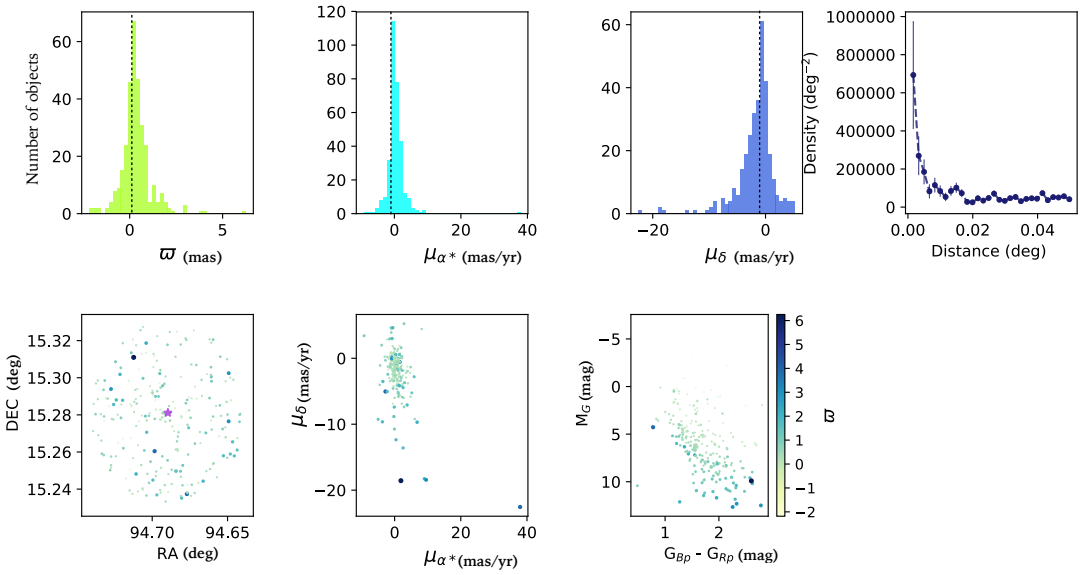


Figure 3.1: First iteration of CEREAL for MWC137. The histograms (top panel) represent the astrometric parameters and the black dashed vertical lines show the position of the known values. The density profile is also shown in the top panel. The bottom panel shows the spatial distribution, CMD and the proper motions with a colour code showing how the parallax values are distributed in the three figures. The purple star in the spatial distribution panel represents the position of the HAeBe.

As was shown in the previous section, for the known clusters (see §2.2), figures 3.2, 3.3 and 3.4 represent the different stages of the selection process using the parallax and proper motions (see appendix D for further examples). The peaks in the parallax and proper motion were expected to be found around the known values of those parameters; for this example these were: 0.11 ± 0.05 mas for the parallax, -0.26 ± 0.09 mas yr^{-1} for the proper motion in RA and at -0.41 ± 0.07 mas yr^{-1} for the proper motion in DEC.

Figure 3.5 shows the final iteration of the selection process made by CEREAL for the HAeBe star MWC137. At this stage, it is still possible that contaminating stars are present, from both the background and the foreground, because the selection is made using large ranges around the known values of the parallax and proper motions.

In the figure it can be observed that the distribution of the parallax and the proper motions appear evenly distributed around the known value. Whilst the density profile showed a smooth decrease in the density as a function of the distance. Also, the spatial and proper motion distributions showed that the companions around the HAeBe star

shared similar parallax values. Moreover, the final iteration of CEREAL had 31 stars in common with the study made by Mehner *et al.* (2016) of the stars.

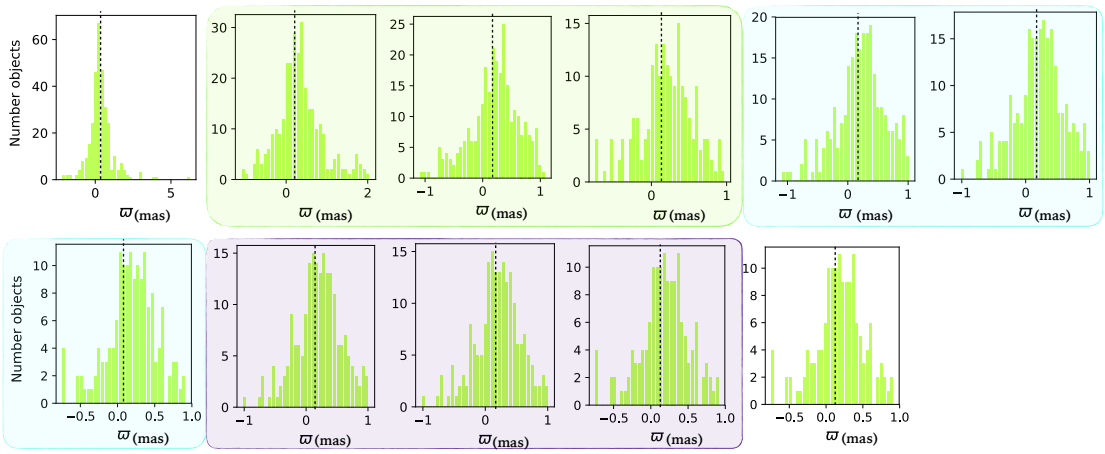


Figure 3.2: Parallax selection steps for MWC137 with the Gaia DR2 data. The expected peak for the parallax is around 0.10 mas. The black dashed vertical line shows the position of the known value of the parallax; and each colour box represents how the parallax is affected during the different stages in the selection made by CEREAL. The colour boxes represent the selection made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple).

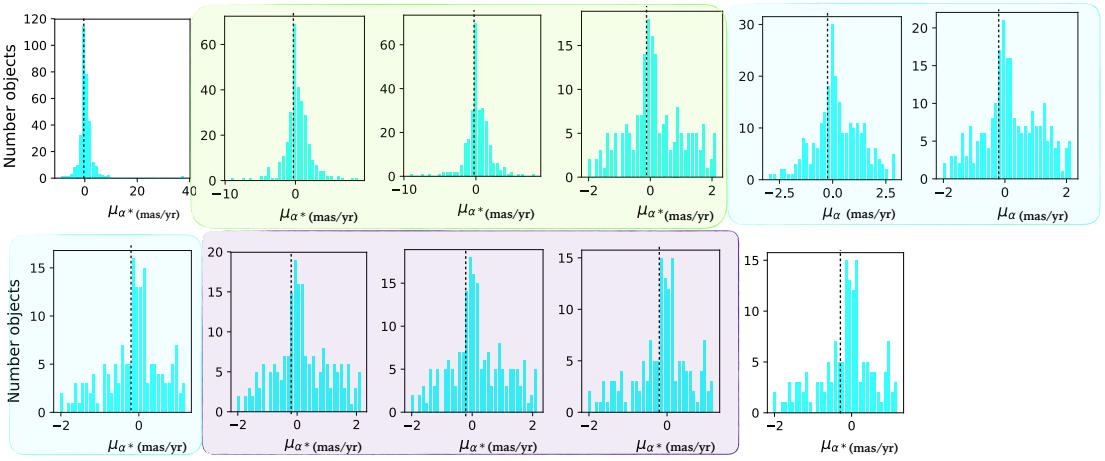


Figure 3.3: Proper motion in RA selection steps for MWC137 with the Gaia DR2 data. The expected peak for the proper motion in RA is around −0.26 mas/yr. The description of the vertical lines and colour boxes is the same as in figure 3.2.

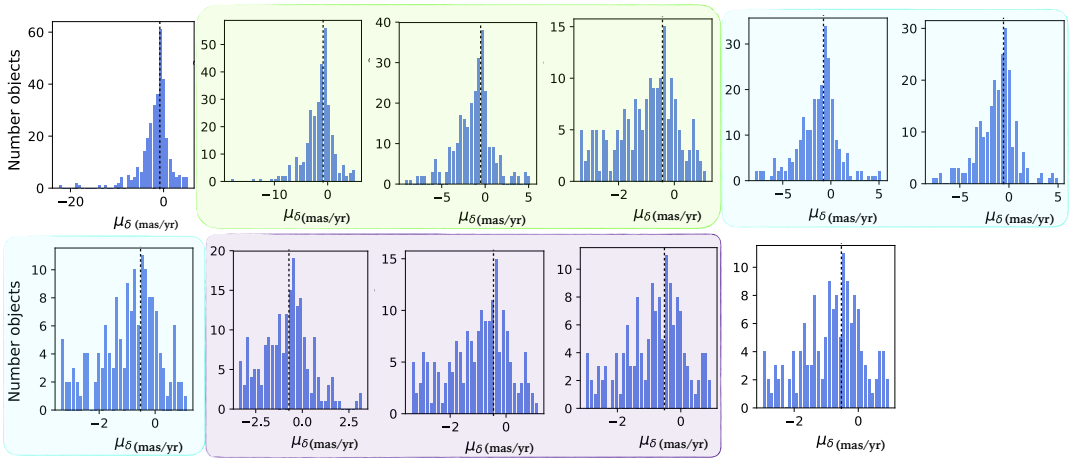


Figure 3.4: Proper motion in DEC selection steps for MWC137 with the Gaia DR2 data. The expected peak for the proper motion in DEC is around -0.40 mas/yr. The description of the vertical lines and colour boxes is the same as in figure 3.2.

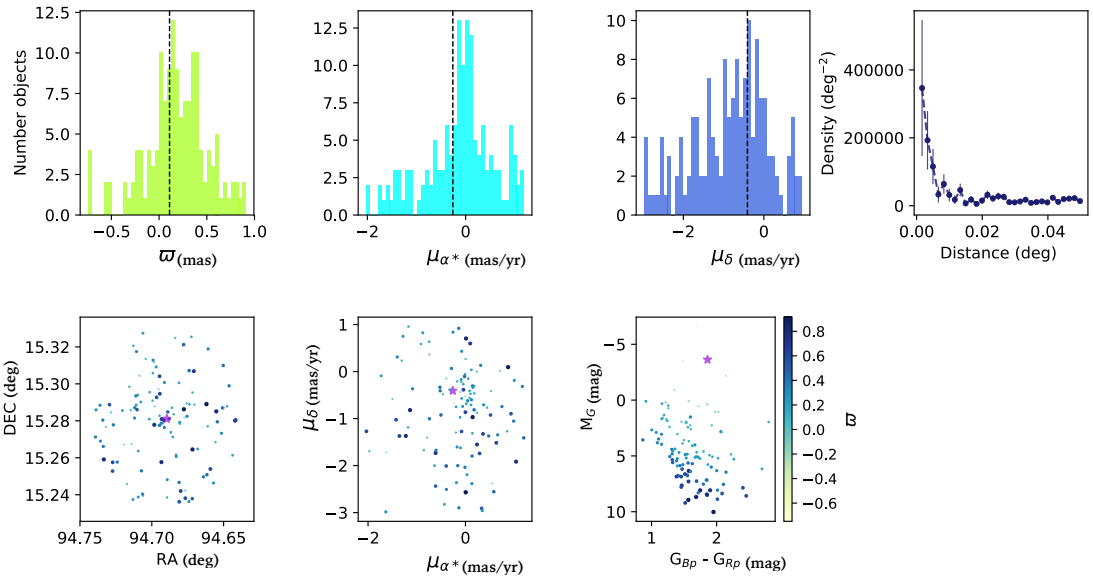


Figure 3.5: Final iteration image from CEREAL to establish if the star was a yes, maybe or not. This star has been classified as a *yes*. This image shows the data for the HAeBe star MWC137. The distribution of the parallax and the proper motions appear evenly distributed around the known value. The density profile shows a smooth decrease in the density as function of the distance. The spatial and proper motion distributions show that the low mass companions around the HAeBe star share similar parallax values. With the same labels as figure 3.1.

3.2.2 *Maybe*: an example where CEREAL assigned a candidate as being potentially in a cluster

Figure 3.6 is an example of a star classified as *maybe* by CEREAL. The differences in the classification between the *yes* and the *maybe* are very subtle. The classification is based principally on the histogram distributions and the density profile. In this case, at least two of the three histograms appear evenly distributed around the known value. Also, for a *maybe* the histograms might have the peak shifted from the known value in one of the parameters, or the peak is not clearly defined. Additionally, their density profile might not show a smooth decrease in the density as function of the distance and might not become completely flat at longer distances. In this case it is possible to see, in the spatial and proper motion distributions, around the HAeBe star a group of stars that shared similar parallax values.

The differences between the *yes* and *maybe* classification might also be due to the number of stars that had been selected, or the HAeBe star might have been surrounded by a cloud or it may have a brighter star in its vicinity that prevents the observation of low mass companions around it.

3.2.3 *NO*: an example where CEREAL assigned a candidate as not being in a cluster

Figure 3.7 is an example of a star classified as *no* by CEREAL. In contrast to the subtle differences between *yes* and *maybe*, the difference between the stars classified as *yes* and the ones classified as *no* were more evident.

In their histograms, it is not clear where the peak in the parameter is located, due to the large number of sources that does not share the same properties as the target, or because the HAeBe stars are too bright that they masking the low mass companions around them.

Additionally, their density profiles could show similar densities around the centre of the stars that then appeared constant over the whole distance. The spatial and proper motion distributions might not show a clear group of stars that shared the same parallax as the HAeBe star.

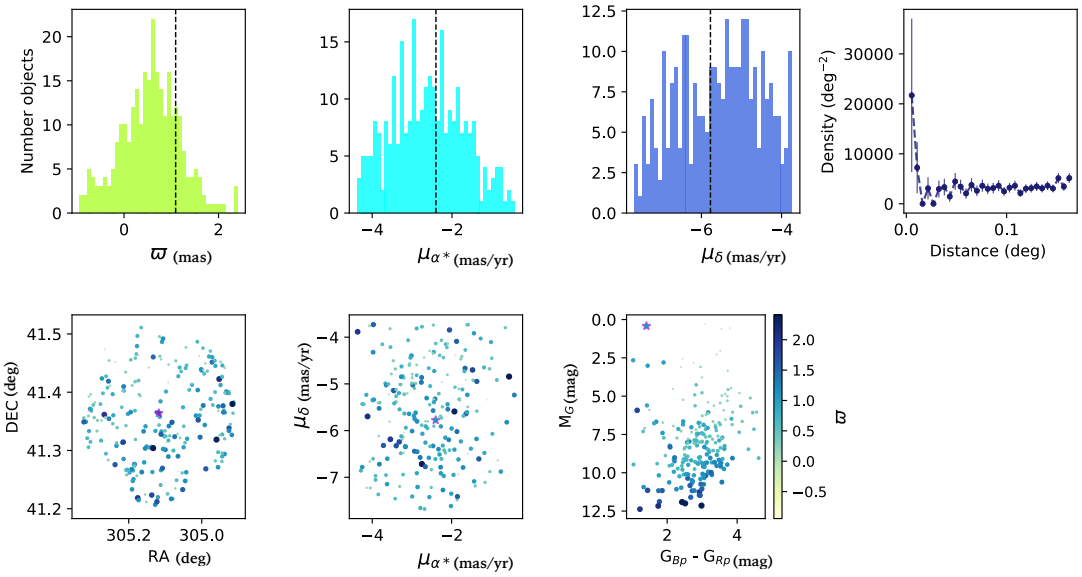


Figure 3.6: Final iteration image from CEREAL to establish if the star was a yes, maybe or no. This image represents a *maybe*. For this HAeBe star, V1685Cyg, the peak in the parallax distribution was shifted from the known value, there was no clear peak in either of the proper motion distributions (the peak in RA was very broad) and the density profile did not move smoothly down to a clear and constant background. But, as in figure 3.5, the spatial and proper motion distribution showed that there were low mass companions around the HAeBe star that shared similar parallax values. Here the same labels are used as in Figure 3.1.

3.2.4 Catalogue of Clusters around HAeBe stars

The classification of the HAeBE stars (as *yes*, *maybe* or *no*) was not trivial because it was not always clear how to decide on the presence of a cluster around a HAeBe star. The classification was based on a combination of the visual inspection of the histogram distributions, the density profiles, and the spatial and proper motion distributions; this made the decision very subjective. To account for the subjectivity of this process, the selections were performed multiple times by different people to confirm the results of the classifications made. Each person focused their inspection on a different part of the data to define the presence of a cluster; one put more emphasis in search for features in the density profiles (as Testi *et al.*, 1997) while the other focused more on the histogram distributions of the parallax and proper motions.

Comparing both the independent classifications, there was an agreement in the classi-

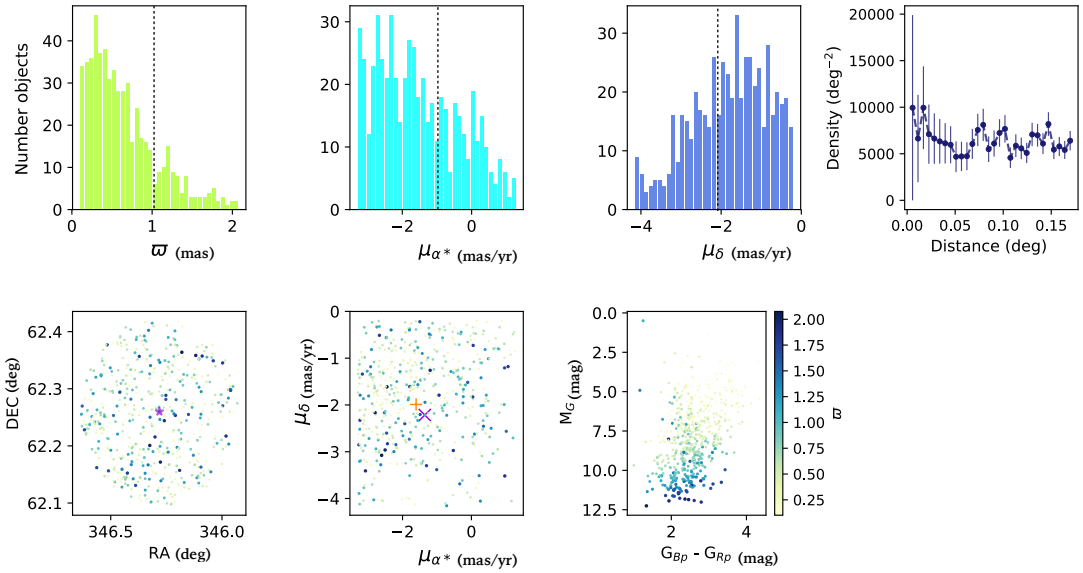


Figure 3.7: The final iterations images from CEREAL to establish if the star was a yes, maybe or no. This image represents a *no*. This image shows the HAeBe star V374Cep. The histogram distributions of all the parameters did not show any clear peak around the known values, this is either due to the large number of sources that do not share the same properties as the target, or because the HAeBe star is too bright that it shadows the low mass companions around it. The density profile shows a similar density which is almost constant across the whole distance range from the central point to long distances. Additionally, as opposed to figure 3.5, the spatial and proper motion distributions do not show any clear group of stars that share the same parallax as the HAeBe star. This figure utilizes the same labels as figure 3.1.

fication for 56% of the stars that were found to be not in clusters (classified as *No*), 10% of the stars found to be in clusters (classified as *Yes*) and 8% of the stars found to potentially be in a cluster (classified as *Maybe*). Whilst, for the remainder of the stars ($\sim 26\%$) there were differences in their classifications. For these objects it was unclear whether they were *Yes*, *Maybe* or *No*; therefore, a second examination of these targets was performed. The final classification of the sample combined the results of both selections to define the presence of clusters around the stars.

Using the classification system (as *yes*, *maybe* or *no*), created to characterise the results found by CEREAL, for the sample of 269 HAeBe stars, $41(15\%)$ stars were found to be in clusters (classified as *Yes*), $35(13\%)$ stars were found to be potentially in a cluster (classified as *Maybe*) and $193(72\%)$ stars were found to not be in clusters (classified as

No).

Table 3.1 shows the 76 stars found to be in or potentially in clusters by CEREAL, i.e., those objects classified as *yes* and *maybe*. This table also shows that 34 stars in this list are associated with known star forming regions in SIMBAD. The remaining stars, those that were not found in clusters, are listed in the appendix E; of these stars, 31 were also reported to be associated with known star forming regions in SIMBAD.

Table 3.1: Catalogue of clusters around the Herbig Ae/Be stars

Star	RA h:m:s	DEC deg:m:s	Spectral type	CEREAL	Know Cluster
A0974-15	05:06:55.50	-03:21:13.0	B3	YES	NGC1788[1]
AS310	18:33:21.20	-04:58:06.0	B1	YES	
AS470	21:36:14.20	+57:21:31.0	A0	MAYBE	
AS477	21:52:34.10	+47:13:44.0	A0	YES	IC5146[2]
BD+30549	03:29:19.80	+31:24:57.0	B8	YES	NGC1333[3]
BD+413731	20:24:15.71	+42:18:01.3	B3	YES	LDN 1641 in Orion A[4]
BFOri	05:37:13.30	-06:35:01.0	A7	YES	
COOri	05:27:38.30	+11:25:39.0	G5	YES	
GSC3975-0579	21:38:08.50	+57:26:48.0	A2	YES	IC1396[2]
GSC5360-1033	05:57:49.50	-14:05:34.0	B5	YES	
HBC217	06:40:42.20	+09:33:37.0	G0	YES	NGC2264[2]
HBC222	06:40:51.20	+09:44:46.0	F8	YES	NGC2264[2]
HBC442	05:34:14.20	-05:36:54.0	F8	MAYBE	Orion Nebula[4]
HBC694	20:24:29.54	+42:14:02.0	A5	YES	GN 20.22.7[5]
HD200775	21:01:36.90	+68:09:48.0	B2	MAYBE	Cep Flare[2]
HD245185	05:35:09.60	+10:01:51.0	A0	YES	Collinder 69 in Lamda Ori Complex[4]
HD287823	05:24:08.00	+02:27:47.0	A0	MAYBE	
HD288012	05:33:04.80	+02:28:10.0	A0	YES	
HD290380	05:23:31.00	-01:04:24.0	F0	YES	
HD290770	05:37:02.40	-01:37:21.0	B9	MAYBE	
HD323771	17:34:04.60	-39:23:41.0	B5	MAYBE	
HD36408	05:32:14.10	+17:03:29.0	B7	MAYBE	
HD36917	05:34:47.00	-05:34:15.0	A0	YES	Orion Nebula[4]
HD36982	05:35:09.84	-05:27:53.2	B1.5	YES	Orion Nebula[4]
HD37371	05:38:09.90	-00:11:02.0	B9	YES	
HD37806	05:41:02.30	-02:43:01.0	B9	MAYBE	
HD46060	06:30:49.80	-09:39:15.0	B8	YES	
HD68695	08:11:44.60	-44:05:09.0	A0	YES	
HD87403	10:02:51.40	-59:16:55.0	A1	MAYBE	Cl Platais 8[6]

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Kos *et al.* (2019), (2) Zucker *et al.* (2020), (3) Belloche *et al.* (2020), (4) Briceño *et al.* (2019), (5) Magakian (2003a), (6) Cantat-Gaudin *et al.* (2018a), (7) Getman *et al.* (2018), (8) Cantat-Gaudin *et al.* (2018b), (9) Mehner *et al.* (2016).

Table 3.2: Table 3.1 continued

Star	RA	DEC	Spectral type	CEREAL	Know Cluster
	h:m:s	deg:m:s			
HD96042	11:03:40.50	-59:25:59.0	B1	YES	
HD97048	11:08:03.20	-77:39:17.0	A0	YES	Cha I in Chan Dark Cloud[2]
HR5999	16:08:34.30	-39:06:19.0	A7	MAYBE	Sco OB[2]
HTCMa	07:02:42.50	-11:26:12.0	A0	YES	
HUCMa	07:04:06.70	-11:26:08.0	B8	YES	
Hen2-80	12:22:23.20	-63:17:17.0	B5	MAYBE	
Hen3-1121	15:58:09.62	-53:51:18.4	B0.5	MAYBE	
Hen3-949	13:57:44.00	-39:58:47.0	B3	MAYBE	
ILCep	22:53:15.60	+62:08:45.0	B2	YES	
LKHa260	18:19:09.40	-13:50:41.0	B6	MAYBE	M 16 in Eagle Nebula[2]
LKHa338	06:10:47.13	-06:12:50.6	B9	YES	GGD 12-15[7]
LkHa215	06:32:41.80	+10:09:34.0	B6	MAYBE	
LkHa257	21:54:18.70	+47:12:09.0	B5	MAYBE	IC5146[2]
LkHa324	21:03:54.20	+50:15:10.0	B9	YES	L988 e[8]
LkHa339	06:10:57.80	-06:14:40.0	A1	YES	GGD 12-15[7]
MWC1021	20:29:26.90	+41:40:44.0	B0	MAYBE	
MWC1080	23:17:25.60	+60:50:43.0	B0	MAYBE	
MWC137	06:18:45.50	+15:16:52.0	B0	YES	SH 2-266[9]
MWC297	18:27:39.50	-03:49:52.0	B1	MAYBE	
NVOri	05:35:31.40	-05:33:09.0	B0	MAYBE	M 42 in Orion Nebula[4]
PDS126	06:13:37.30	-06:25:02.0	A7	YES	
PDS129	06:31:03.60	+10:01:13.0	F5	YES	
PDS134	07:32:26.60	-21:55:36.0	B6	MAYBE	
PDS277	08:23:11.80	-39:07:01.0	F3	YES	
PDS324	10:57:24.30	-62:53:13.0	B1	MAYBE	
PDS415N	16:18:37.22	-24:05:18.5	F0	MAYBE	Sco OB II[2]
PDS520	18:30:06.20	+00:42:34.0	F3	MAYBE	Serpens Cloud[2]
PDS551	18:55:22.98	+04:04:35.1	O9	MAYBE	
RRTau	05:39:30.50	+26:22:27.0	A0	MAYBE	
TOri	05:35:50.50	-05:28:35.0	A3	MAYBE	M 42 in Orion Nebula[4]
UYOri	05:32:00.30	-04:55:54.0	B9	MAYBE	
V1478Cyg	20:32:45.50	+40:39:37.0	B0	MAYBE	
V1685Cyg	20:20:28.20	+41:21:51.0	B3	MAYBE	
V1787Ori	05:38:09.30	-06:49:17.0	A5	YES	LDN 1641 in Orion A[4]
V1977Cyg	20:47:37.50	+43:47:25.0	B8	MAYBE	
V2019Cyg	20:48:04.80	+43:47:26.0	B2	MAYBE	
V346Ori	05:24:42.80	+01:43:48.0	A8	YES	

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Kos *et al.* (2019), (2) Zucker *et al.* (2020), (3) Belloche *et al.* (2020), (4) Briceño *et al.* (2019), (5) Magakian (2003a), (6) Cantat-Gaudin *et al.* (2018a), (7) Getman *et al.* (2018), (8) Cantat-Gaudin *et al.* (2018b), (9) Mehner *et al.* (2016).

Table 3.3: Table 3.1 continued

Star	RA h:m:s	DEC deg:m:s	Spectral type	CEREAL	Know Cluster
V350Ori	05:40:11.80	-09:42:11.0	A1	MAYBE	
V361Cep	21:42:50.20	+66:06:35.0	B2	YES	NGC7129[8]
V373Cep	21:43:06.80	+66:06:54.0	B5	YES	NGC7129[8]
V380Ori	05:36:25.40	-06:42:58.0	A1	YES	LDN 1641 in Orion A[4]
V388Vel	08:42:17.30	-40:44:10.0	A1	MAYBE	[DBS2003]21[8]
V586Ori	05:36:59.25	-06:09:16.3	A2	YES	LDN 1641 in Orion A[4]
V590Mon	06:40:44.60	+09:48:02.0	B7	YES	NGC2264[2]
V594Cas	00:43:18.30	+61:54:40.0	B8	YES	
V599Ori	05:38:58.60	-07:16:46.0	F0	MAYBE	LDN 1641 in Orion A[4]
WWVul	19:25:58.80	+21:12:31.0	A3	YES	

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Kos *et al.* (2019), (2) Zucker *et al.* (2020), (3) Belloche *et al.* (2020), (4) Briceño *et al.* (2019), (5) Magakian (2003a), (6) Cantat-Gaudin *et al.* (2018a), (7) Getman *et al.* (2018), (8) Cantat-Gaudin *et al.* (2018b), (9) Mehner *et al.* (2016).

3.3 Comparison between CEREAL and clustering algorithms from the literature

In recent years, there has been an increase in the volume of astronomical data available. An example of this is the Gaia DR2 catalogue, which provides accurate data for more than a billion stars. In parallel, a wide variety of different methodologies have begun to be implemented for the analysis of these larger samples, such as Machine Learning techniques (Alpaydin, 2014; Géron, 2019).

The *Density-based clustering* algorithms are a sub-type of machine learning algorithms dedicated to the study of overdensities. In these algorithms, the overdensities represent the possible clusters in the sample. A cluster in these algorithms is defined as an area of higher density of an arbitrary shape with respect to the remainder of the data (Pedregosa *et al.*, 2011).

Some representative density-based clustering algorithms (DCAs) are: *DBSCAN* (Ester *et al.*, 1996), *OPTICS* (Ankerst *et al.*, 1999) and *HDBSCAN* (Campello *et al.*, 2015; McInnes *et al.*, 2017). These DCAs were chosen from the literature to compare their efficiency in finding clusters against CEREAL. More information about the density-based

clustering algorithms is provided in appendix F.

255 out of the sample of 269 stars were analysed by each of the DCAs following the procedure from Cánovas *et al.* (2019), which used the position and, at least, two astrometric parameters (ϖ, $\mu_{\alpha*}$ and μ_δ). The remaining 14 stars of the sample could not be analysed because they did not have all the data for their astrometric parameters (ϖ, $\mu_{\alpha*}$ and μ_δ) needed to process the sample with the DCAs.

Each method applies a technique to extract only sources that have high quality astrometry. CEREAL applies a value of RUWE$\leq$ 1.40 (see section §2.1) to obtain a clean sample. The DCAs also use the value of RUWE$\leq$ 1.40, and additionally, removed any objects with parallax signal to noise $>$ 10 and visibility periods $\leq$ 7 (Arenou *et al.*, 2018; Cánovas *et al.*, 2019). For this analysis the quality conditions apply for the DCAs where removing any object with RUWE$\leq$ 1.40 and parallax signal to noise $>$ 10.

Furthermore, to quantify how many stars were found in a cluster through the DCAs, a homogeneous classification system was applied to the results obtained. This system was the same as the one used for CEREAL (see section §2.1), where the stars that were considered to be in a cluster were represented by the flag *yes*; the stars that were potentially considered to be in a cluster were represented by the flag *maybe*; and, the stars that were not considered to be in a cluster were represented by the flag *no*. These flags were added to the stars analyzed by the DCAs to simplify the comparison with CEREAL.

The results from the three DCAs were combined and compared with the classifications made by CEREAL for the same sample. The results for the 255 HAeBe stars, with each independent technique, showed that both CEREAL and the DCAs retrieved a similar number of stars in clusters, possibly in clusters and found to not be in clusters. However, when these methods both detected a cluster around the same HAeBe star they did not necessarily find the same number of low mass companions in the cluster. Excluding the $14(5\%)$ stars unclassified due to the absence of their parallax and proper motions, CEREAL found: $41(15\%)$ as *yes*, $35(13\%)$ as *maybe* and $179(67\%)$ as *no*; and, the three DCAs combined found: $39(14\%)$ as *yes*, $47(17\%)$ as *maybe*, $169(63\%)$ as *no*.

3.3.1 Cases where CEREAL and DCAs found a cluster around a HAeBE star

Although CEREAL and DCAs applied a different technique to select the low mass companions around the HAeBe stars, it was possible for both methods to find clusters around the same stars, which happened on 28 occasions.

Figure 3.8 gives a representative example where both methods found a cluster around the same star. The figure shows a closer look at the distributions of the parallax and proper motions for the HAeBe star ILCep (spectral type B2).

However, the number of possible candidates for cluster members found by each method was different; only a subset of these were found in common by both methods. CEREAL found the possible candidate members for each clusters by defining tighter ranges around the known values of each astrometric parameter (ϖ, $\mu_{\alpha*}$ and μ_δ), where the peaks around the known values are still visible and well defined (as are show in figure 2.2). In the previous example, for the HAeBe star ILCep CEREAL found 520 stars and DCAs found 24, with 20 stars in common in both samples.

This difference between both methods may be due to CEREAL concentrating the selection of the possible members of the cluster within a tight range around the known values of the parallax and proper motion in RA and DEC for the HAeBe star. Those selected stars, which represented the final selection made by CEREAL, have similar values to the known values of the parallax and proper motion in RA and DEC for the HAeBe star.

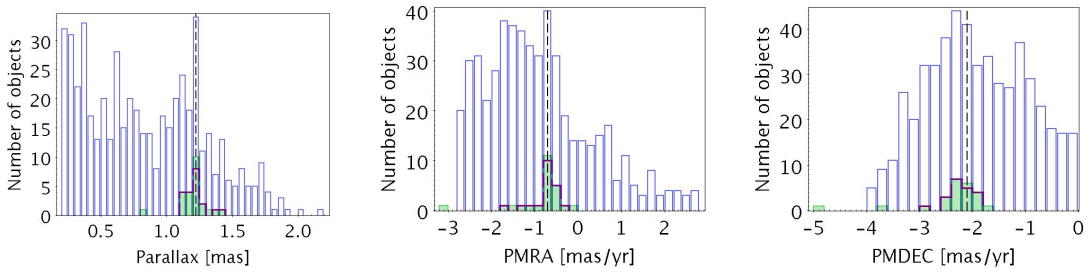


Figure 3.8: Comparison between a HAeBe star found in clusters by CEREAL and DCAs. From left to right, the panels in each figure represent the parallax, the proper motion in RA and the proper motion in DEC for the HAeBe star ILCep. In the figure, the open blue histogram represents the final selection of stars made by CEREAL, the green bars represent the stars selected by the DCAs and the open purple histogram represents the stars in common between the selection made by CEREAL and the DCAs. The vertical dashed line in each histogram highlights the known value of each parameter for the HAeBe star ILCep.

The DCAs instead tend to define the presence of a cluster around a HAeBe star only using two parameters, the parallax or the proper motion in RA or DEC. With this method, the DCAs often found stars with a larger range of values, these objects might have parallax and proper motions similar to the HAeBe stars, but might not actually be part of the cluster.

3.3.2 Case when CEREAL or DCAs found a cluster around a HAeBE star while the other algorithm does not

In addition to the differences in the objects found to be in the clusters found by both methods, there were examples where one method found a cluster whilst the other found the same object to not be in a cluster. The first type of these that will be discussed is when CEREAL was able to find a cluster around the HAeBe stars when the DCAs found the star to not be in a cluster. This happened on 4 occasions; around the HAeBe stars HD68695, MWC137, PDS277 and V586Ori.

An example of which is the HAeBe star MWC137 which CEREAL found to be in a cluster and the DCAs found to be not in a cluster. Figure 3.9 shows the distribution of the parallax and proper motions for the HAeBe star MWC137 (with spectral type B0).

To obtain this result, CEREAL started its analysis of MWC137 with a sample from Gaia DR2 of $\sim$ 40000 stars with 0.049 degrees. After all the iterations made over the parallax and proper motion in RA and DC, CEREAL found 149 candidate members for the cluster around MWC137.

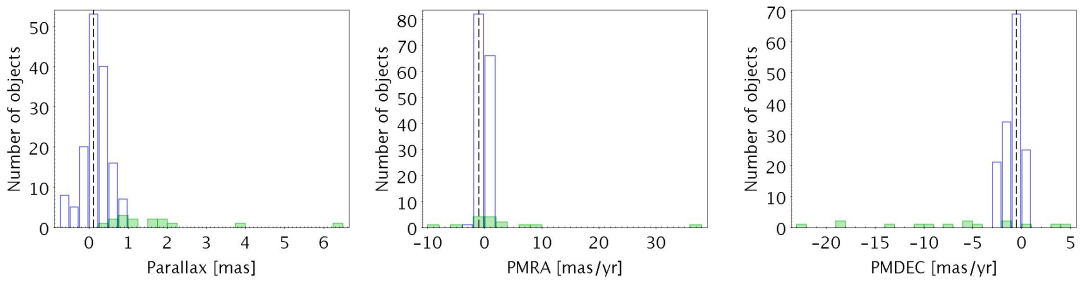


Figure 3.9: Comparison between the stars found in a cluster by CEREAL and not in a cluster by DCAs. CEREAL finds a cluster for the HAeBe star MWC137 and the DCAs do not. The open blue histogram represent the stars selected with CEREAL and, the green bars represent the stars selected with the DCAs. The vertical dashed line in each histogram points out the known value of each parameter for the HAeBe star.

This result is contrary to that found by the DCAs for the same target. The DCAs methods started however, with a sample of only 15 stars from which they were unable to find a cluster. The 15 stars were chosen by the rigorous pre-selection criteria from the $\sim$ 40000 stars selected from Gaia DR2, specifically, in the parallax (parallax signal to noise $>$ 10) which removed all the targets in the sample that were at the same distance as the HAeBe star. The 15 stars in their input sample did not contain any stars close to the known value of the parallax and proper motions of MWC137 (see figure 3.9). The differences in the results obtained by each algorithm could therefore be traced back to the selection criteria applied to reduce the data before applying the methods.

The opposite observation to that above was also made, where CEREAL did not find a cluster around the HAeBe stars but the DCAs did. This happened on 3 occasions; around the HAeBe stars GSC5379-0359, HD155448 and HD37357. Figure 3.10 shows the distribution for the parallax and proper motion of the HAeBe star HD37357 (spectral type A0).

During the selection process, CEREAL was unable to find stars that shared similar values for their parallax and proper motions to the star HD37357. This made it impossible for CEREAL to define a cluster for the stars in the sample (see figure 3.7). This can occur because the HAeBe stars were embedded deeply in a high density region, or in other cases, the HAeBe star might not be the most massive object in the possible cluster.

Contrary to this result, the DCAs appear to have selected the bright stars in the vicinity of the HAeBe star, these objects do not necessarily share all the same astrometric

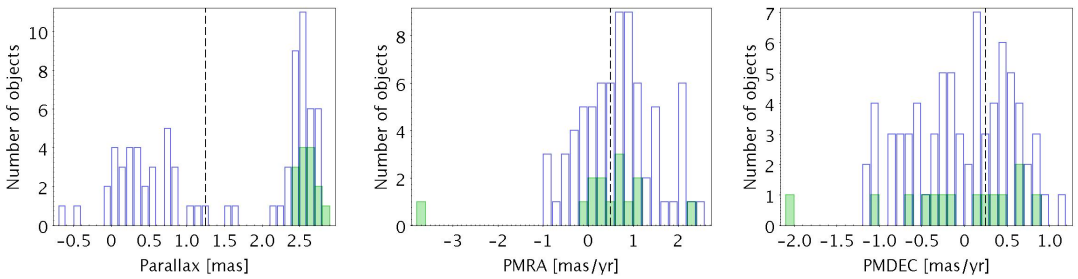


Figure 3.10: Comparison between the stars found in clusters by DCAs and not in clusters by CEREAL. The DCAs find a cluster for HD37357 while CEREAL does not. The open blue histogram represents the stars selected by CEREAL and the green bars represent the stars selected by the DCAs. The vertical dashed line in each histogram points out the known value of each parameter for the HAeBe star HD37357.

parameters as the HAeBe stars.

Figure 3.11 shows a colour image around the HAeBe star HD37357 (spectral type A1V) and highlights the stars selected by CEREAL and the DCAs. This figure shows that the HAeBe star HD37357 is surrounded by two other stars that appear to be more massive. The first one, and nearby, is HD37373 which have a spectral type B8V (Houk & Swift (1999); SIMBAD) and associated with the molecular cloud LDN1641 (Briceño *et al.* (2019); Großschedl *et al.* (2019); SIMBAD); and the second star is HD37481 with spectral type B2IV/V (Houk & Swift (1999), SIMBAD). This shows that the HAeBe star is not the most massive object in the possible cluster.

Even with the differences described above when comparing the clusters found by each method, it was possible to combine the results for the HAeBe stars found to be in a

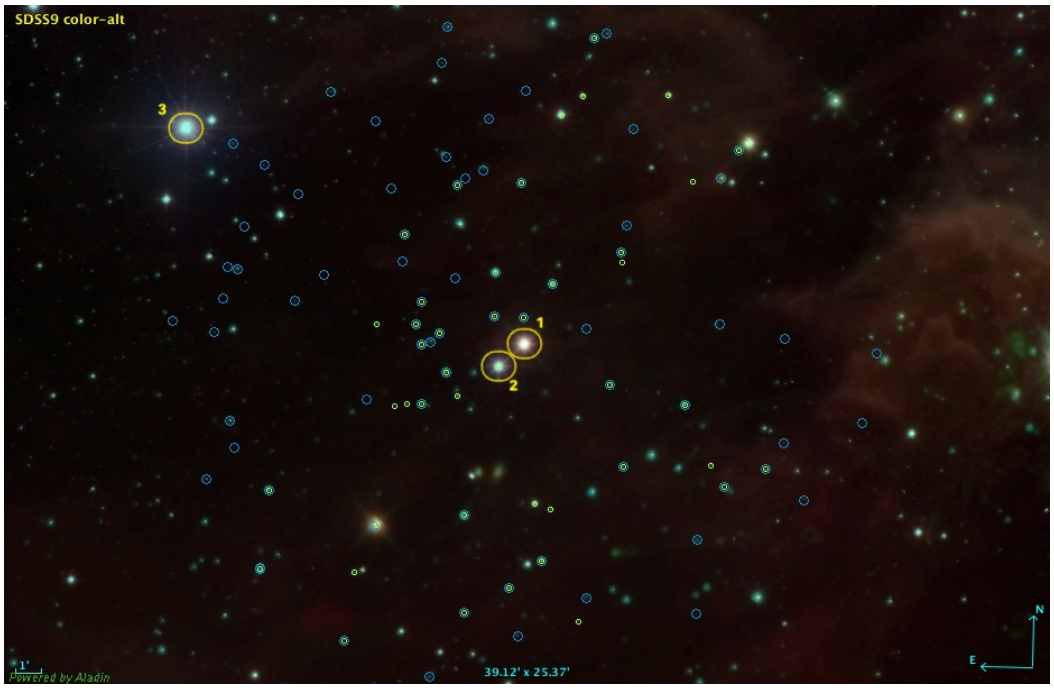


Figure 3.11: Visual inspection of the HAeBe star HD37357 found in a cluster by DCAs and not by CEREAL. The image is a colour composition created with the survey SDSS9, DSS2 and All WISE in Aladin (Bonnarel *et al.*, 2000). The open blue circles represents the stars selected by CEREAL and the open green circles represents the stars selected by the DCAs. The open yellow circles represent: 1) HAeBe star HD37357 with spectral type A1V; 2) Star HD37373 with spectral type B8V; and 3) Star HD37481 with spectral type B2IV/V. The HAeBe star might not be the most massive object in the possible cluster.

cluster, and create two separate samples. The first one, named *gold sample*, contain all those stars classified as *yes (Y)* by CEREAL and the DCAs (Y:Y). The second sample, named *silver sample*, contains the stars which were found to maybe be in a cluster by both methods or to be in a cluster by one method and maybe in a cluster by the other (Y:M, M:Y, M:M).

Table 3.4 contains the 28 stars of the gold sample and the 23 stars of the silver sample. The appendix G shows a table with the number of stars found by each DCA and CEREAL for each star in the gold and silver sample.

Table 3.4: Stars found in clusters by CEREAL and the Density-based clustering algorithms

Star	RA	DEC	Spectral type	CEREAL Class	DCAs Class	Clusters Flag
	h:m:s	deg:m:s				
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A0974-15	05:06:55.50	-03:21:13.0	B3	YES	YES	G
AS310	18:33:21.20	-04:58:06.0	B1	YES	MAYBE	S
AS477	21:52:34.10	+47:13:44.0	A0	YES	YES	G
BD+30549	03:29:19.80	+31:24:57.0	B8	YES	YES	G
BD+413731	20:24:15.71	+42:18:01.3	B3	YES	YES	G
BFOri	05:37:13.30	-06:35:01.0	A7	YES	YES	G
COOri	05:27:38.30	+11:25:39.0	G5	YES	MAYBE	S
GSC3975-0579	21:38:08.50	+57:26:48.0	A2	YES	YES	G
GSC5360-1033	05:57:49.50	-14:05:34.0	B5	YES	YES	G
HBC217	06:40:42.20	+09:33:37.0	G0	YES	YES	G
HBC222	06:40:51.20	+09:44:46.0	F8	YES	YES	G
HBC442	05:34:14.20	-05:36:54.0	F8	MAYBE	YES	S
HBC694	20:24:29.54	+42:14:02.0	A5	YES	YES	G
HD200775	21:01:36.90	+68:09:48.0	B2	MAYBE	YES	S
HD245185	05:35:09.60	+10:01:51.0	A0	YES	MAYBE	S
HD287823	05:24:08.00	+02:27:47.0	A0	MAYBE	MAYBE	S
HD288012	05:33:04.80	+02:28:10.0	A0	YES	MAYBE	S
HD290380	05:23:31.00	-01:04:24.0	F0	YES	MAYBE	S
HD290770	05:37:02.40	-01:37:21.0	B9	MAYBE	MAYBE	S
HD36917	05:34:47.00	-05:34:15.0	A0	YES	YES	G
HD36982	05:35:09.84	-05:27:53.2	B1.5	YES	MAYBE	S
HD37371	05:38:09.90	-00:11:02.0	B9	YES	YES	G
HD46060	06:30:49.80	-09:39:15.0	B8	YES	YES	G
HD87403	10:02:51.40	-59:16:55.0	A1	MAYBE	YES	S
HD96042	11:03:40.50	-59:25:59.0	B1	YES	MAYBE	S
HD97048	11:08:03.20	-77:39:17.0	A0	YES	YES	G
HTCMa	07:02:42.50	-11:26:12.0	A0	YES	YES	G
HUCMa	07:04:06.70	-11:26:08.0	B8	YES	YES	G
Hen3-1121	15:58:09.62	-53:51:18.4	B0.5	MAYBE	YES	S

Column 7 represent a flag for the gold (**G**) and silver (**S**) sample describe in section 3.3

Table 3.5: Table 3.4 continued

Star	RA	DEC	Spectral type	CEREAL Class	DCAs Class	Clusters Flag
	h:m:s	deg:m:s				
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hen3-949	13:57:44.00	-39:58:47.0	B3	MAYBE	YES	S
ILCep	22:53:15.60	+62:08:45.0	B2	YES	YES	G
LKHa260	18:19:09.40	-13:50:41.0	B6	MAYBE	MAYBE	S
LKHa338	06:10:47.13	-06:12:50.6	B9	YES	MAYBE	S
LkHa257	21:54:18.70	+47:12:09.0	B5	MAYBE	MAYBE	S
LkHa324	21:03:54.20	+50:15:10.0	B9	YES	YES	G
LkHa339	06:10:57.80	-06:14:40.0	A1	YES	YES	G
PDS126	06:13:37.30	-06:25:02.0	A7	YES	YES	G
PDS129	06:31:03.60	+10:01:13.0	F5	YES	YES	G
PDS324	10:57:24.30	-62:53:13.0	B1	MAYBE	YES	S
PDS415N	16:18:37.22	-24:05:18.5	F0	MAYBE	YES	S
PDS520	18:30:06.20	+00:42:34.0	F3	MAYBE	YES	S
TOri	05:35:50.50	-05:28:35.0	A3	MAYBE	MAYBE	S
V1685Cyg	20:20:28.20	+41:21:51.0	B3	MAYBE	MAYBE	S
V1787Ori	05:38:09.30	-06:49:17.0	A5	YES	YES	G
V346Ori	05:24:42.80	+01:43:48.0	A8	YES	MAYBE	S
V361Cep	21:42:50.20	+66:06:35.0	B2	YES	YES	G
V373Cep	21:43:06.80	+66:06:54.0	B5	YES	YES	G
V380Ori	05:36:25.40	-06:42:58.0	A1	YES	YES	G
V590Mon	06:40:44.60	+09:48:02.0	B7	YES	YES	G
V594Cas	00:43:18.30	+61:54:40.0	B8	YES	YES	G
WWVul	19:25:58.80	+21:12:31.0	A3	YES	YES	G

Column 7 represent a flag for the gold (**G**) and silver (**S**) sample describe in section 3.3

3.3.3 Case when CEREAL uses the quality conditions of the DCAs to find clusters around the HAeBe stars

To complement the results obtained in sections § 3.3.1 and § 3.3.2, the effect of applying the same quality conditions used by the DCAs on the Gaia DR2 data is taken into consideration in this section for the analysis of this data with CEREAL. This study was carried out in order to see whether the same stars in the clusters around the HAeBe stars were recovered, when using the DCAs conditions with CEREAL.

To begin the analysis, the quality conditions of the DCAs (Cánovas *et al.*, 2019) were applied to the raw Gaia DR2 data. These conditions kept only objects with RUWE$\leq$1.40 and parallax signal to noise $>$10. A table which contains the number of stars obtained

when the different quality conditions were applied is given as appendix G.

The stars obtained after applying the quality conditions of the DCAs (Cánovas *et al.*, 2019) were then re-analysed using the same selection ranges for parallax and proper motions used by CEREAL in the first analysis (see section § 3.2). This was carried out to allow for the comparison of CEREAL and the DCAs for the same sample, because it was not practicable to reanalyse the original CEREAL sample again with of each of the DCAs considered in Cánovas *et al.* (2019).

As an input for this analysis the gold and silver samples shown in table 3.4 were used. These samples contained 51 HAeBe stars that were classified as yes and maybe by CEREAL and the DCAs in section § 3.3; this sample contains 28 stars from the gold sample and 23 stars from the silver sample.

Tables 3.6 show a comparison of the results obtained by CEREAL and each DCAs (*DBSCAN*,*OPTICS* and *HDBSCAN*; Ankerst *et al.*, 1999; Campello *et al.*, 2015; Ester *et al.*, 1996; McInnes *et al.*, 2017, respectively) for the gold and silver samples. This table shows that CEREAL found a cluster for each of the stars in the sample while the DCAs, *DBSCAN*, *OPTICS* and *HDBSCAN*, found clusters around 42, 46 and 51 cluster stars, respectively. A summary which contains the number of stars obtained by each DCA is given as appendix G.

The different numbers of clusters found by the DCAs might be due to the selection process used by each algorithm. It is of particular note that *HDBSCAN* (Campello *et al.*, 2015; McInnes *et al.*, 2017) is the only algorithm that found the same number of clusters as CEREAL. *HDBSCAN* is the only algorithm that can identify clusters with different densities using only one parameter and is sensitive to the density gradients inside a cluster (see appendix F) which is similar to the process used by CEREAL, searching for over-densities in the samples.

The result obtained from this analysis was that 29 HAeBe stars were classified as *yes* and 22 as *maybe* from the sample of 51 HAeBe stars, with no stars found to be *no*, not in a cluster. This result shows that it was still possible to find stars in clusters using the original selection ranges with CEREAL and the quality conditions of the DCAs. This result is shown in table 3.7, where the columns represent the cluster classification found by CEREAL and the DCAs.

Figure 3.12 shows an example of this analysis for the HAeBe star BD+30549 classified as yes by both methods. The plots show the histogram distributions of the parallax and

Table 3.6: Clusters found by CEREAL and each DCA

Star	CEREAL	DBSCAN	OPTICS	HDBSCA	Star	CEREAL	DBSCAN	OPTICS	HDBSCA
A097415	Y	Y	Y	Y	HTCMa	Y	Y	Y	Y
AS310	Y	Y	Y	Y	HUCMa	Y	Y	Y	Y
AS477	Y	Y	Y	Y	Hen31121	Y	Y	Y	Y
BD+30549	Y	Y	Y	Y	Hen3949	Y	Y	Y	Y
BD+413731	Y	Y	Y	Y	ILCep	Y	Y	Y	Y
BFOri	Y	Y	Y	Y	LKHa260	Y	Y	Y	Y
COOri	Y	N	Y	Y	LKHa338	Y	Y	Y	Y
GSC3975-0579	Y	Y	Y	Y	LkHa257	Y	N	N	Y
GSC5360-1033	Y	Y	Y	Y	LkHa324	Y	Y	Y	Y
HBC217	Y	N	Y	Y	LkHa339	Y	Y	Y	Y
HBC222	Y	N	Y	Y	PDS126	Y	Y	Y	Y
HBC442	Y	N	N	Y	PDS129	Y	Y	Y	Y
HBC694	Y	Y	Y	Y	PDS324	Y	Y	Y	Y
HD200775	Y	Y	Y	Y	PDS415N	Y	N	N	Y
HD245185	Y	Y	Y	Y	PDS520	Y	Y	Y	Y
HD287823	Y	Y	Y	Y	TOri	Y	Y	Y	Y
HD288012	Y	N	N	Y	V1685Cyg	Y	Y	Y	Y
HD290380	Y	N	N	Y	V1787Ori	Y	Y	Y	Y
HD290770	Y	Y	Y	Y	V346Ori	Y	Y	Y	Y
HD36917	Y	Y	Y	Y	V361Cep	Y	Y	Y	Y
HD36982	Y	N	Y	Y	V373Cep	Y	Y	Y	Y
HD37371	Y	Y	Y	Y	V380Ori	Y	Y	Y	Y
HD46060	Y	Y	Y	Y	V590Mon	Y	Y	Y	Y
HD87403	Y	Y	Y	Y	V594Cas	Y	Y	Y	Y
HD96042	Y	Y	Y	Y	WWVul	Y	Y	Y	Y
HD97048	Y	Y	Y	Y					

proper motions for the original CEREAL and DCAs selection; the CEREAL selection using the DCAs quality conditions and the stars in common in all the catalogues. The figure also shows the position of the known values of each parameter, where the peak of the distribution is concentrated. The figure shows an agreement in the results between the previous selection and this analysis, where it is possible to see a peak in the distribution of each sample around the same area. The number of stars found by each method were 99 stars for the original CEREAL selection; 53 stars for the DCAs selection; 51 stars for the CEREAL selection using the DCAs quality conditions; of these stars 32 are found in common in the 3 catalogues.

For both the original CEREAL results and those found using the DCAs quality conditions it is clear that there are some outliers present in the data; this is due to the large selection ranges used in the analysis. However, this does not change the overall results since most of the objects are concentrated close to the known values. Additionally,

Table 3.7: New CEREAL classification for the sample.

Star	Spectral type	CEREAL class	DCAs class	CEREAL DCAs class	Star	Spectral type	CEREAL class	DCAs class	CEREAL DCAs class
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A0974-15	B3	YES	YES	YES	HTCMa	A0	YES	YES	YES
AS310	B1	YES	MAYBE	YES	HUCMa	B8	YES	YES	YES
AS477	A0	YES	YES	YES	Hen3-1121	B0.5	MAYBE	YES	MAYBE
BD+30549	B8	YES	YES	YES	Hen3-949	B3	MAYBE	YES	MAYBE
BD+413731	B3	YES	YES	YES	ILCep	B2	YES	YES	YES
BFOri	A7	YES	YES	YES	LKHa260	B6	MAYBE	MAYBE	MAYBE
COOri	G5	YES	MAYBE	MAYBE	LKHa338	B9	YES	MAYBE	MAYBE
GSC3975-0579	A2	YES	YES	YES	LkHa257	B5	MAYBE	MAYBE	YES
GSC5360-1033	B5	YES	YES	MAYBE	LkHa324	B9	YES	YES	YES
HBC217	G0	YES	YES	YES	LkHa339	A1	YES	YES	YES
HBC222	F8	YES	YES	MAYBE	PDS126	A7	YES	YES	YES
HBC442	F8	MAYBE	YES	MAYBE	PDS129	F5	YES	YES	YES
HBC694	A5	YES	YES	YES	PDS324	B1	MAYBE	YES	MAYBE
HD200775	B2	MAYBE	YES	MAYBE	PDS415N	F0	MAYBE	YES	MAYBE
HD245185	A0	YES	MAYBE	YES	PDS520	F3	MAYBE	YES	MAYBE
HD287823	A0	MAYBE	MAYBE	MAYBE	TOri	A3	MAYBE	MAYBE	MAYBE
HD288012	A0	YES	MAYBE	MAYBE	V1685Cyg	B3	MAYBE	MAYBE	MAYBE
HD290380	F0	YES	MAYBE	YES	V1787Ori	A5	YES	YES	MAYBE
HD290770	B9	MAYBE	MAYBE	MAYBE	V346Ori	A8	YES	MAYBE	YES
HD36917	A0	YES	YES	MAYBE	V361Cep	B2	YES	YES	YES
HD36982	B1.5	YES	MAYBE	MAYBE	V373Cep	B5	YES	YES	YES
HD37371	B9	YES	YES	YES	V380Ori	A1	YES	YES	MAYBE
HD46060	B8	YES	YES	YES	V590Mon	B7	YES	YES	YES
HD87403	A1	MAYBE	YES	YES	V594Cas	B8	YES	YES	YES
HD96042	B1	YES	MAYBE	MAYBE	WWVul	A3	YES	YES	YES
HD97048	A0	YES	YES	YES					

CEREAL found a similar number of clusters to the DCAs for the objects in the sample in both cases; the main difference is in the number of stars recovered in these clusters.

Figure 3.13 shows another example of this analysis for the HAeBe star AS477, with spectal type A0. This HAeBe star was classified as yes by CEREAL (original selection and using the DCAs quality conditions) and DCAs.

As in the previous example, the plot shows the histogram distributions of the parallax and proper motions for the original CEREAL and DCAs selection; the CEREAL selection using the DCAs quality conditions and the correlation between the catalogues. Figure 3.13 also shows the position of the known value of each parameter, where the peak of the distribution is concentrated. The number of stars found by each method were 802 by the original CEREAL selection; 45 by the DCAs selection; 75 by the CEREAL selection using the DCAs quality conditions; and 36 stars in common in the 3 previous catalogues. In addition, the plot shows the spatial and proper motion distributions of the sample.

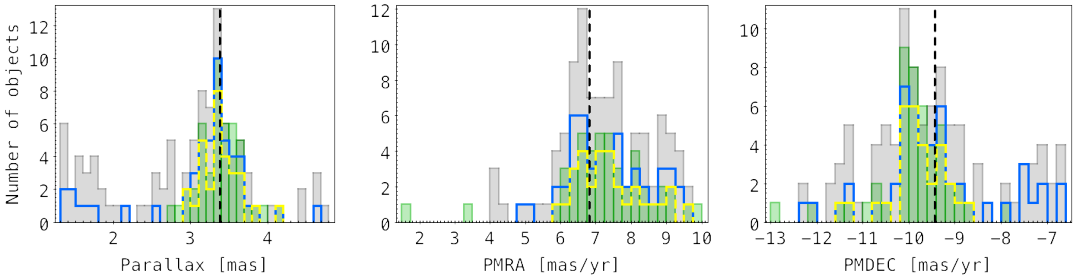


Figure 3.12: Comparison between a HAeBe star found in a cluster by CEREAL and DCAs using the quality conditions (RUWE$\leq$ 1.40 and parallax signal to noise $>$ 10). The HAeBe star BD+30549 was found to be in a cluster by both methods. From left to right the histogram distributions of the parallax and proper motions in RA and DEC are shown. The full grey steps represent the original sample selected by CEREAL; the green bars show the stars selected by the DCAs; the open blue steps represent the stars selected by CEREAL using the DCAs quality conditions; and the open dashed yellow steps show the stars that were found in common between the three catalogues. The vertical dashed line in each histogram points out the known value of each parameter for the HAeBe star.

The histogram distributions show a good agreement between all the catalogues by showing a peak around the known value of the HAeBe star, even when the distribution of the parallax for the CEREAL samples shows a large number of objects with parallax values $\leq$ 0.8 mas, which could be outliers, other clusters or background stars. This might also be due to the HAeBe star AS477 not being located at the centre of the cluster and due to the larger ranges used to select the data.

3.3.4 Final remarks section § 3.3

In general, CEREAL provides greater freedom than DCAs in parameter selection, when detecting clusters; although this then necessitates the investment of more time in their detection and classification. In addition, CEREAL uses the known values of each parameter as a starting point in the evaluation of the presence of a cluster around a HAeBe stars; whereas the DCAs search, over the whole sample, for patterns of over-densities without taking into consideration the known values of each parameter.

Both DCAs and CEREAL apply restrictions to the data used in their analysis that can also affect their final results. CEREAL only applies one restriction, on RUWE when extracting sources from the main sample, the DCAs apply another restrictions along with RUWE when extracting sources for the sample (parallax signal to noise $>$ 10; Cánovas

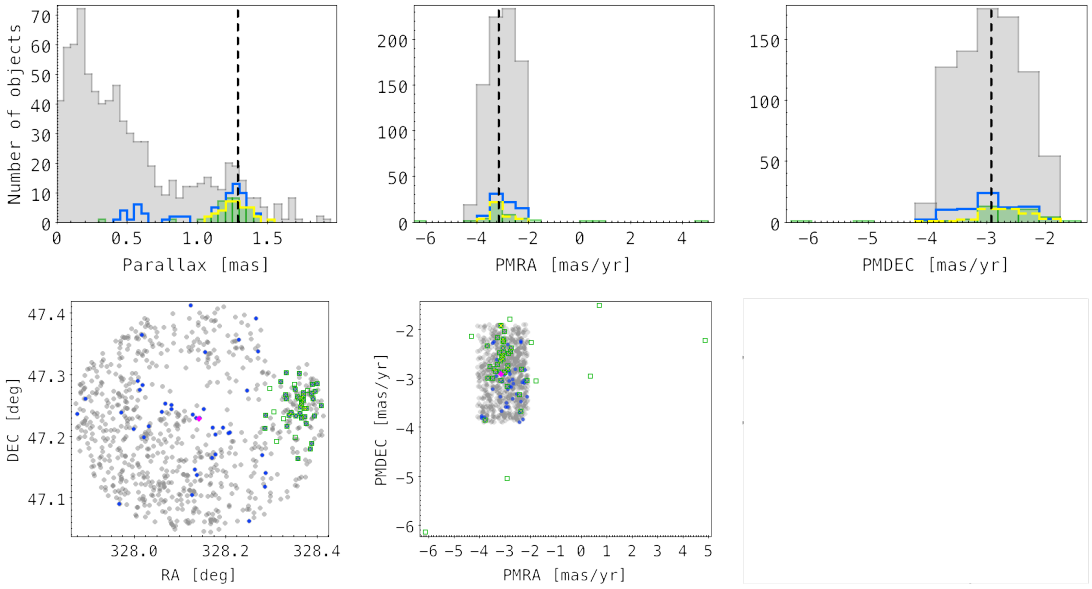


Figure 3.13: Comparison between a HAeBe star found in a cluster by CEREAL and DCAs using the quality conditions: RUWE$\leq$ 1.40 and parallax signal to noise $>$ 10, but not located at the centre of the cluster. The HAeBe star AS477 was found to be in a cluster by both methods. The top panels are the same as figure 3.12. The bottom panels show the spatial and proper motion distributions; the full grey circles represent the original sample selected by CEREAL; the open green square show the stars selected by the DCAs; the full blue circles represent the stars selected by CEREAL using the DCAs quality conditions; and the yellow x's show the stars that were found in common between the three catalogues. The HAeBe star AS477 is represent by the full magenta diamond.

et al., 2019). However, even when the samples are analysed with CEREAL using the same quality conditions as the DCAs, CEREAL still finds a larger number of low mass companions around the HAeBe stars. CEREAL also finds low mass companions around A-type stars which share a similar parallax and proper motion to the HAeBe star, in contrast to the DCAs which found stars that do not all appear to have the same proper motion as the HAeBe star. It is possible that these stars found by the DCAs might only represent the brightest stars around the HAeBe star (see figure 3.13).

The addition of the DCAs quality control restrictions can result in over-cleaning the initial samples, to the point that they can eliminate possible members of a cluster before even starting any analysis (see appendix G). This effect was seen for several HAeBe stars in (e.g MWC137 that was already known to belong to the cluster SH2-266) which the DCAs did not place in a cluster.

This comparison has shown the reliability of CEREAL as a method to find clusters, of any size, from any sample. Unlike the DCAs, that have been shown to be more efficient at searching for clusters in large samples of objects ($\geq$ 10000; Cánovas *et al.*, 2019) and obtain clusters of medium or large size. However, the clusters found by the DCAs may not always be associated with the HAeBe star.

3.4 Analysing the results obtained by Testi *et al.* (1999) using CEREAL

The work of Testi *et al.* (1997, 1998, 1999) represents a starting point for the search of stellar groups around Herbig Ae/Be stars. They studied the embedded population of young stars around 44 Herbig Ae/Be stars with spectral types ranging from O9 to A7, which ensured a good coverage of the intermediate mass range (Testi *et al.*, 1997, 1999). In their sample, the clusters found around the Herbig Ae/Be stars appeared only for those stars earlier than B5-B7 (Testi *et al.*, 1997). Additionally, Testi *et al.* (1999) reported the age for 25 HAeBe stars. Their ages were calculated from the location of the star in a H-R diagram, using the evolutionary tracks and isochrones from Palla & Stahler (1993).

Table 3.8 shows the 43 stars of the Testi *et al.* (1999) sample in common with CEREAL[1].

It should be noted that both studies used a different field of view to search for clusters around the HAeBe stars. Testi *et al.* (1997) used a radius of $\sim 0.06 - 0.11$ degrees which was big enough to be able to count some stars in the field close to the HAeBe star. CEREAL instead searched within a radius of ~ 0.5 degrees which covered a much larger area around the HAeBe star, and allowed the finding of more low mass companions in the vicinity of the HAeBe star which shared similar properties, such as parallax and proper motions.

Therefore, to make a direct comparison of the sample between the two methods, only the spectral types and the richness indicator (I_C) reported by Testi *et al.* (1999) were used. I_C is a very good indicator to measure how the density increases around the HAeBe star and corrects for the local background/foreground contamination.

[1]The missing star from Testi *et al.* (1999) sample is RNO 1B (V710 Cas), which has a spectral type of F8II (Samus' *et al.*, 2017) and is located in the dark cloud LDN 1287 (Juárez *et al.*, 2019). This star is also considered as a variable star of Orion Type (SIMBAD, Samus' *et al.* (2017))

Table 3.8: Common stars between Testi *et al.* (1999) and CEREAL

Testi *et al.* (1999) ID	CEREAL ID	Spectral type	I_C	CEREAL Class
ABAur	ABAur	A0	3.0 ± 6	NO
AS310	AS310	B0	70.0 ± 17	YES
BD+404124	V1685Cyg	B2	11.0 ± 3	MAYBE
BD+61154	V594Cas	B8	-1.4 ± 3	YES
BD+651637	V361Cep	B2	75.0 ± 5	YES
BD-061253	V380Ori	B9	-2.0 ± 2	YES
BFOri	BFOri	A7	1.1 ± 1	YES
BHJ71	V374Cep	B0	4.0 ± 3	NO
Elias1	V892Tau	A6	2.0 ± 3	NO
HBC202	VYMon	B8	23.2 ± 5	NO
HBC282	VVSer	B9	16.9 ± 5	NO
HBC312	LkHa257	B8	5.5 ± 6	MAYBE
HBC329	VXCas	A0	4.5 ± 4	NO
HD200775	HD200775	B3	1.9 ± 1	MAYBE
HD216629	ILCep	B2	34.0 ± 6	YES
HD245906	RRTau	A4	0.8 ± 6	MAYBE
HD250550	HD250550	B7	2.2 ± 2	NO
HD259431	HD259431	B5	0.9 ± 2	NO
HD293782	UXOri	A2	-0.3 ± 1	NO
HD37490	HD37490	B3	9.9 ± 3	NO
HD52721	GUCMa	B2	20.5 ± 4	NO
IPPer	IPPer	A3	5.3 ± 4	NO
LkHA218	HTCMa	B9	2.0 ± 5	YES
LkHA233	V375Lac	A7	1.0 ± 1	NO
LkHA25	V590Mon	B7	14.5 ± 5	YES
LkHa198	V633Cas	A5	-10.6 ± 11	NO
LkHa208	LkHa208	A3	2.2 ± 5	NO
MWC1080	MWC1080	B0	31.0 ± 3	MAYBE
MWC137	MWC137	B0	76.0 ± 9	YES
MWC297	MWC297	O9	20.4 ± 1	MAYBE
MWC300	V431Sct	Be	21.0 ± 8	NO
MWC480	HD31648	A2	5.0 ± 6	NO
MWC497	HKOri	A4	2.2 ± 1	NO
MWC758	HD36112	A3	3.4 ± 1	NO
MWC763	TOri	A3	1.0 ± 2	MAYBE
MacCH12	HBC1	A5	5.1 ± 1	NO
NGC2245	LkHa215	B7	3.9 ± 1	MAYBE
RMon	RMon	B0	-12.8 ± 3	NO
RNO6	HBC334	B1	11.0 ± 1	NO
V1012Ori	V1012Ori	B9	1.9 ± 2	NO
V1271Ori	HD245185	A2	4.5 ± 5	YES
V645Cyg	V645Cyg	O7	29.5 ± 2	NO
XYPer	XYPere+w	B6	11.3 ± 3	NO

Spectral type and I_C values were taken from Testi *et al.* (1999).

This indicator (I_C) is found by measuring the local density of a source within a specific radius from the position of the HAeBe star. The result obtained shows that some HAeBe stars are located in areas with a density enhancement; while, other areas present a almost constant density (Testi *et al.*, 1997).

Figure 3.14 reproduces the results obtained by Testi *et al.* (1999) and plots the richness indicator (I_C) as function of spectral type for the 43 HAeBe stars of the sample in common with CEREAL, which are also represented in the figure. The figure is divided into three sections which show the limits defined by Testi *et al.* (1999) indicating whether a HAeBe star was in an rich cluster ($I_C \gtrsim 40$), a small cluster ($10 \lesssim I_C \lesssim 40$) or in small aggregates or are background stars ($I_C \lesssim 10$; Testi *et al.*, 1999).

For the zone of rich clusters ($I_C \gtrsim 40$) both catalogues appeared to agree that the stars were located in clusters. These stars were 3 B-type stars that were likely to be

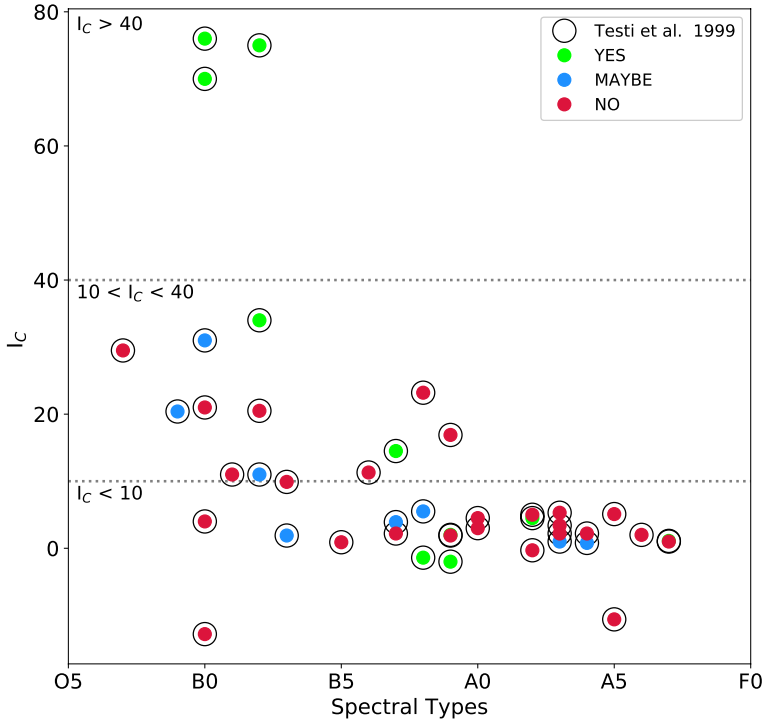


Figure 3.14: Correlation between Spectral type and the richness indicator, I_C, from Testi *et al.* (1999). Open black circles are the stars from Testi *et al.* (1999). The CEREAL classification is represented by: full green circles (*YES*), full blue circles (*MAYBE*) and the full red circles (*NO*).

the most massive stars in their clusters. Also, it is known that 2 of these (MWC137 and V361Cep) are in fact in known clusters (SH 2-266 and NGC 7129, respectively; SIMBAD, Mehner *et al.* (2016)).

For the other two areas the situation is a bit different. The samples cover the ranges of spectral types from O9 to B9 which included the definition by Testi *et al.* (1997) where the clusters appeared only around stars earlier than B5-B7. CEREAL classifies 5 out of 12 stars in this range to be in a cluster, one of these is located in a known cluster (V590Mon in NGC 2264; SIMBAD). The remaining 7 stars have been classified as not being part of a cluster, while Testi *et al.* (1999) reported that these could be part of a small cluster. Of these, 2 out of 7 have been reported to be part of a molecular cloud and star forming region (HBC334 and VVSer, respectively; SIMBAD).

Finally, for the other regions where $I_C \lesssim 10$, CEREAL found 10 of the 28 stars to be in a cluster, of which 4 were A-type stars and 2 of them were part of a known cluster (HD245185 and TOri, Collinder 69 and Orion Nebula, respectively; SIMBAD).

Overall, CEREAL found similar results to those reported by Testi *et al.* (1999), for the sample covering the spectral types from O9 to A7 (see table 3.8). Up to a certain point, the results of CEREAL agreed with the spectral type limit defined by Testi *et al.* (1999), that stars earlier than B7 could possibly have a cluster around them and that stars later than B7 were not associated with a cluster.

From the sample of 43 stars, CEREAL found 18 stars in clusters, classified as yes and maybe, of which 13 stars were B-type stars and 4 stars are A-type. However, there were differences with the results found by Testi *et al.* (1999) around the A-type stars. Testi *et al.* (1999) analysed 16 A-type stars which appeared in the region which defined the HAeBe star to be in a small aggregate or a background star ($I_C \lesssim 10$; see figure 3.14). This implies that Testi *et al.* (1999) did not find any rich clusters around the A-type stars. In contrast, CEREAL found 4 of 16 A-type stars in clusters, where 2 of these stars (HD245185 with spectral type A2 and TOri with spectral type A3) were located in known clusters (Collider 69 and M42, respectively; SIMBAD).

This result of finding clusters around the A stars might be related to the position of these stars in a known star forming region, or it may be the discovery of a new aggregation where the A star is located. Another possibility is that the A stars were not necessarily the most massive star in their cluster, which means they were not the centre of the cluster. Further analysis will be carried out later in this work to explore this point in detail.

3.5 Conclusion

The capability of CEREAL at finding clusters using Gaia DR2 data has been shown, in the analysis carried out here around the sample of 269 known HAeBe stars.

From this sample, $76(28\%)$ stars were either found to be in or to potentially be in a cluster. It was non trivial to classify these stars because there was not an agreed well defined formula on which to categorize the presence of clusters around HAeBe stars. Even if the number of HAeBe stars found to be in clusters might appear small in number, this still represents a big improvement on the 15 stars found by Testi *et al.* (1999, § 3.4) to be in either rich ($I_C \gtrsim 40$) or small clusters ($10 \lesssim I_C \lesssim 40$).

The Herbig Ae/Be stars in table 3.1 represent the first catalogue of HAeBe stars found in clusters by applying a single tool over the same data homogeneously.

The comparison between CEREAL and the DCAs show that using the same data but different quality controls, might lead to different results, such as finding the same clusters but not the same member stars or not finding a cluster by applying different quality techniques to extract the sources before the analysis.

This comparison also showed that using the known value of each parameter as a starting point gave CEREAL an advantage in the assessment of the presence of clusters around the HAeBe stars. In addition, CEREAL gave more freedom to the user for the selection of the parameters to analyse the sample. Further, the reliability of the algorithm as a method to find clusters, of any size, from any sample has been clearly demonstrated.

Another important difference was that CEREAL was able to find clusters around 4 A-type stars in the common sample with Testi *et al.* (1999), where 2 of these stars were located in known clusters. The 4 stars were part of the 16 A-type stars included in the sample from Testi *et al.* (1999) which were considered in their analysis to be either in a small aggregate or to be background stars ($I_C \lesssim 10$; see figure 3.14).

The different result obtained for the sample from Testi *et al.* (1999) by CEREAL might be due to the selection of a larger radius around the central sample or the position of the HAeBe star in known star formation region. Moreover, using the data from Gaia DR2 and a large cone search around the HAeBe stars with CEREAL has proved that it was possible to find more clusters around the same objects in the sample from Testi *et al.* (1999).

Chapter 4

Cluster properties of the Herbig Ae/Be stars

This chapter investigates the properties of the clusters found around HAeBe stars, focusing the analysis on the relationship between the spectral types of the targets and their parameters such as their parallax and G magnitude. In addition, it will explore whether a change in the parallax or G magnitude (or a combination of both), affects the cluster properties of the sample.

Finally, this chapter investigates if there is a difference between the clustering fractions of the B and A type stars. Then, it will consider whether the position of stars within their clusters has any effect on their cluster properties.

4.1 Parallax

This section will be focused on the influence of parallax on the clustering properties. Figure 4.1 represents the parallax distribution of 255 stars from the sample, whose parallax values were also reported by Gaia DR2, and were studied with `CEREAL`. It is apparent in figure 4.1 that, across the full range of parallax values, the majority of stars were found to not be in clusters.

Although, the majority of stars at all parallaxes appeared to not be in clusters, there did appear to be a parallax dependence for the stars found to be either in or maybe in a

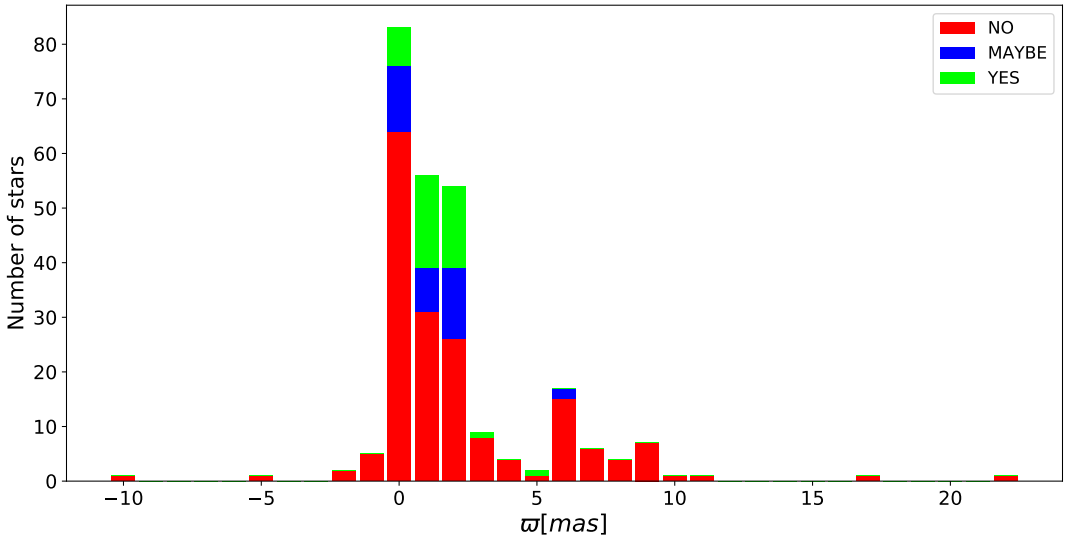


Figure 4.1: Histogram of the parallax distribution of the Herbig Ae/Be stars. The colours represent the different classifications made by CEREAL for the stars found in clusters (green bars), possibly found in a cluster (blue bars) and not in a cluster (red bars).

cluster. The stars that are found to be located in clusters, or possibly in clusters, appear to be located in a specific region of the graph, between 0 and 7 *mas*.

There is a possibility that clusters, or possible clusters, can only be detected in a narrow range of parallaxes, due to a bias in the method with which CEREAL evaluate the sample. It is possible that the algorithm has been less effective at finding clusters around objects that are either very far away or very close by.

Another way to consider the effect of parallax on the clustering fractions within the sample was to consider the number of stars found to be in, maybe be in, or to not be in clusters. This is displayed in figure 4.2, which presents the parallax distribution of the fraction of stars that were classified by CEREAL to be part of a cluster or not, with statistical error bars. For the stars that were found to not be in clusters, their fractions appear to be similar within their error bars over all parallaxes. The stars classified as yes and maybe appear to be located in a specific range of parallax values. However, additional analysis is required to verify if the fraction of stars that were found to be, or possibly be, in clusters being observed only within a narrow range of parallaxes was statistically significant or not.

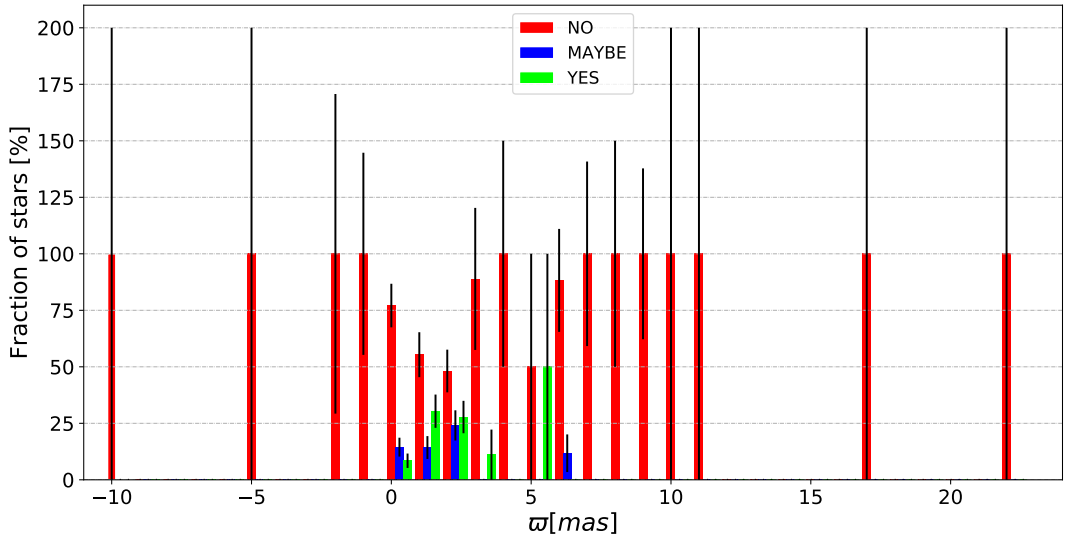


Figure 4.2: Parallax distribution of the fraction of the stars classified by CEREAL. The colour represents the different classifications made by CEREAL for the stars found in clusters (green bars), possibly found in a cluster (blue bars) and not in a cluster (red bars).

As it is not clear from figure 4.2 if there is a parallax dependence for the stars found in clusters and those possibly found to be in clusters, a statistical test is required. The test chosen to investigate this uncertainty, is the Kolmogorov-Smirnov test (KS test; Field, 2016; Hodges, 1958; Wall & Jenkins, 2003) which was applied to the sample of stars found to be in clusters and found to possibly be in clusters. In general, the KS test is used to examine whether a distribution is significantly different from a normal distribution (Field, 2016). The result of this statistical analysis is then used to identify if a potential bias is present in the sample in parallax.

For this study, the package *ks2samp* from *scipy statistic modulo* was used (Hodges, 1958). This package tests the null hypothesis that two independent samples are drawn from the same continuous distribution. The results obtained from this test are the maximum difference between the two cumulative distributions *(D)*, and a statistical probability value *(pvalue)*. If the *pvalue* is large ($\geq$ 0.10), and the D is small, then it is likely that one sample is consistent with having been drawn from the same distribution as the other sample. If however, the *pvalue* is small ($\leq$ 0.10) and the D value is large, then the sample does not appear to be drawn form the same continue distributed.

Figure 4.3 shows the cumulative distribution of the B-type and A-type stars, compared with each other and their clustering properties found by CEREAL. The left panel shows the distribution of B stars against A stars. It is evident that the A stars exhibit larger parallaxes ($>$ 10 mas, which meant that they were located closer) than the B stars. The distribution of the B and A stars appeared to be different due to their small pvalue of 2.7x10^{-09}.

In Figure 4.3, the middle and right panel present the distributions of the cluster properties of B-type and A-type stars. This figure shows that the stars that were found to not be in clusters had a different parallax distribution to the stars that were found to be present in clusters. In particular, those objects found not to be in clusters, were present at both low and large parallax values, where those found to be in clusters, were not.

When comparing the parallax values of the B and A type stars found to be in clusters and their KS test results (pvalues of 0.0155 and 0.0143, respectively), these samples appeared different and not to be drawn from equivalent distributions. These differences prevented accurate comparison of the cluster properties of the B-type and A-type stars being made without further analysis. These differences can be due to an observational selection effect, where the bright stars ($\sim\leq$ 8 mag) do not have a cluster around them or

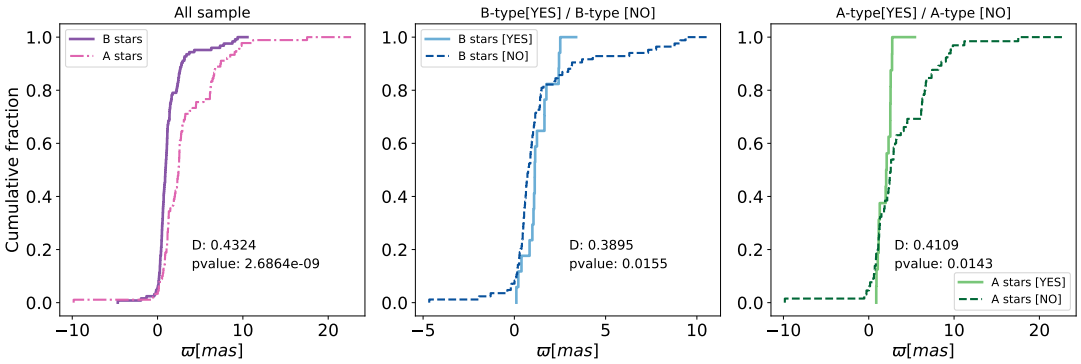


Figure 4.3: KS test for the parallax distribution of the sample of B and A stars. The distribution of the B-type and A-type stars were compared with each other and by their clustering properties found by CEREAL. In the left panel, the solid line represents the B stars and the dashed-dotted line represents the A stars. In the middle and right panels, the solid line represents the B and A stars found in clusters (classified as *yes* by CEREAL) and the dashed line represents the B and A stars found to not be in clusters (classified as *no* by CEREAL).

their brightness outshines their less luminous neighbours; or due to that the number of bright stars in the sample is too small. Perhaps, this effect can be also related to the faint stars, that are to far away and is not possible to detect the possible clusters members around the HAeBe stars.

To have a fair comparison between the stars found in clusters and the stars found not to be in clusters, it is important that the stars compared have the same parallax distribution. It is therefore necessary to remove almost all the nearby ($\geq$7 *mas*) and distant ($\leq$ 0 *mas*) objects from figure 4.3. With this selection, it was possible to obtain a new sample where the B-type and A-type stars, both in and not in clusters, had the same parallax distribution, which helped to remove this potential bias in the sample. This new range where the parallax distribution was truncated between 1 and 6 *mas* reduced the sample from 255 stars to 125 stars.

Figure 4.4 represents the cumulative function for the sample of B-type and A-type stars which have been selected with parallaxes in-between 1 and 6 *mas*. The left panel of 4.4, still showed that the B and A stars had different population distributions (pvalue=0.0018), which meant they were different samples. This result could be explained by the fact that the B stars were typically located further away than the A stars.

Although, the left panel in figure 4.4 still showed the B and A stars appeared to be differently distributed in parallax, the middle and right panels (which compared the stars

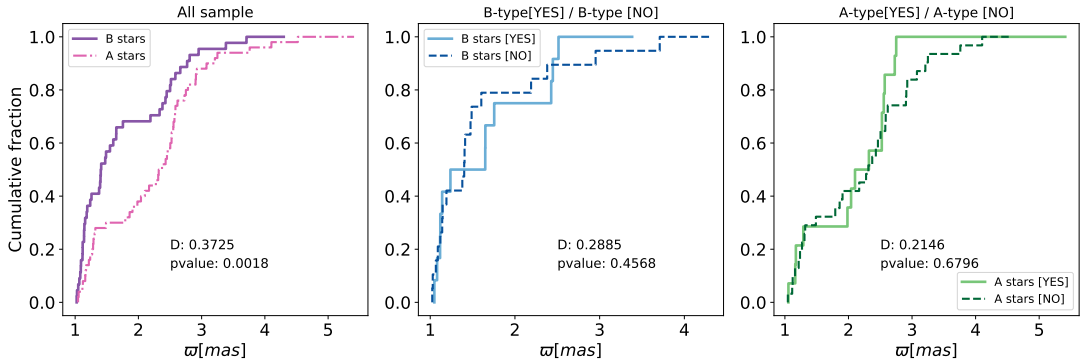


Figure 4.4: KS test for the parallax distribution of the sample of B and A stars selected between 1 to 6 *mas*. The distribution of the B-type and A-type stars are compared with each other and by their clustering properties found by `CEREAL`. The lines have the same meaning as the ones used previously in figure 4.3.

found to be in or not be in clusters, for the B and A types respectively) showed a large positive change when compared with the same panels in figure 4.3. It appeared that due to the reduction in the sample, these stars could now be considered to be consistent with being drawn from the same parent distribution, given that their pvalues were much larger than 0.10 (0.4568 and 0.6796, respectively).

The selection made in parallax, between 1 and 6 *mas*, reduced the possibility of a bias in this parameter affecting the sample. Therefore, this selection produced a sample of B and A stars that was more homogeneous, which could then be better compared with each other.

4.2 G magnitude

Following the analysis carried out in the previous section (§4.1), an assessment for the presence of any bias was also performed for the G magnitudes. Figure 4.5 shows the G magnitude distribution of the 269 Herbig Ae/Be stars that were classified by `CEREAL`.

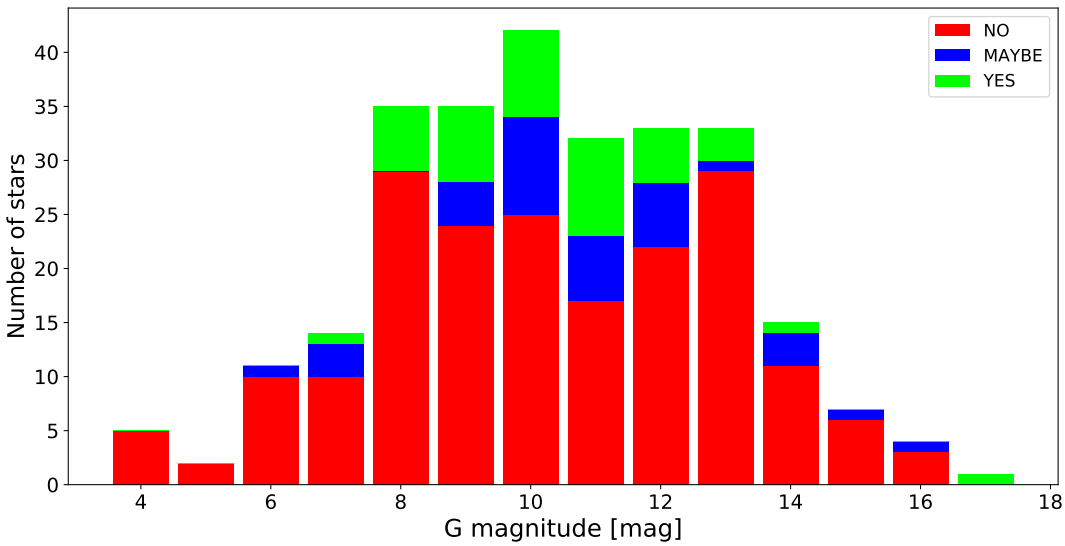


Figure 4.5: Histogram of the G magnitude distribution of the sample of Herbig Ae/Be stars. The colours represent the different classifications made by `CEREAL` for the stars found in clusters (green bars), possibly found in a cluster (blue bars) and found to not be in a cluster (red bars).

The stars found to be in clusters and those found to not be in clusters both covered the same range of magnitudes; this is similar to that observed for the analysis of the influence of parallax in the sample carried out earlier. Here, similar to the observation for the effect of parallax above, the objects found to be in clusters seemed to be found preferentially within a specific magnitude range of the graph (between 6 and 18 mag). A more detailed examination of how the G magnitude affected the clustering properties was therefore also required.

Figure 4.6 presents the G magnitude distribution of the stars in clusters that were classified by CEREAL to be, to possibly be, or to not be part of a cluster (Yes, Maybe, No). The fraction of stars that were found to not be in clusters in figure 4.6 was similar over all G magnitude values, within their error bars. Although, it did appear visually that the stars found to be in or to be possibly in a cluster, were primarily within the magnitude range between 6 and 18 mag. Therefore, it was necessary to perform further analysis to verify if there was a statistically significant difference in the distribution of the stars classified as yes and maybe in G magnitude.

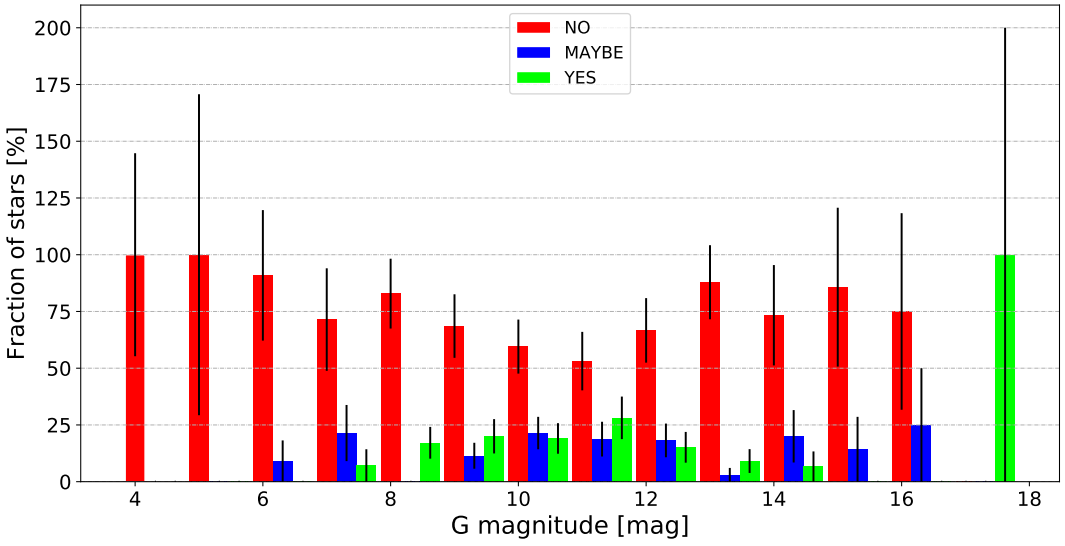


Figure 4.6: G magnitude distribution of the fraction of the stars classified by CEREAL. The colours represent the different classifications made by CEREAL for the stars found in clusters (green bars), possibly found in a cluster (blue bars) and not in a cluster (red bars).

Only finding clusters within a specific range of G magnitude might be interpreted as a bias, this could be caused by the algorithm being potentially less effective at finding clusters around targets that were either too bright or too faint. Here, as in section 4.1 the KS test was again used to analyse the presence of a bias in the sample.

Figure 4.7 shows the cumulative distribution of the B-type and A-type stars, compared with each other and their clustering properties found by CEREAL. The left panel shows the distribution of the B stars and the A stars which appeared to be drawn from the same distribution (pvalue $= 0.2475$). The middle and right panels present the distributions of the cluster properties of the B and A stars. Although, the distributions of the B stars found to be in clusters appeared visually to present a narrower range of G magnitude than those found to not be in clusters, statistically they are consistent with being drawn from the same sample, due to their pvalue being ≥ 0.10 (0.4413). However, the A-type stars that were found to be in or to not be in clusters both visually and statistically appeared to have a different distribution, with their pvalue ≤ 0.10 (0.0362). This difference in the apparent distribution of the A stars, may be due to a limitation in the algorithm at finding clusters around the bright A stars or it could be interpreted as the presence of a different bias in the sample.

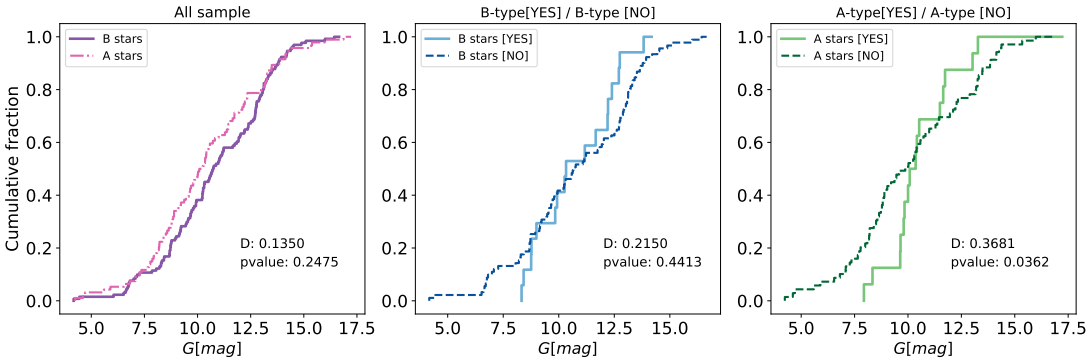


Figure 4.7: KS test for the G magnitude distributions of the sample of B and A stars. The distributions of the B-type and A-type stars are compared with each other, and the distributions of the stars found to be in and to not be in clusters by CEREAL are also compared for both types. In the left panel, the solid line represents the B stars and the dashed-dotted line represents the A stars. In the middle (B type) and right (A type) panels, the solid lines represent the stars found to be in clusters (classified as *yes* by CEREAL) and the dashed lines represent the stars found to not be in clusters (classified as *no* by CEREAL).

To obtain a sample of stars which had the same G magnitude distribution, whether they were located in clusters or not, it was necessary to remove almost all the bright (≤ 7 mag) objects. With this selection, it was possible to find a new sample where the cluster properties were similar. This occurred when the G magnitude was restricted to between 7 and 17.5 mag. The faint limit was fixed to 17.5 mag as this was the magnitude of the faintest object in the sample.

Figure 4.8 represents the cumulative fraction for the sample of B-type and A-type stars which were selected with G magnitudes between 7 and 17.5 mag. This selection slightly reduced the original sample from 269 stars to 251 stars.

In figure 4.8, the left and middle panels appear to show a similar behavior to the equivalent panels in figure 4.7. The B and A stars still had similar population distributions (pvalue $= 0.3150$) and the distribution of the B stars found to be in and to not be in clusters were again similar (pvalue $= 0.3174$), which meant they appeared to be drawn from the same sample. The right panel of figure 4.8 differs from that in figure 4.7, with the distribution of A stars, found to be or not to be in clusters, now appearing to be consistent with being drawn from the same sample (pvalue $= 0.5742$), following the selection between 7 and 17.5 mag.

The selection made in the G magnitude, between 7 and 17.5 mag, was use to explain the selection effect on the data and to reduced the possibility of finding any bias, due

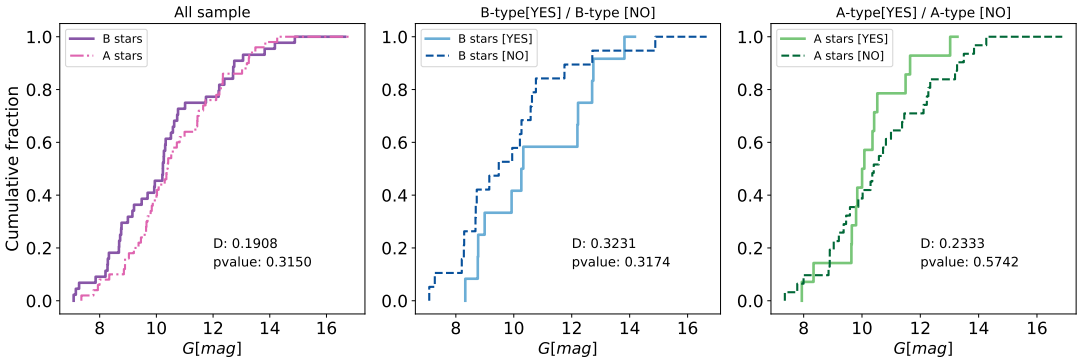


Figure 4.8: KS test for the G magnitudes of the sample of B and A stars selected between 7 to 17.5 mag. The distribution of the B-type and A-type stars are compared with each other and then by their clustering properties found with CEREAL. The lines have the same meaning as the ones in figure 4.7.

to the G magnitude distribution, in the sample of B and A stars and produced a more homogeneous distribution that could then be better compared with each other.

4.3 Combining the parallax and G magnitude selection ranges

The analysis made in the previous sections for the parallax (§ 4.1) and G magnitude (§ 4.2) showed that it was possible to find a sub-sample of the B-type and A-type stars that could then be compared with each other. The next question was then, what happened when both selections, in parallax and G magnitude, were combined?

The following section therefore, analyzes a new sub-sample formed from the limits chosen previously in both parallax (1 to 6 mas) and G magnitude (7 to 17.5 mag).

This sub-sample is represented in figure 4.9, where each histogram shows the parallax and G magnitude of the stars classified by `CEREAL` to be either part of, or not part of a cluster. As in the previous analysis, the stars that were classified as no appeared to cover the full range of each parameter. However, the parallax distribution indicate that the stars classified as yes and maybe seem to be confined to a smaller range. In contrast, in the G magnitude distributions the stars classified as yes and maybe also appear to cover the full range in magnitude, as was seen for those found to not be in clusters.

A similar behaviour was also observed in figure 4.10 which represented the distribution of the fraction of the stars that were classified by `CEREAL` to be in (*yes*), potentially be in (*maybe*) or not be in (*no*) a cluster. The fraction of stars that were found to not be in clusters were similar over all parallaxes and G magnitudes, within their error bars. For those stars that were classified as yes and maybe further analysis was again required, to confirm if their behavior was similar to the results obtained in their previous sections (§ 4.1 and § 4.2). As in the previous sections, the KS test was used to confirm if the combination of the selection ranges for both parameters, between 1 to 6 mas and 7 to 17.5 mag, produced a sample that had any statistically significant differences in its distributions that could be interpreted as a bias in the sample in either parallax or G magnitude.

Figure 4.11 shows the result obtained with the selections made in both parallax and G magnitude. The upper panel shows the result of the KS test for the parallax. On the left panel, the distribution of the B-type and A-type stars still shows that their samples appeared to come from different distributions (pvalue = 0.0018) whilst in the middle and right panels, the stars in clusters and not in clusters had similar distributions (pvalue=

0.4568 and 0.6796, for the B and A stars respectively). The bottom panel in figure 4.11 shows the results of the KS test for the G magnitude where all the distributions had a pvalue larger than 0.10 (pvalue= 0.3150, 0.3174 and 0.5742; respectively), which meant they were consistent with being drawn from the same parent population.

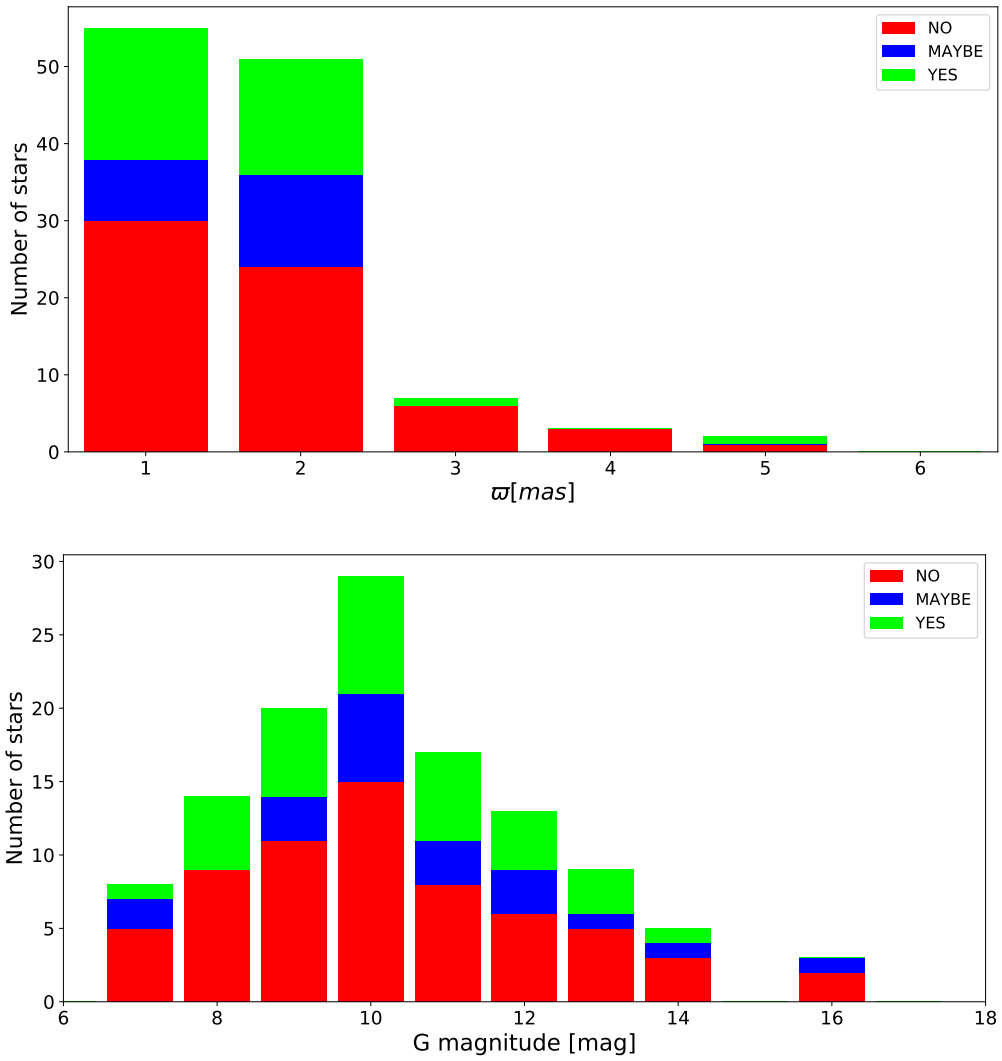


Figure 4.9: Histogram distributions for the objects selected between 1 to 6 mas and 7 to 17.5 mag. The colour of the bars represent the clustering classification made by CEREAL; green bars for *yes*, blue bars for *maybe* and red bars for *no*. The *upper panel* presents the parallax and the *bottom panel* presents the G magnitude.

This prove that the combination of the selections provided similar results to those obtained in the individual analysis of the parallax and G magnitude, where the possibility of a bias in the sample is reduced and a more homogeneous distribution of stars is generated; where the stars found to be in a clusters or not, appeared to have the same distributions and can therefore be compared. However, these new selections reduced the original sample from 269 to 118 objects.

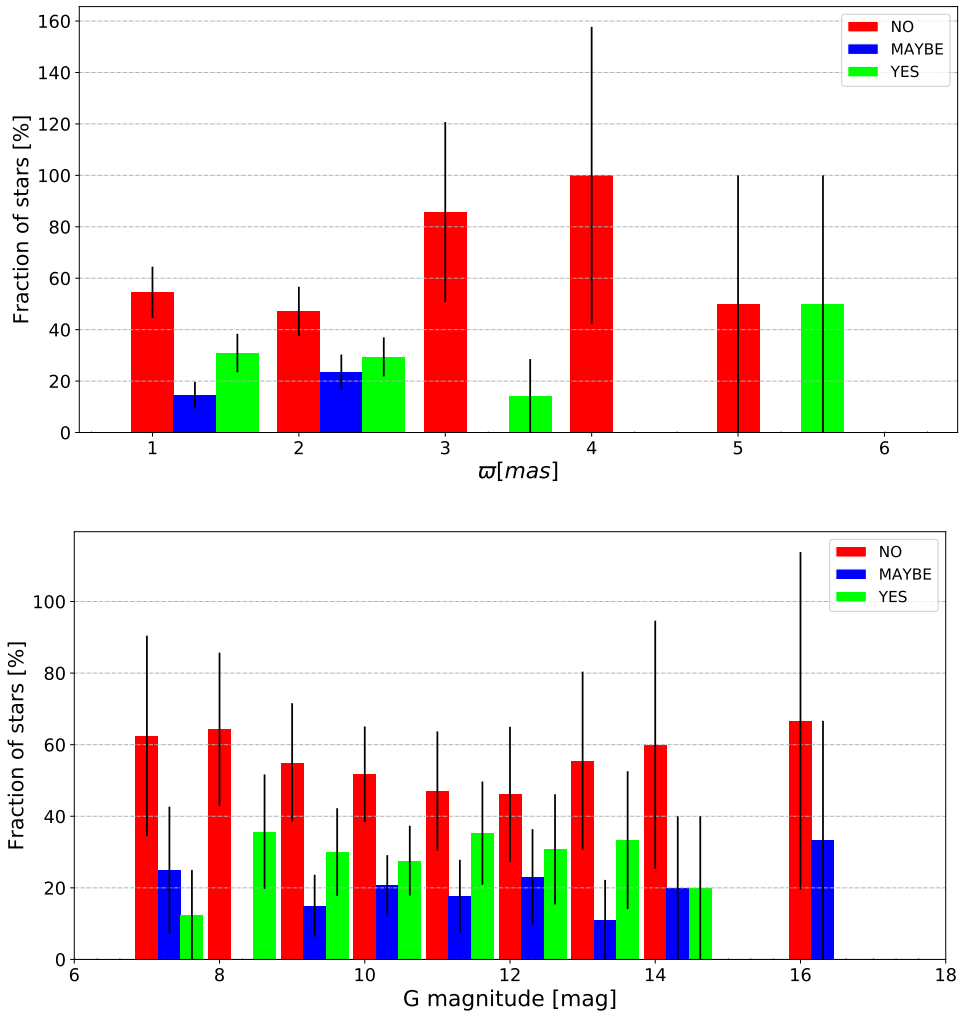


Figure 4.10: Distributions of the fraction of the stars in clusters selected between 1 to 6 mas and 7 to 17.5 mag. The *upper panel* represents the parallax and the *bottom panel* represents the G magnitude. The colours of the bars represent their clustering classification made by CEREAL; the green bars for *yes*, blue bars for *maybe* and red bars for *no*.

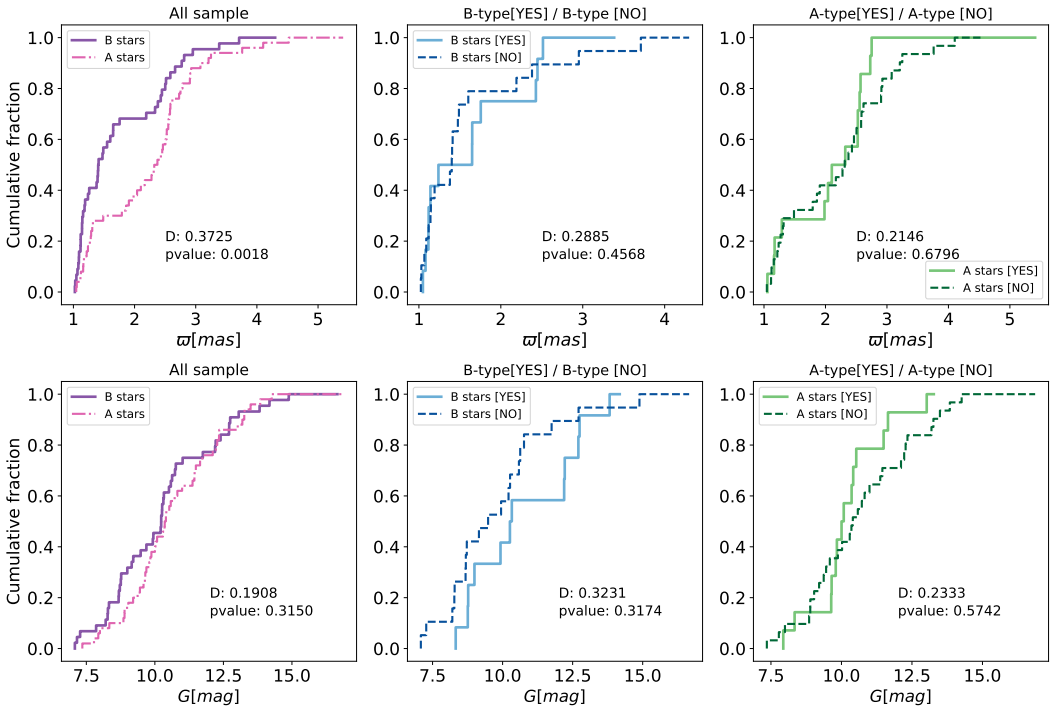


Figure 4.11: KS test for the sample of B and A stars selected between 1 to 6 mas and 7 to 17.5 mag. The distribution of the B-type and A-type stars are compared with each other and by their clustering properties, found by CEREAL. The distributions for Parallax are located in the upper panel and for G magnitude in the bottom panel. In the left panel, the solid line represents the B stars and the dashed-dotted line represents the A stars. In the central and right panels the solid lines represent the, B and A, stars found in clusters (classified as YES by CEREAL) and the dashed lines represent the, B and A, stars found to not be in clusters (classified as NO by CEREAL).

4.4 Differences between the clustering fraction of the B and A stars

From this point onward, the analysis of the objects will be split into 4 groups: the full sample of 269 Herbig Ae/Be stars; the sample selected in parallax between 1 to 6 mas; the sample selected in G magnitude between 7 to 17.5 mag; and, the combined sample selection with a parallax between 1 and 6 mas and with a G magnitude between 7 and 17.5 mag. In addition, the clustering properties found by CEREAL for the stars found to be in clusters (*yes*) and possibly to be part of a cluster (*maybe*), were combined, and both

were subsequently considered as stars found in clusters (*yes*).

The analysis performed previously in this chapter (§4) has shown that the clustering fractions of the sample of HAeBe stars have similar distributions (see figures 4.2 and 4.6). However, a deeper examination of the B type and A type stars have shown slight differences between these types. These differences were corrected for by making a selection in both parallax (§4.1) and G magnitude on the sample (§4.2). The result of these selections is that the clustering fractions remained similar (see figure 4.10) even when the sample was reduced from 269 to 118 objects (§4.3).

The fraction of stars classified by CEREAL to be yes or no are given in figure 4.12 for the full and reduced samples. With the upper panel representing the the full sample of B and A stars and the bottom panel those selected between 1 to 6 mas in parallax and between 7 to 17.5 mag in G magnitude.

In both samples, the fraction of stars that were classified as yes and no were similar for both the B and A stars, within their error bars. This result allowed for further analysis to be performed where both samples could be directly compared.

4.4.1 Investigating the position of B and A stars within their clusters

The analysis carried out in the previous sections of this chapter has been focused on understanding the clustering properties of the sample of the HAeBe stars. In particular, this analysis was focused on the B and A stars. It has been shown that by making different selections for the parallax and G magnitude (§ 4.1, 4.2 and 4.3) of the B and A stars did not appear to change the clustering properties (§ 4.4) of the stars. These selections then produced a sample of stars with which it was possible to make a direct comparison between the B and A type stars.

The results from CEREAL appear markedly different from those obtained by Testi *et al.* (1997, 1998, 1999, see section §3.4). This is illustrated in figure 4.12, where the fraction of clusters for the B and A stars appear similar, within their error bars. However, Testi *et al.* (1997) found that clusters were only detected for stars with spectral type B7 or earlier, and were not detected around A stars.

Testi *et al.* (1997) confirmed the presence of clusters using radial profiles, which they searched for clusters in small areas (4-7 arcmin) around the HAeBe stars. In contrast, CEREAL searched for clusters using larger areas ($>$ 30 arcmin). Therefore, the results

of Testi *et al.* (1997) were more sensitive to the location of the stars, as they expected that these stars were the centre of their own clusters. In contrast, while CEREAL used the HAeBe stars as the centre of the search, in processing Gaia DR2 data, it does not assume that they were the centre of their possible clusters. This difference then give rise to the question, is the location of a star within its cluster dependent on it spectral type?

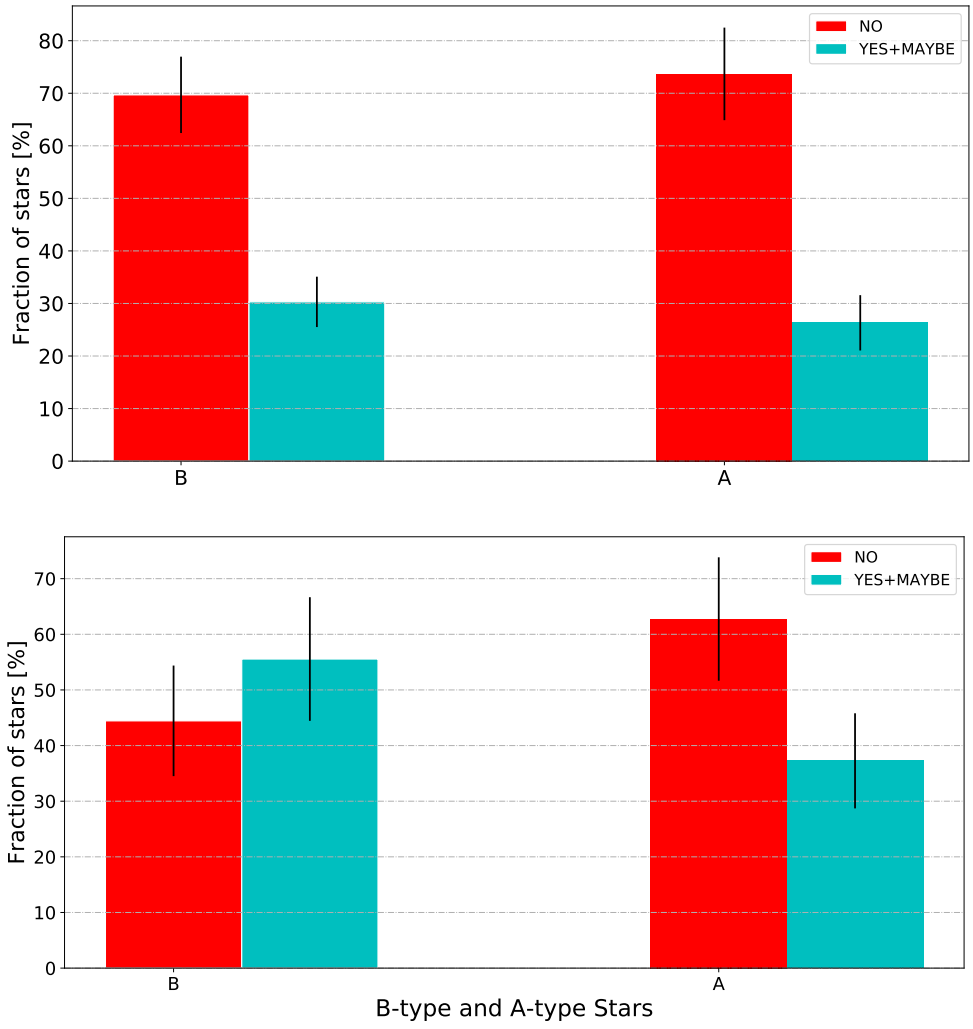


Figure 4.12: The distribution of the fraction of B and A stars in clusters found by CEREAL. The red bars are the stars found to not be in clusters and the cyan bars are the stars found to be in clusters (*yes*) or to be possibly in clusters (*maybe*). The *upper panel* presents the full sample and the *bottom panel* presents the sample selected between 1 to 6 mas and 7 to 17.5.

To examine whether the spectral type of the HAeBe star had any effect on their position within a cluster, two different analyses were performed. Firstly, a re-analysis of the radial profiles obtained with `CEREAL`, and; secondly, a combination of the five astrometric parameters of the HAeBe stars obtained from Gaia DR2 (position (RA, DEC), parallax and proper motion in RA and DED) were used to evaluate the position of the HAeBe stars in their clusters

Radial profiles

The first method was the re-analysis of the radial profiles obtained with `CEREAL`. The radial profiles were constructed by measuring the local densities of the sources in an annular radius from the position of the HAeBe star.

Figure 4.13 shows 3 examples of radial profiles from the sample. This figure shows the typical behaviour observed for 3 different HAeBe stars. These radial profiles were used to evaluate the presence of clusters around the HAeBe stars. Each point in the graph represents the density found at that annular radius.

Panel **(A)** in figure 4.13 shows the radial profile of the HAeBe star MWC137. This

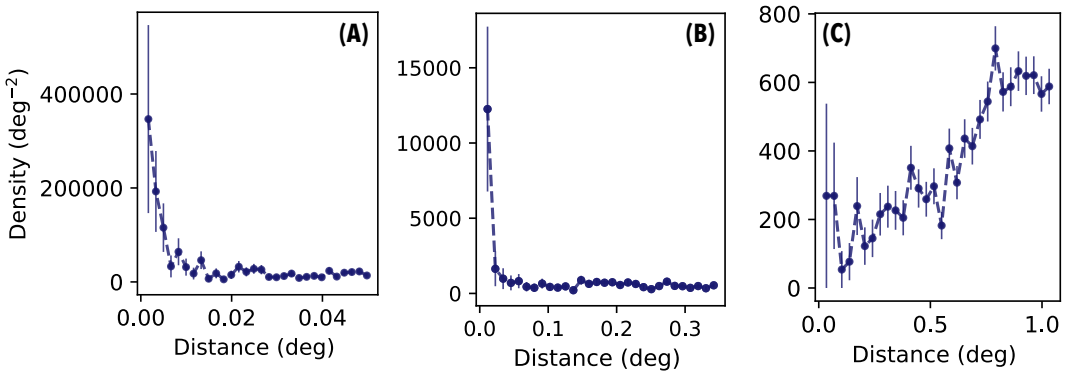


Figure 4.13: Different radial profiles of the HAeBe stars. These profiles were used to confirm whether the HAeBe stars were located at the centre of their cluster or not. Panel **(A)** shows the star MWC137, which has an increased density beyond the initial point. Panel **(B)** shows the star HD290770 which has an extreme drop in the density after the central source, with the remainder of the density seeming to be constant; and, in panel **(C)**, the star HD37411 shows an unexpected increase in the density,that is potentially unrelated to the presence of a cluster.

panel shows an overdensity around the central star. At each point on the graph, the density decreases as the distance from the HAeBe star increases. This could mean that the HAeBe star was located at the centre of its cluster, and was surrounded by a considerable number of low mass companions. Panel **(B)** shows a density profile for a star that might not have a cluster around it. There is a drop in the density after the central point, while the remainder of the points have much lower densities, which seem to be an almost constant baseline. This could mean that the HAeBe star was not located at the centre of a cluster and that the other stars might represent foreground and/or background stars.

Finally, panel **(C)** showed an unusual behaviour in the density profile, which increased with the distance. This did not seem to be related to the presence of a cluster. Instead, it could be due to the presence of another (brighter) object, or other sources of contamination that hindered the selection of the possible lower mass companions around the HAeBe star.

Combination of the five astrometric parameter

The second method combined the five astrometric parameters of the B and A stars, position (RA, DEC), parallax (ϖ) and proper motion ($\mu_{\alpha*}$ and μ_δ). Figure 4.14 shows the spatial distribution and proper motions of two HAeBe stars classified by `CEREAL` to be part of a cluster; the left panels show the absolute values of the proper motion and the right panels show the relative proper motion values (median subtracted), both in grey arrows.

From this plot it appeared clear that both of the HAeBe stars were located in different positions within their clusters. One of the stars was located in the centre of a cluster and it was surrounded by other stars which had similar parallax values, whilst the other one was not located in the centre of a cluster of stars which appeared to have similar parallax value to the HAeBe.

In the left top panel of figure 4.14, the HAeBe star V361Cep, has a parallax value of 1.12mas. Based on their colours the stars around the HAeBe star look like they have similar parallax values (the scale for the parallax values is given by the coloured bar). In addition, the proper motion arrows are all pointing in the same direction. The HAeBe star is located in the centre of the cluster and is surrounded by a large number of low mass stars.

In the left bottom panel, the HAeBe star AS477, has a parallax value of 1.29mas. The stars that share the same parallax value as the HAeBe star are not grouped at the centre,

but are on the left hand side of the plot. The proper motion arrows for this group and the HAeBe star do however, point in the same direction. This showed that the HAeBe star is associated with the cluster, but is not located in the centre.

Additionally, both right panels show the relative proper motion of each HAeBe star (grey arrows). These values were obtained by subtracting the median proper motion value of each HAeBe star from the proper motion of the sample. The median value for the proper motion in RA and DEC for the HAeBe star V361Cep (top right) were -1.95mas/yr and -3.07mas/yr, respectively; and for the HAeBe star AS477 (bottom right)

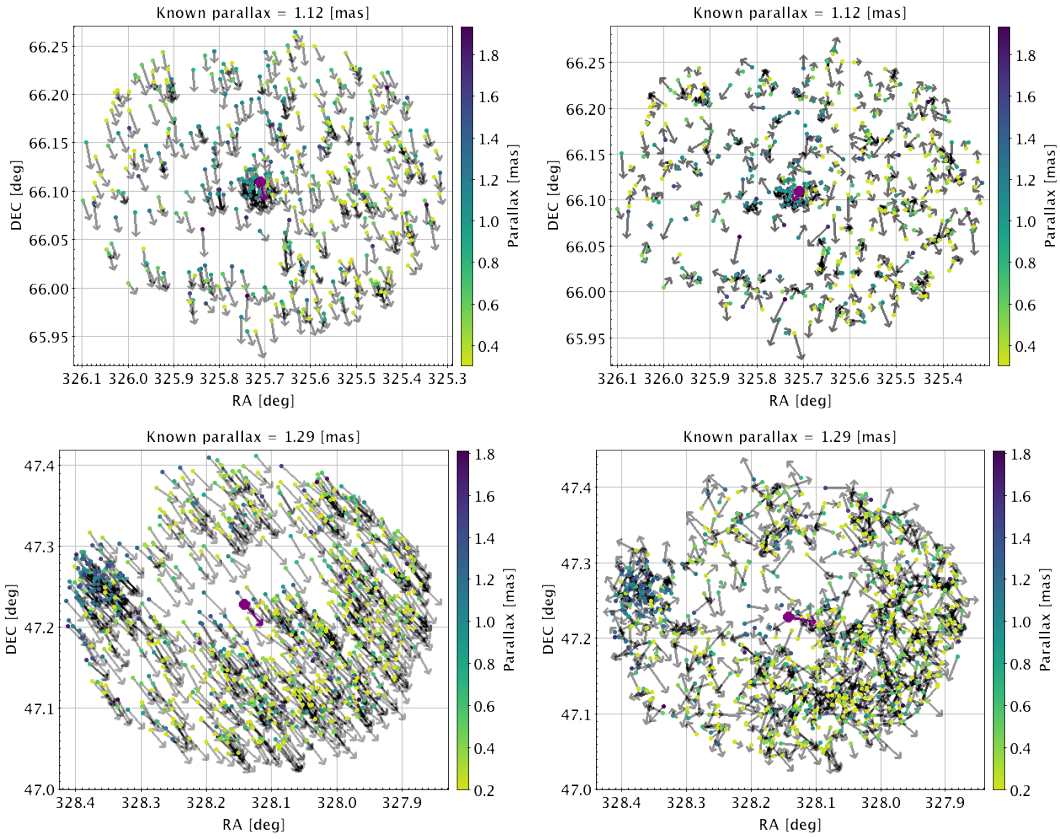


Figure 4.14: Spatial distributions of two HAeBe stars, used to confirm their locations in their clusters. The grey arrows on the left panels represent the absolute proper motion and the right panels show the proper motion of each star when the mean proper motion was subtracted. The coloured bar represents the key for the colour code used to represent the parallax for each star. In the *top panels*, the filled purple circle represents the star V361Cep, which seems to be located at the centre of its cluster; and, on the *bottom panels* the filled purple circle represents the star AS477, which seems not to be located at the centre of the cluster.

were -2.97mas/yr and -2.88mas/yr, respectively. This new arrow represent the relative proper motion of all stars which move in different directions but still some stars that share similar kinematic properties to the HAeBe star are moving in the same direction.

Result

Both of the methods used in figures 4.13 and 4.14 are now applied to the whole sample of HAeBe stars that were found to be in clusters. The result of this evaluation is shown in figure 4.15 which gives the distribution of the fraction of B and A stars classified by `CEREAL` to be either in a cluster or not. This figure also shows whether the B and A stars were located in the centre of a cluster (*IN*) or not (*OUT*). From this figure it was clear that the B stars are much more likely to be located at the centre of their clusters than the A stars. For the B stars, 38 were found to be located in the centre and 1 was not in the centre; and, for the A stars, 5 stars were found to be located in the centre, and 21 were not in the centre.

This shows that although B and A type stars were both found to be present in clusters with a similar distribution, their positions within these clusters differed. The B stars were primarily located at the centre of their cluster. This could mean that as was expected, the B stars tended to be the most massive objects in their clusters, whereas the A stars were more likely to not be located in the centre of their clusters. This could mean that the A stars were not the most massive stars in their clusters.

The result that A stars tended to not be located in the centre of their clusters, coupled with the larger search areas utilized here ($>$ 10 arcmin) compared with Testi *et al.* (1997, 4-7 arcmin), may explain why clusters were observed for A type stars in this study, but were absent in their work.

Following this, a subsequent breakdown of the B and A stars into sub spectral types was carried out and this is given in figure 4.16. This was done to further examine the dependence of the fraction of stars found to be in clusters and their *IN*/*OUT* positioning on spectral type.

From figure 4.16 it was evident that for the sample of B stars found to be in clusters, each sub spectral type was located at the centre of their cluster. Unlike the A stars, where for all sub types the majority that were found to be in clusters, were not then in the centre of these clusters.

In addition to showing the difference between the behaviour of the B and A sub types, figure 4.16 showed no sub spectral type dependence for the clustering fractions of B stars within their errors. This again appears to differ from the results found earlier by Testi *et al.* (1997), where clusters were only found for stars with spectral type B7 or earlier.

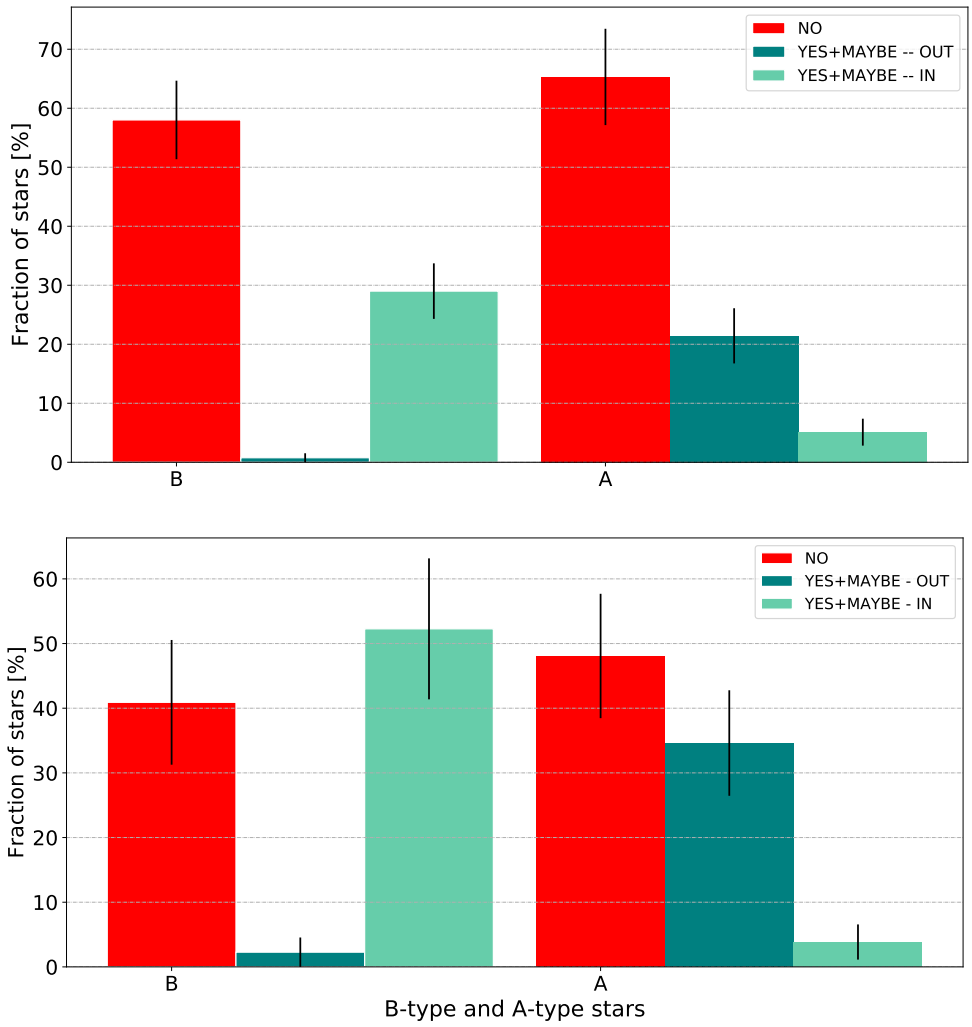


Figure 4.15: The distribution of the fraction of B and A stars located either in the centre of a cluster or not and their clustering classifications given by CEREAL. The red bars represent the stars found to not be in clusters. The dark green bars represent the stars in clusters that were not located in the centre (OUT) and; the light green bars represent the stars in clusters that were located in the centre (IN). The *upper panel* represents the full sample of HAeBe stars and the *bottom panel* represents the sample of HAeBe stars selected between 1 to 6 mas and 7 to 17.5 mag.

Further analysis of the B sub spectral types was then carried out to evaluate the presence of any possible trends in the clustering fractions observed here. This was to determine if there was any difference in the clustering behaviour within a specific sub spectral type range, as was seen by Testi *et al.* (1997), or if there was no dependence on sub spectral type here.

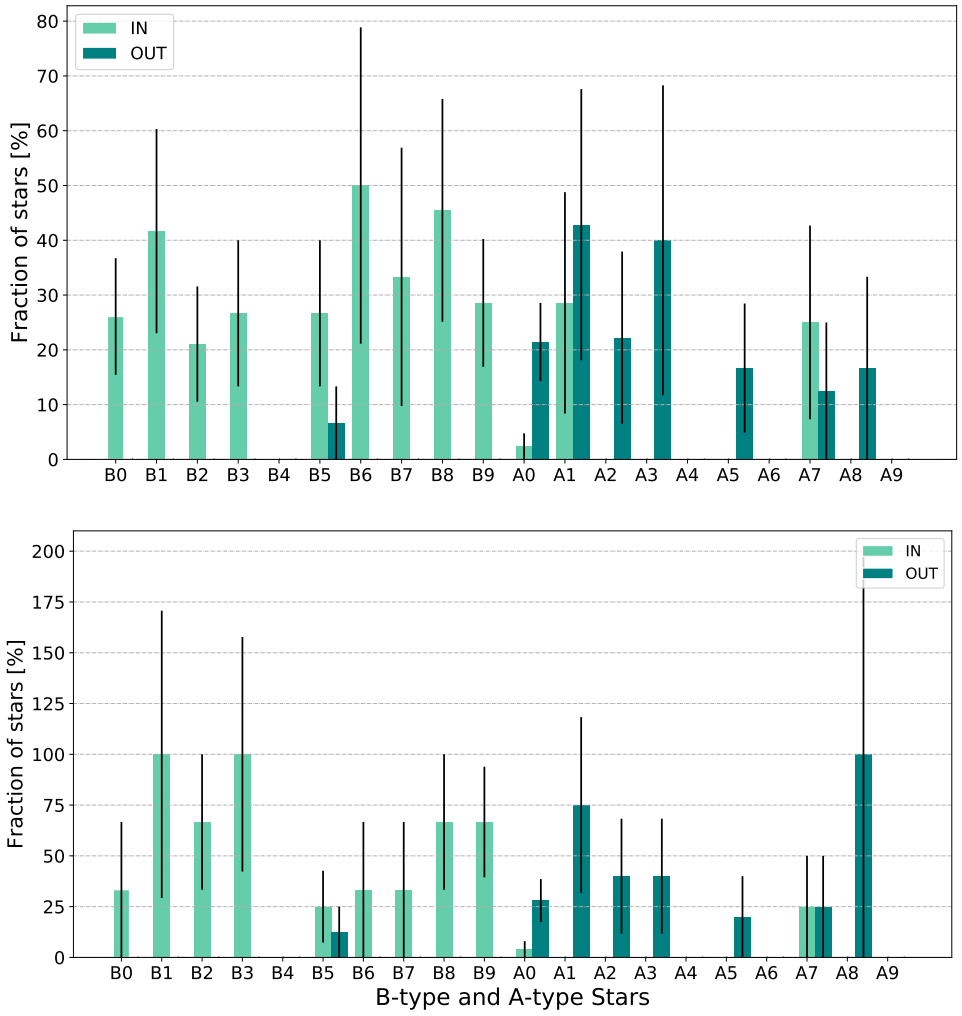


Figure 4.16: The distribution of the fraction of B and A stars located at the centre or not at the centre of their cluster, for all their sub-spectral types. The light green bars represent the stars found in clusters that were located in the centre (IN) and the dark green bars represent the stars in clusters that were not located in the centre (OUT). The *upper panel* represents the full sample of HAeBe stars and the *bottom panel* represents the sample of HAeBe stars selected between 1 to 6 mas and 7 to 17.5 mag.

4.4.2 Examining the sub spectral type distribution of the B stars in clusters.

The analysis carried out in the previous sections showed that the clustering fraction of B stars appeared evenly distributed for all the sub spectral types, within their errors (see figure 4.16). This analysis was carried out both on stars located in the centre of their clusters and those who were not. This result appears to be in contrast with the earlier study made by Testi *et al.* (1997), which defined a detectable level for the presence of clusters for HAeBe stars limited to spectral types B7 and earlier.

Their work used a richness index, I_C, to define how abundant or not the clusters around the HAeBe stars were (see section §3.4). However, when CEREAL analysed the sample of HAeBe stars, which included the targets studied by Testi *et al.* (1999), CEREAL did not find a sharp transition between rich and less rich clusters, as was defined by Testi *et al.* (1997).

To examine the differences between these studies, a closer inspection of the B type stars was made. This was performed to search for any sub spectral type trend in the clustering properties of the B stars, as would be expected from the work of Testi *et al.* (1997). The results of this analysis are shown in 4.17, where the sub spectral types, located in the centre clusters and not, are then sequentially grouped together to evaluate any presence of trends in the data.

The first row appeared to have a homogeneous distribution for all the sub spectral types from B0 to B9, within their error bar. In the following rows, two to four, the sub spectral types were then combined into groups of two, three and four sub spectral types. These plots also showed a homogeneous distribution of the sub spectral type of the B stars, within their error bars. This was the case for the full sample (plots on the left side) and for the sub sample selected in parallax (1 to 6 mas) and G magnitude (7 to 17.5 mag, right panel).

The observation that the fraction of B stars found by CEREAL to be in clusters appeared to be evenly distributed over all the sub spectral types is in contrast with the work of Testi *et al.* (1997). Testi *et al.* (1997) found that clusters around HAeBe stars were limited to the sub spectral type range from B7 and earlier. This difference was not caused by CEREAL observing later spectral type B stars in clusters but not in the centre of their clusters; as all but 1 (LkHa257) of the B stars found to be in clusters were in the centre of their cluster. The difference between the results here and the observation by Testi *et al.*

(1997) may be explained by the larger sample analysed by CEREAL (132 B stars), or may be due to the different sampling methods used (as was discussed earlier in 3.4).

4.5 Conclusion

This chapter has explored the clustering classifications of the sample of HAeBe stars, specifically the B and A stars. The analysis performed has taken into account the influence that certain parameters, such as the parallax and G magnitude, could have had on the cluster classification found for each of the HAeBe stars using CEREAL.

First the impact of the astrometric parameters on the classification of stars found to not be in clusters by CEREAL was considered. The distribution of these stars seemed to cover the full parallax and G magnitude ranges and their fraction appeared constant, within their error bars. This differed with respect to the stars classified by CEREAL to be in, or to potentially be in a cluster; the distribution of these stars appeared to be found in a preferential range in both the parallax and G magnitude.

A deeper analysis of the preferential ranges, where the stars found to be in a cluster were located, was carried out to interpret if this was due to the presence of a bias in the sample. These potential biases were removed by generating a sub sample, containing only those objects selected within specific ranges of parallax (1 to 6 mas) and G magnitude (7 to 17.5). These ranges were chosen due to the nature of the B-type and A-type stars and the data available from Gaia DR2. For parallax, in general the B stars tended to be located at larger distances than the A stars. For the G magnitude, it was not possible to find clusters around HAeBe stars with magnitudes brighter than 6 mag. This could have been due to the nature of bright stars, that they hamper the detection of less luminous neighbours (Ascenso, 2018) and therefore it was not possible to see those.

Subsequently, a statistical analysis, using the KS test, on the parallax and G magnitude distributions of the sample of B and A stars, that were found to be in a cluster or not, was made. This analysis showed that B and A stars appeared to be drawn from the same distribution and could therefore be compared with each other.

This study then investigated if the fraction of B and A stars in clusters differed with the spectral type of the targets. The fraction of B and A stars found to be in, or not in a cluster looked similar. In addition, the fraction of the stars in clusters appear to be distributed evenly over all their sub spectral types, within their error bars.

The effect of spectral type on the position of the HAeBe star within its cluster was also considered. This analysis showed that the objects classified by CEREAL to be in clusters around B stars tended to be at the centre of their clusters and A stars tended to not be in the centre. The A stars in general appeared to be part of a sub-cluster. This was interpreted to mean that the B stars might be the most massive objects in their clusters. The A stars found to be in clusters were mostly not located in the centre of their own cluster, but were present in known star formation regions or known clusters. A comparison of the A type stars found to be in a cluster with SIMBAD showed that 6% of these were found to belong to a known cluster.

Further analysis of each sub spectral type confirmed that over all sub types B stars tend to be located in the centre of their clusters; where for the A stars, over all their sub spectral type they appeared in general to be part of a sub-cluster. The sub spectral type analysis also showed that clusters around the B stars could be found over the full range of sub spectral types; not just around B7 or early, as mentioned by Testi *et al.* (1997). An additional difference between the results found by Testi *et al.* (1997) and this work was that CEREAL found clusters around some A type stars; where Testi *et al.* (1997) found no clusters for spectral types later than B7; these difference may be due to the combination of CEREAL searching larger areas around the HAeBe stars, and A stars in general not being at the centre of their own clusters. The analysis in this work does not find a sudden break where after a specific sub spectral type clusters could no longer be found; this work has in fact shown an almost uniform distribution of the fraction of stars found to be in clusters around B-type stars and the A-type stars, and across all their sub spectral types.

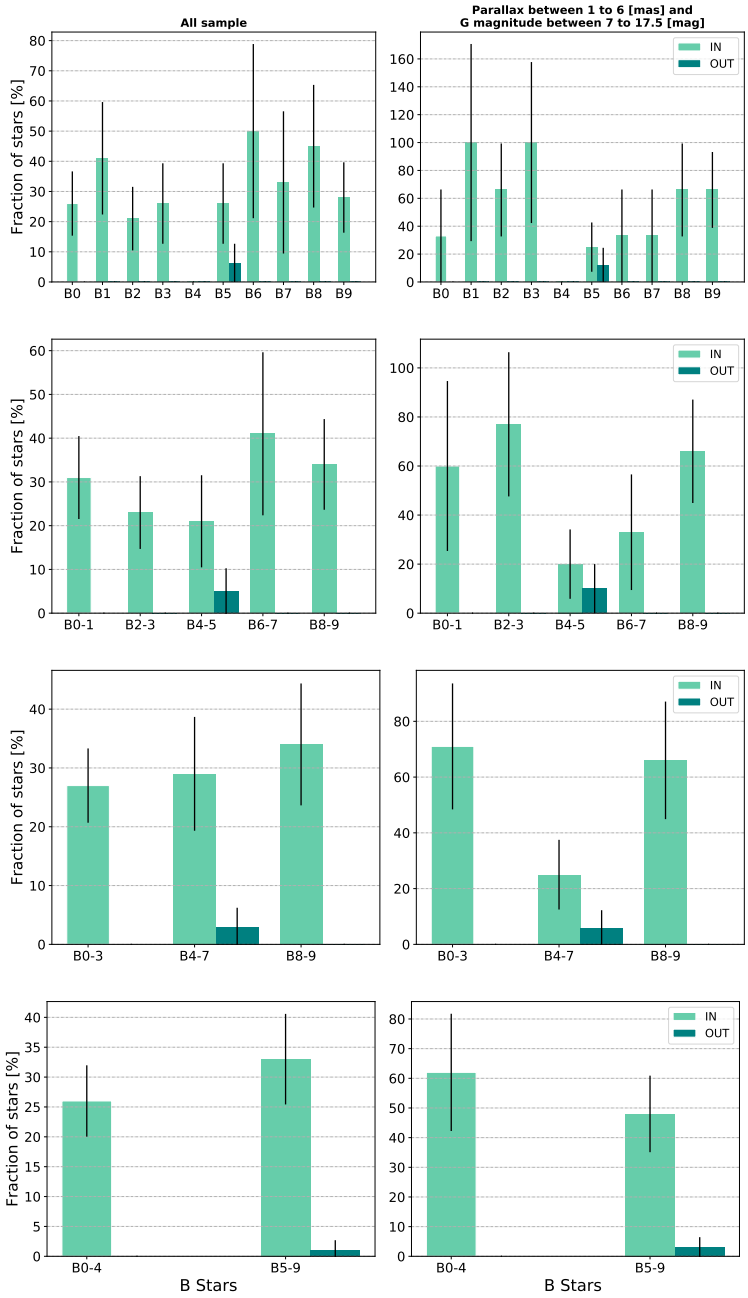


Figure 4.17: Distribution of fractions of stars found to be in clusters, for all the sub-spectral types of the B stars located at the centre of their cluster or not. The light green bars represent the stars found to be in clusters that were located in the centre (IN) and the dark green bars represent the stars in clusters that were not located in the centre (OUT). Each row shows groupings of the sub-spectral types of the B stars. *Left column* shows the sub spectral types of the B stars of the full sample. *Right column* shows the sub spectral types of the B stars that were selected with parallaxes between 1 and 6 mas and G magnitude between 7 and 17.5 mag.

Chapter 5

Cluster formation around B stars: observations versus theory

Be humble because you are made of earth.
Be noble, for you are made of stars.

— Serbian proverb.

It is known that stars, including all massive stars, are born in dense stellar clusters. These clusters form part of giant molecular clouds, that are aggregates of cold and dense molecular clouds distributed across the galaxy (Bonnell *et al.*, 2003; Wright, 2015). The formation of these clusters involves a large number of different processes such as stellar feedback, turbulence, large-scale instabilities, and the agglomeration of smaller clouds. (Lada *et al.*, 1993; Wright, 2015).

The stars formed within a cluster share similar properties, such as their chemical composition and their kinematics. Knowing these properties for stars can be crucial to understanding their formation processes. In addition, knowing their kinematic properties makes it possible to distinguish if the stars have been formed in isolation, or to identify the possible members of the star cluster and background stars.

There have been several studies carried out using different theoretical models of star formation to simulate how stars were formed within stellar clusters.

This chapter will compare the clusters found around HAeBe stars by CEREAL, specifically those B stars located in the centre of their cluster and with densities statistically significantly higher than the background density (3σ level, see appendix B) with theoretical models from the literature. This comparison was carried out with the aim of evaluating whether the clusters around the massive stars, in this sample, appeared to follow any of these models.

5.1 Theoretical models

Studying the formation of massive stars is complicated by the presence of both observational and theoretical challenges. These challenges are due to the nature of these objects, as high-mass stars tend to be embedded in rich clusters, that are located at large distances (Beltrán, 2018), and their evolution happens at a fast enough rate that they reach the zero age main sequence whilst they are still embedded in their molecular clouds and are still undergoing active accretion (Beltrán, 2018). Additional problems are related to the presence of uncertainties in obtaining properties for the cloud in which the massive stars have formed. Further to this, ascertaining whether the observed properties are due to the initial conditions of the cloud, or due to its subsequent evolution is often not possible (Krumholz & Bonnell, 2007).

Several studies have been carried out to analyse and explain the formation of massive stars in clusters. The formation of massive stars can be explained by using two theories: Monolithic collapse (or core accretion; Beltrán, 2018; Krumholz & Bonnell, 2007; Louvet, 2018; Rivilla *et al.*, 2013, references therein) and Competitive Accretion (or Hierarchical; Bonnell *et al.*, 2003; Krumholz & Bonnell, 2007; Louvet, 2018; Rivilla *et al.*, 2013; Vázquez-Semadeni *et al.*, 2019, references therein). These models differ from each other due to the method for the accretion onto the stars (Rivilla *et al.*, 2013).

The Monolithic collapse, or core accretion theory, leads to star formation on a timescale of the order 10^5 yrs, which initiates with a strongly peaked density distribution. This model predicts that the massive stars form from massive cores, which collapse from their natal molecular cloud and produce a single massive object, or only a few, instead of many low mass stars. The newborn star would gather its mass only from the massive core. (Beltrán, 2018; Louvet, 2018; Rivilla *et al.*, 2013, references therein).

The competitive accretion theory predicts that the molecular cloud fragments into low mass cores of the order of their Jeans mass, which compete to accrete un-bound gas from the whole cloud (Beltrán, 2018; Louvet, 2018, references therein). This high fragmentation scenario would then naturally produce a whole cluster of low mass stars. In this model, the massive stars form in a clustered environment which suggests that the massive objects would be born in the densest central part of the low-mass star cluster (Louvet, 2018; Rivilla *et al.*, 2013; Wright, 2015).

5.2 Clusters found with CEREAL versus Theory

The previous chapters have described how for a sample of known HAeBe stars, CEREAL was able to assess the presence of clusters around these targets. For those stars found to be in clusters the more massive, in this case the B stars, tended to be in the centre of their clusters. This in part confirms the results of Testi *et al.* (1997, 1998, 1999), where B stars were more likely to be in the centre of their clusters.

So far, the knowledge obtained from the B stars has raised questions about how these clusters formed. In particular, do these clusters follow the core accretion, or the competitive accretion model? Examples from the literature have suggested that the monolithic model can accurately reproduce the observed properties of the ONC, Pleiades, R136 and NGC 3603 (Banerjee & Kroupa, 2017b, references therein). There is also observational evidence suggesting that the Serpens cluster could be formed following the model of hierarchical fragmentation (Testi *et al.*, 2000).

To understand how the clusters found by CEREAL were formed, a sample of 15 B stars[1], that were found to be in and located in the centre of a cluster and have a density that is statistically significantly higher than the background density (3σ level, see §2.2.1 and appendix B); these stars were selected in order to be compared with the theoretical models of star cluster formation. The sample contains objects that were also analysed by the DCAs as described earlier (§ 3.3).

[1]The star GSC5360-1033 was excluded from this analysis because did not have a reported value for the mass and age (Vioque *et al.*, 2018).

5.2.1 Cluster size

The cluster radii determined in section §2.2.1 and appendix B will be used in the analysis to follow. The cluster sizes used in this work are larger than those used in the previous studies (Hillenbrand *et al.*, 1995; Testi *et al.*, 1997, 1998, 1999, 0.15 pc and 0.2 pc, respectively). As discussed previously the use of larger cluster sizes will lead to lower stellar densities being found than if a smaller cluster radius was used.

5.2.2 Minimum Mass

The next step for the comparison with the theoretical models was to calculate the stellar density of the clusters. Before doing so, an estimate for the lower mass limit detected for each cluster was required, to do this the masses of the objects within the cluster needed to be calculated. To find the mass of the objects within the cluster, a similar method to the one used by Oudmaijer & Parr (2010) was applied in this analysis. In that work the mass ratio was calculated by using the K magnitude difference of their sample; which did not find any dependence on the spectral types within their sample of stars. Oudmaijer & Parr (2010) found a linear relationship when plotting the logarithm of the mass of the stars against their K magnitude.

Following a similar method, the lower mass limit detected for each cluster was found by plotting the logarithm of the masses of the full sample of HAeBe stars (with has masses reported by Vioque *et al.*, 2018, see §3.2) and their absolute magnitude.

Figure 5.1 shows the relation between the logarithm of the mass and absolute magnitude of the HAeBe stars[1]; despite the scatter of the data, the stars show a linear relation that can be represented as $log(M) = m * M_G + c$, with M in solar units, m and c as -0.131 ± 0.002 and 0.532 ± 0.006, respectively. This relationship translates to a mass ratio as $\frac{M}{M_{HAeBe}} = 10^{-0.131 \Delta M_G}$, where ΔM_G is the magnitude difference between the object in the cluster and the Herbig Ae/Be star. All the possible cluster members found around the HAeBe stars have a fainter magnitude compared to the HAeBe stars. In this analysis, extinction corrections (see appendix H) were not applied to the possible cluster members, because the value of the A_G of the cluster member is unknown.

[1]The absolute magnitude of the HAeBe stars were corrected by A_G using the extinction relation found by Wang & Chen (2019, $A_G = 0.789 A_V$). The A_V values for the Herbig stars were taken from Vioque *et al.* (2018).

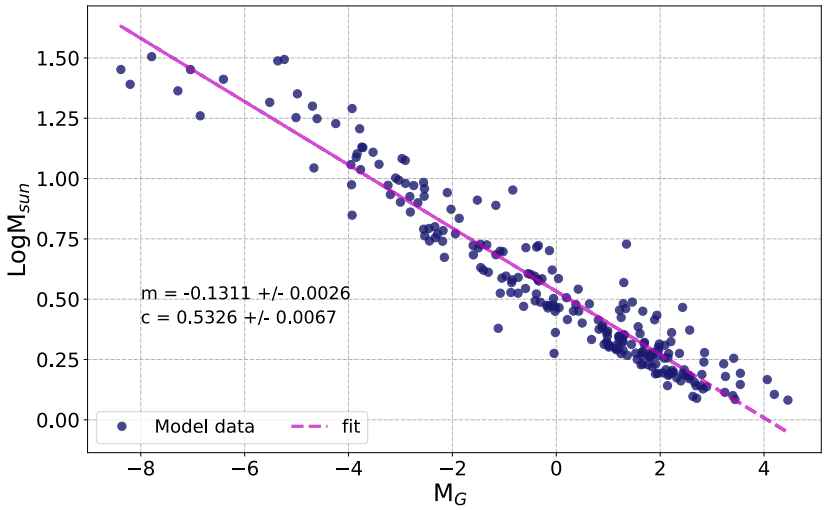


Figure 5.1: Mass ratio relation for HAeBe stars. The full blue circles represent the masses of the full sample of HAeBe stars (see §3.2) and their absolute magnitudes corrected for extinction. The data appears to be linearly distributed and first order fit of the form $log(M) = m * x + c$ (magenta dashed line) is shown.

It is expected that the extinction of the HAeBe stars should be larger than for their low mass companions due to the inclusion of circumstellar extinction. Therefore, the decision was taken to calculate ΔM_G using the extinction corrected absolute magnitudes of the HAeBe stars and the uncorrected absolute magnitudes of their low mass companions. The correction of the absolute magnitudes of the HAeBe stars and not their low mass companions will have led to overestimating ΔM_G. Although the resultant ΔM_G is likely closer to the true ΔM_G than if the absolute magnitudes of the HAeBe stars were not corrected for extinction. The case where neither the absolute magnitudes of the HAeBe stars and their low mass companions are corrected for extinction is provided in appendix H, the true value of ΔM_G should lie between these limits.

Figure 5.2 provides an example calculation of the lower mass limit and shows the fainter stars found in the cluster around the HAeBe star. These values were estimated using the mass ratio relation together with the data obtained from the clustering analysis made by CEREAL.

The result obtained from this analysis was that, for each cluster it was possible to extract the minimum mass observed within them. This value varied from target to target

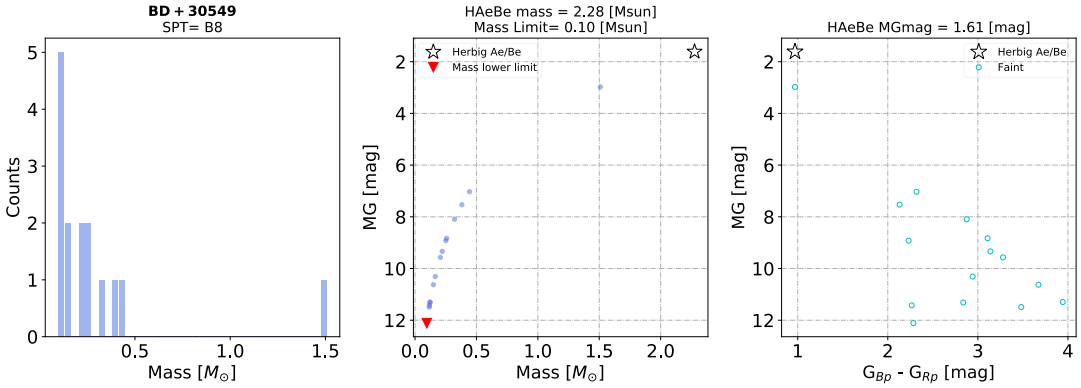


Figure 5.2: Lower mass and magnitude limits for a HAeBe star calculated using the Mass ratio relation. *Left panel:* represents the mass distribution of the low mass companions found with CEREAL around the HAeBe star BD+30549. *Middle panel:* shows, in full blue circles, the relation between the mass of the low mass companions and their G magnitude. The position of the lower mass detected in this cluster by CEREAL, is also indicated by the inverse red triangle. *Right panel:* shows a colour-magnitude distribution for the low mass companions found around the HAeBe star BD+30549 with a cluster size of 0.7 ± 0.2pc (see section §5.2.1).

since it was related to the number of stars found in each cluster and their magnitudes. The lowest and largest minimum masses found here were $\sim 0.1 M_\odot$ and $\sim 0.7\ M_\odot$, respectively; where the mean minimum mass observed was $\sim 0.2\ M_\odot$.

5.2.3 Completeness correction

An analysis of the completeness of the sample was made to take into account the number of stars that might have been missed for each cluster as function of magnitude. For this analysis the photometric data provided by the Gaia DR2 catalogue was used, specifically the G magnitude.

It is known that the Gaia DR2 catalogue[1] contains sources with G magnitudes limited to between $\approx$ 3 and 21. The Gaia DR2 catalogue is essentially complete between G magnitudes 12 to 18 but is still incomplete at the bright end and has an ill-defined faint magnitude limit which depends on celestial positions (Arenou *et al.*, 2018). In dense areas of the sky, the magnitude limit of the Gaia DR2 catalogue is 18 mag (Gaia Collaboration *et al.*, 2018b). This completeness represents a huge improvement on the values shown for

[1]Visit Gaia Data Release 2 site for more information what the survey contents.

the first Gaia data release but, there is still some uncertainty in the full completeness of the survey.

As is mentioned by Gaia Collaboration *et al.* (2018b) and Arenou *et al.* (2018), the completeness of Gaia DR2 is affected by each individual data process used and the overall validation of the data. Arenou *et al.* (2018) also identified that the completeness was affected by a variety of truncations and filters applied to the different types of data[1]. This suggests there is not a direct way to know the completeness of the clusters found around the HAeBe stars using the Gaia DR2 data. However, to obtain a reasonable understanding about the completeness of the Gaia DR2 catalogue the completeness was inferred from the work included in Boubert & Everall (2020)[2].

Boubert & Everall (2020) predict the completeness of Gaia DR2 by exploiting the fact that it only contains sources with at least five astrometric detections and predict the number of times that each source was observed by Gaia and assume that the probability of detection is a function of magnitude. They found that Gaia DR2 is mainly complete over the range 7 $<$ G magnitude $<$ 20, but that the completeness falls to 0% over the range 20 $<$ G magnitude $<$ 21.3. They also found that, as predicted by Gaia Collaboration *et al.* (2018b) and Arenou *et al.* (2018), Gaia DR2 cannot see the extreme bright (G magnitude $<$ 3) or faint objects (detection falls from 100% to 0% between 20 $<$ G magnitude $<$ 21.5) stars. Figure 5.3 shows the figure 5 from Boubert & Everall (2020) which was used to estimate the completeness of the sample of clusters around the HAeBe stars. Additionally, the effect of applying a filter to the sample was also considered. In this case, the filter used was RUWE $<$1.40 (see §2.1.1 which was applied over the raw Gaia DR2 data before the cluster assessment). The impact of this filter is shown in left panel of figure 5.4 where the number of stars found at each G magnitude before and after RUWE was applied are given. In the right panel of the figure the fraction of stars that remained at each G magnitude after RUWE was applied is also shown, these were then used to correct for the influence this selection had on the number of stars obtained.

The results obtained by Boubert & Everall (2020) show that Gaia DR2 is $\sim$99% complete between 7 $<$ G magnitude $<$ 20 which includes all the stars found in the clusters around the HAeBe stars. The G magnitude of the minimum mass objects observed here

[1] Gaia Collaboration *et al.* (2018b), Arenou *et al.* (2018) and the Gaia DR2 documentation provide more information about the completeness.

[2] See Completeness of the Gaia-verse site for more information.

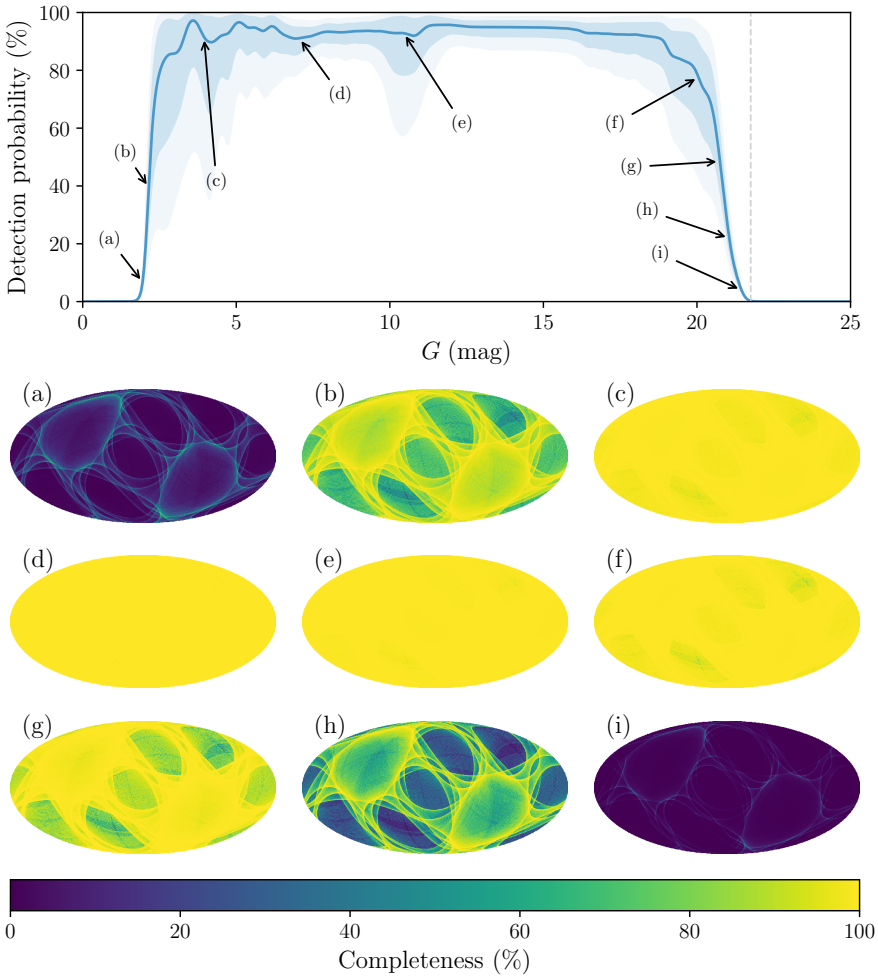


Figure 5.3: Distribution of probabilities that a source is detected by Gaia as a function of magnitude. *Top panel* shows the detection probability used to calculate the completeness as a function of magnitude. *Bottom panel* shows how the completeness changes at each point shown in the detection probability plot on the left and the colour bar represents the completeness observed. Figure taken from Boubert & Everall (2020).

in clusters, varied from 17.21 mag to 20.26 mag; where the mean G magnitude observed for the minimum mass objects in the clusters was $\sim$ 18.60 mag. The mean minimum mass G magnitude found here is in agreement with the expected upper limit for G magnitude observable due to photometric precision, suggested by Arenou *et al.* (2018) when using clusters with little extinction (G$\sim$18).

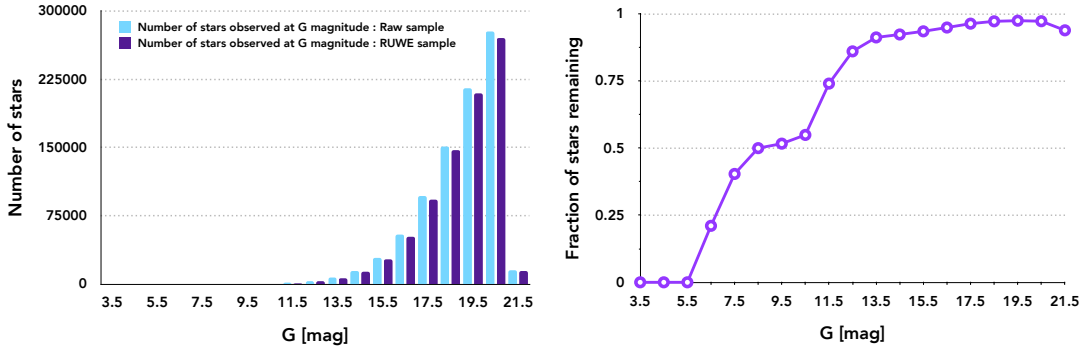


Figure 5.4: The distribution of the probability that a source detected by Gaia was removed by the cut in RUWE as a function of magnitude. *Left panel* represents the distribution of the number of stars in the sample at each G magnitude before and after the cut for RUWE <1.40 was applied. *Right panel* shows the fraction of the stars remaining at each G magnitude afterRUWE was applied.

5.2.4 Stellar density

A calculation of the stellar density was made for the sample of 15 B stars[1] found to be in clusters, that were also located in the centre of their clusters and were corrected for their completeness as function of the magnitude as described previously. The results from this calculation are presented in figure 5.5, which shows the relationship between the age and the stellar densities of the clusters found around the B stars. The stellar densities were calculated using the number of stars found in each cluster with the clusters considered to be spheres. The volume of each cluster was obtained by using the cluster size (§ 5.2.1), estimated before, as the radius of the sphere.

Figure 5.5 compares the stellar densities of the clusters found around the HAeBe stars, that were classified as yes and maybe, with the age of the HAeBe star (§ 4.4 discusses the reasoning behind treating stars classified as maybe as stars found to be in clusters). The ages of the stars were taken from the catalogue of Vioque *et al.* (2018). There is no clear trend in the stellar densities with age for the stars classified as maybe in the top panel with the densities of these falling between 4.5 and 3.8x10^1 stars/pc^3 (the average stellar density across the whole range being 1.4x10^1 stars/pc^3). However, in the lower panel where the rolling average is presented a general decrease in the density with age is apparent falling from >2.0x10^1 stars/pc^3 for ages younger than below 1x10^5 yrs

[1]The star GSC5360-1033 was excluded from this analysis because it did not have a reported value for the mass and age (Vioque *et al.*, 2018).

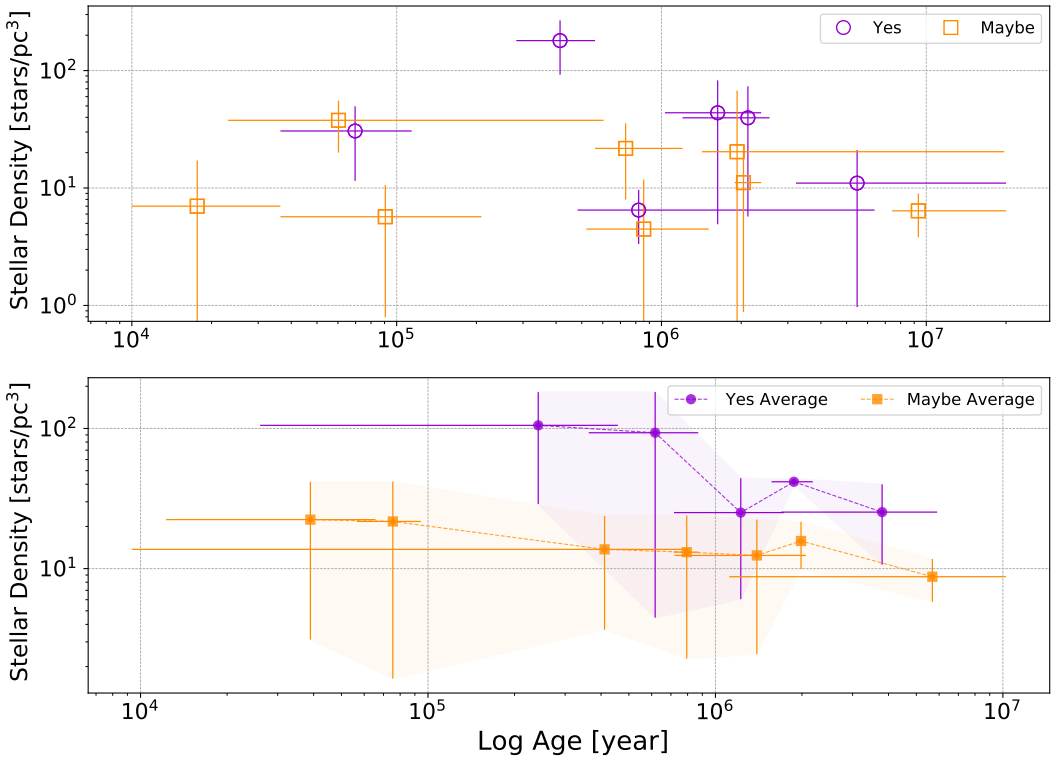


Figure 5.5: Stellar density vs Age. The top panel compares the stellar densities of the B stars found to have a density significantly higher than the background density (3σ level) and have been classified by CEREAL as yes (purple open circles) and maybe (orange open squares). The bottom panel represents the rolling average (for every 2 points) of the B stars shown in the top panel. The purple and orange shadows represents the error of the rolling average.

to $<$1.0x10^1 stars/pc^3 at 5x10^6 yrs. This general trend is mainly based on the density around AS310 leading to the first two points of the rolling average being above the average across the whole range. The trend for the stellar density to decrease with ages is more apparent for the stars classified as yes in the top panel. Although, here there is also an initial increase in stellar density from 3.1x10^1 stars/pc^3 at 7x10^4 yrs to 1.8x10^2 stars/pc^3 at 4x10^5 yrs; followed by, the general trend of stellar density to decrease with age falling to 1.1x10^1 stars/pc^3 at 5x10^6 yrs. As was seen for the stars classified as maybe this general trend is mainly based around a single cluster around V361Cep having a stellar density far above the average density across the whole range of 5.1x10^1 stars/pc^3.

When looking at the lower panel the rolling average of the stellar densities with age the influence of V361Cep leads to two regions one at higher stellar densities below 1x10^6 yrs and a second region with lower stellar densities for clusters around stars older than 1x10^6 yrs, with both of these regions being flat. The apparent difference in the trends of the stars classified as yes and those classified as maybe based on their rolling averages is mainly based around the density of the cluster found around V361Cep. This can be seen in the top panel of the figure where there is a high degree of overlap between the stellar densities of the clusters around stars classified as yes and maybe. It is notable that any apparent difference in trends between stars classified as yes and maybe is mainly based on the stellar densities calculated around 3 stars V361Cep, MWC137, HD46060; if MWC137 and V361Cep were not included in this analysis then the differences in trends at young ages ($<$1x10^6 yrs) would no longer be present. However, a difference in average densities between the clusters found as yes and maybe would be present. Additionally, it should be noted that all of the clusters observed here did not appear to have high densities (3.0x10^1 stars/pc^3) with an average number of cluster members of only 18 stars.

The stellar densities of each cluster, presented in figure 5.5, do not show a complete representation of all the possible members of these clusters, due to each cluster having a limit on the mass below which smaller objects can not be observed. This effect could lead to the low stellar densities observed for the clusters here, and this effect therefore needed to be accounted for, beyond the corrections made before based on the completeness carried out in section §5.2.3.

5.2.5 Initial Mass Function (IMF)

It can be assumed that from a given molecular cloud several stars will be born, at the same time, with different masses (Prialnik, 2000). The cluster sample analysis carried out in this study however had a lower limit under which masses could not be observed within them. To correct for these missing small objects, it was necessary to use the initial mass function of the clusters (IMF; Salpeter, 1955) to estimate the number of cluster members that could not be observed (due to these objects being under the minimum mass limit), to obtain a more representative value of the stellar density of the clusters members around the HAeBe stars.

The IMF theoretically predicts the number of stars with a given mass to be born in a molecular cloud (Salaris & Cassisi, 2005). Salpeter (1955) determined an empirical expression which fitted the available observational data, at the time, with a power law between $0.4 M_{\odot}$ and $10\ M_{\odot}$ (Ward-Thompson & Whitworth, 2011). This equation is represented by $\frac{d_n}{d_M} = CM^{-x}$, where d_n is the number of stars born with masses between M and $M + d_M$, $x = 2.35$ and C is a constant.

For this analysis, the IMF was obtained using the scipy package from python[1] by integrating the Salpeter (1955) IMF between 0.07 $M_{\odot}$ (Bonnell *et al.*, 2003) to the mass of the HAeBe stars (Vioque *et al.*, 2018).

To find the ratio of objects over the minimum mass limit as a fraction of the total number of cluster member using the IMF, two integrals needed to be calculated. The first one ($INT1$) used the minimum mass found in each cluster by the mass-magnitude relation as a lower limit and used the mass of the HAeBe star, taken from Vioque *et al.* (2018), as the upper limit, as these were likely to be the most massive objects in the cluster. The second integral ($INT2$) used the whole mass range up to that of the HAeBe stars, using 0.07 $M_{\odot}$ as the lower limit, as this was the minimum stellar mass computed by Bonnell *et al.* (2003).

The IMF calculations predicted how many stars, of each mass, were formed in the clusters. The integrals gave the total population (INT2) and the population of the observable range of masses (INT1), these were then used to calculate an IMF ratio of observable objects to those under the minimum mass for the cluster (Ratio=$\frac{INT2}{INT1}$), which could then be used to estimate the "real" (or complete) stellar density of each cluster (Stellar Density * Ratio = Corrected Stellar Density), for the subsequent cluster analysis performed on this sample. A representation of the two integrals of the IMF that were calculated is shown in figure 5.6.

5.2.6 Corrected stellar densities

Figure 5.7 shows the relationship between the age and the "real" stellar densities of the clusters found around the B stars. The stellar density was calculated by multiplying the IMF ratio by the stellar densities calculated previously (which were presented in figure 5.5).

[1] Using General integration (quad) from scipy package and the tutorial by Zach Pace, Lia Corrales, Stephanie T. Douglas on to use scipy.integrate

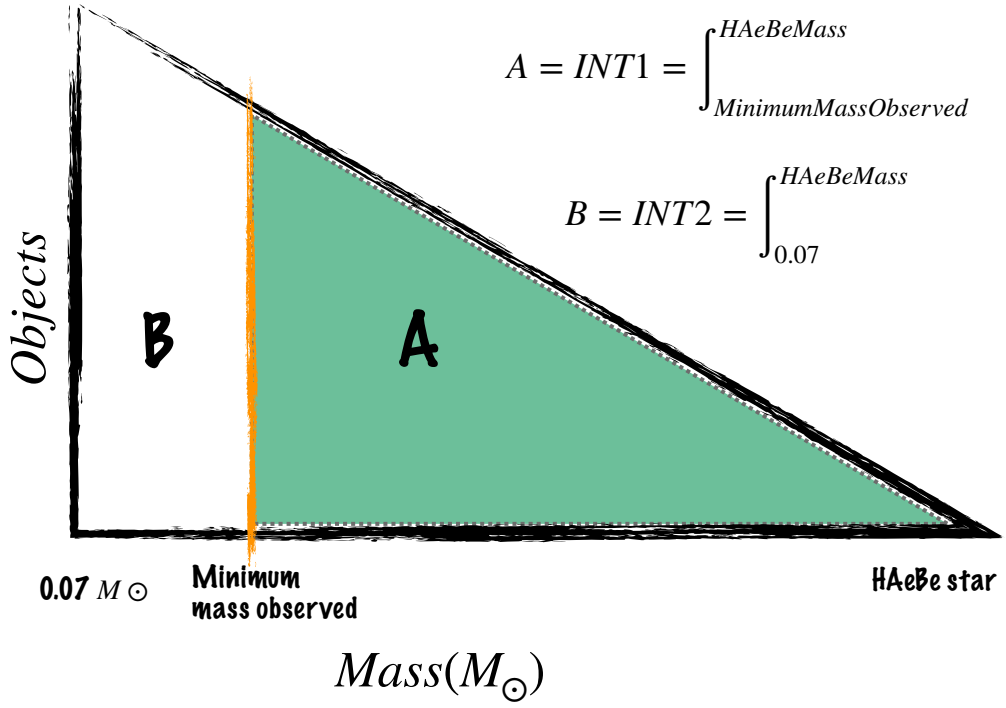


Figure 5.6: A schematic representation of the estimation of the IMF ratio, which is calculated from the the population of stars that are predicted to fall within the observable range of masses (A) and the total population of stars that are found between the minimum stellar mass of $0.07M_\odot$ and the mass of the HAeBe star (B), which are calculated in INT1 and INT2 respectively.

Figure 5.7 shows a more accurate representation of the stellar densities of the clusters found around the HAeBe stars, that were classified as yes and maybe, as a function of the age of the HAeBe stars. In this figure there has been an increase in the stellar densities, due to the completeness in mass obtained from the IMF for each cluster, when compared with figure 5.5; the mean stellar density now observed is 194 stars/pc^3, which is much larger than the mean stellar density detected earlier of 30 stars/pc^3 without the IMF corrections. As was seen for the results discussed previously prior to the IMF correction, the distributions of the stars in clusters classified as both *yes* and *maybe* with age appear very similar with a high degree of overlap in the top panel. When the rolling averages in the lower panel are considered both show a trend from higher densities over $4x10^2$ stars/pc^3 at earlier times to lower densities less than $8x10^1$ stars/pc^3 at later ages. There

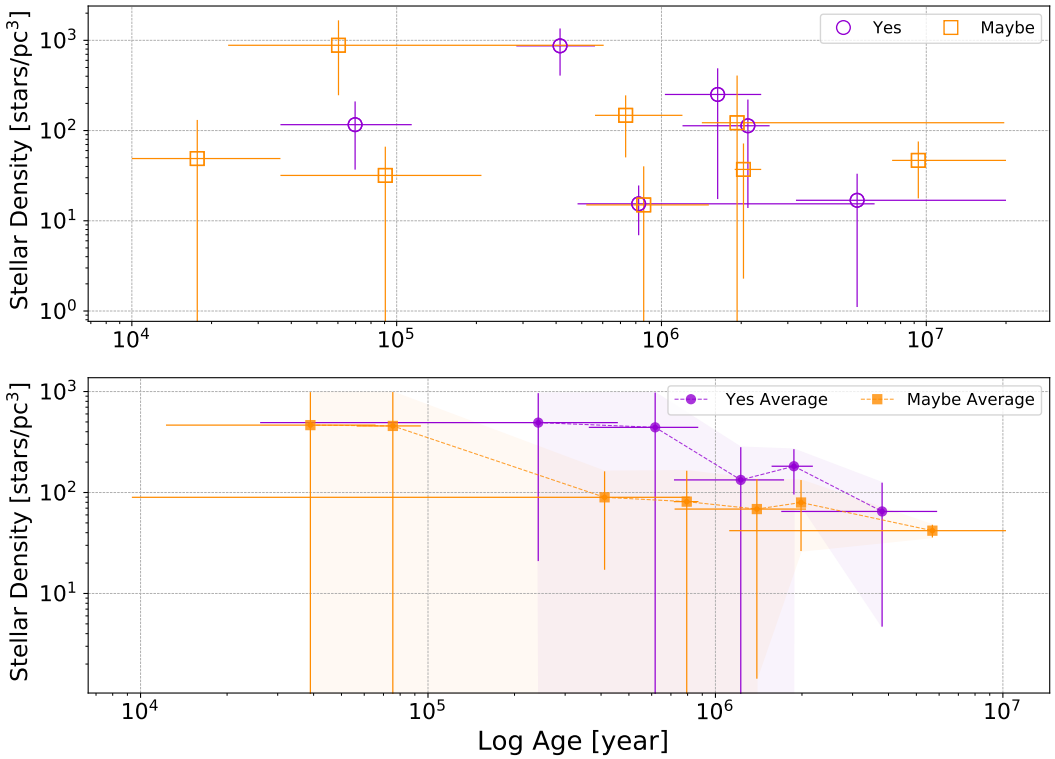


Figure 5.7: Real Stellar density vs Age. The "real" stellar density was calculated using the value obtained with the IMF correction for the number of objects below the minimum observable mass for the clusters analysed in this sample. The figure has the same description as figure 5.5.

is however, a difference in when this transition occurs with the rolling average of the stellar densities for stars classified as maybe beginning to fall earlier (stars $>1x10^5$ yrs) than for those classified as yes ($>7x10^5$ yrs).

That both the stars classified as *yes* and *maybe* have the same trend, higher densities for the earliest ($<5x10^5$ yrs), followed by a trend down in stellar density (between $4x10^5$ to $1x10^6$ yrs) and then a relatively stable behaviour for ages over $1x10^6$ yrs is not unexpected. The similarity of both the *yes* and *maybe* samples additionally allows for the potential grouping of both these data sets. It is important to note, that the differences between the classification of targets as either *yes* or *maybe* were very subtle (§3.2.2), which meant that the stars classified as *maybe* could also be considered to be *yes*. The combination of these two samples reduces the statistical errors in the average stellar densities with

age providing a better sample to compare with the results of the theoretical models. The results obtained in the previous sections (§5.2.2, §5.2.5, §5.2.4 and §5.2.6) have been summarised in table 5.1.

Table 5.1: Summary of the cluster parameters

Star	Spectral type	Distance	Mass	Age	Cluster Radius	Num Stars	Low Mass Limit	IMF ratio	Stellar Density	Complete Stellar Density
		pc	$M_\odot$	Myr	pc		$M_\odot$		stars pc^{-3}	stars pc^{-3}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
A0974-15	B3	$397.6^{+9.6}_{-8.9}$	$2.71^{+0.36}_{-0.23}$	$2.0^{+18.0}_{-1.0}$	0.34 ± 0.26	4 ± 2	$0.257^{+0.036}_{-0.025}$	$6.0^{+1.9}_{-1.6}$	20.0 ± 47.0	$120.0^{+280.0}_{-280.0}$
AS310	B1	$2110.0^{+360.0}_{-240.0}$	$11.9^{+4.8}_{-3.4}$	$0.06^{+0.55}_{-0.04}$	0.387 ± 0.043	9 ± 3	$0.71^{+0.29}_{-0.21}$	$23.0^{+18.0}_{-13.0}$	38.0 ± 18.0	$880.0^{+780.0}_{-640.0}$
BD+30549	B8	$295.0^{+13.0}_{-11.0}$	$2.28^{+0.37}_{-0.19}$	$5.0^{+15.0}_{-2.0}$	0.73 ± 0.21	18 ± 4	$0.096^{+0.017}_{-0.01}$	$1.53^{+0.51}_{-0.31}$	11.0 ± 10.0	$17.0^{+16.0}_{-16.0}$
HD36982	B1.5	$408.0^{+19.0}_{-16.0}$	$5.2^{+0.42}_{-0.29}$	$0.73^{+0.47}_{-0.17}$	0.64 ± 0.13	24 ± 5	$0.285^{+0.028}_{-0.023}$	$6.8^{+1.4}_{-1.2}$	22.0 ± 14.0	$147.0^{+98.0}_{-97.0}$
HD37371	B9	$411.0^{+19.0}_{-16.0}$	$3.85^{+0.63}_{-0.67}$	$0.86^{+0.65}_{-0.34}$	0.67 ± 0.36	6 ± 2	$0.17^{+0.03}_{-0.032}$	$3.4^{+1.1}_{-1.2}$	4.5 ± 7.3	$15.0^{+25.0}_{-25.0}$
HD46060	B8	$933.0^{+96.0}_{-71.0}$	$9.6^{+3.4}_{-2.4}$	$0.09^{+0.12}_{-0.05}$	1.07 ± 0.3	29 ± 5	$0.249^{+0.089}_{-0.065}$	$5.6^{+3.7}_{-2.7}$	5.7 ± 4.9	$32.0^{+34.0}_{-31.0}$
HUCMa	B8	$1174.2^{+100.0}_{-77.0}$	$3.02^{+0.15}_{-0.15}$	$2.04^{+0.34}_{-0.15}$	0.81 ± 0.24	25 ± 5	$0.169^{+0.013}_{-0.013}$	$3.34^{+0.6}_{-0.6}$	11.0 ± 10.0	$37.0^{+35.0}_{-35.0}$
Hen3-949	B3	$643.0^{+33.0}_{-28.0}$	$4.18^{+0.74}_{-0.52}$	$0.8^{+5.6}_{-0.3}$	0.9 ± 0.13	20 ± 4	$0.132^{+0.025}_{-0.019}$	$2.38^{+0.83}_{-0.62}$	6.5 ± 3.1	$15.4^{+9.2}_{-8.5}$
ILCep	B2	$805.0^{+31.0}_{-27.0}$	$9.9^{+2.7}_{-1.3}$	$0.07^{+0.044}_{-0.033}$	0.63 ± 0.13	32 ± 6	$0.188^{+0.053}_{-0.028}$	$3.8^{+2.0}_{-1.1}$	31.0 ± 19.0	$116.0^{+94.0}_{-79.0}$
LKHa338	B9	$885.0^{+63.0}_{-50.0}$	$1.885^{+0.094}_{-0.094}$	$9.0^{+11.0}_{-2.0}$	1.05 ± 0.13	31 ± 6	$0.29^{+0.018}_{-0.018}$	$7.3^{+3.4}_{-3.4}$	6.4 ± 2.6	$47.0^{+29.0}_{-29.0}$
LkHa324	B9	$605.0^{+16.0}_{-14.0}$	$2.82^{+0.61}_{-0.2}$	$2.12^{+0.44}_{-0.92}$	0.47 ± 0.13	17 ± 4	$0.151^{+0.033}_{-0.014}$	$2.9^{+1.2}_{-0.6}$	40.0 ± 34.0	$110.0^{+110.0}_{-100.0}$
MWC137	B0	$2910.0^{+600.0}_{-400.0}$	$23.1^{+10.6}_{-6.5}$	$0.018^{+0.019}_{-0.008}$	0.75 ± 0.36	12 ± 4	$0.29^{+0.14}_{-0.09}$	$7.0^{+5.9}_{-3.8}$	7.0 ± 10.0	$49.0^{+82.0}_{-76.0}$
V361Cep	B2	$893.0^{+35.0}_{-31.0}$	$5.31^{+0.69}_{-0.48}$	$0.41^{+0.15}_{-0.13}$	0.288 ± 0.041	18 ± 4	$0.223^{+0.032}_{-0.025}$	$4.8^{+1.3}_{-1.0}$	180.0 ± 88.0	$870.0^{+480.0}_{-460.0}$
V373Cep	B5	$922.0^{+33.0}_{-29.0}$	$3.18^{+0.51}_{-0.39}$	$1.63^{+0.75}_{-0.6}$	0.291 ± 0.073	5 ± 2	$0.251^{+0.042}_{-0.033}$	$5.7^{+1.9}_{-1.6}$	44.0 ± 39.0	$250.0^{+240.0}_{-230.0}$

Columns (3), (4) and (5) taking from Vioque *et al.* (2018)

Columns (7): Number of stars found using column (6) (§5.2.1) and Gaia DR2 completeness (§5.2.3).

Columns (8): Low mass limit detected for each cluster (§5.2.2).

Columns (9): Initial mass function ratio of each cluster (§5.2.5).

Columns (10): Stellar density for each cluster (§5.2.4).

Columns (11): Stellar density for each cluster corrected by their IMF (§5.2.6).

5.2.7 Comparison between the observations and the theory

The measurement of the IMF provided a more accurate representation of the stellar densities of the clusters found around the HAeBe stars (see figure 5.7). The increase in the stellar densities revealed the large number of missing stars that were not detected during the cluster assessment with CEREAL (CEREAL observed under 16% of the total cluster members calculated once the IMF corrections were made). As was mentioned previously in section §5.2.5, the IMF correction utilised in this analysis was the Salpeter (1955) IMF. This was chosen along with the minimum mass (see figure 5.6) to allow for direct comparison with the modelling results of Bonnell *et al.* (2003), as this was the IMF and minimum mass utilised in their work. The calculations made in section 5.2 along with the results of section 3.2 and the main parameters of the HAeBe stars, taken from Vioque *et al.* (2018), will now be used to compare the clusters found around the B-type star with theoretical models of stellar formation.

The numerical simulations with which the results of this study are being compared, were carried out by Bonnell *et al.* (2003). Their simulations followed the fragmentation of a turbulent molecular cloud, and the subsequent formation and early evolution of a stellar cluster containing more than 400 stars. Their model used a high-resolution smoothed particle hydrodynamics (SPH; Bonnell *et al.*, 2003, references therein) simulation to follow the fragmentation of an initially uniform-density molecular cloud, containing 1000 $M_{\odot}$ in a diameter of 1 pc and a gas temperature of 10 K. The results obtained from these simulations were that the clusters formed through a hierarchical process of several sub-clusters, which grew and merged through the subsequent dynamics to form the much larger final cluster. Figure 5.8 shows different snapshots of the stellar cluster formed through hierarchical fragmentation, from the simulation made by Bonnell *et al.* (2003).

Figure 5.9 shows the evolution of the local (hierarchical model) and global (monolithic model) stellar number-densities for the cluster as function of the time estimated by Bonnell *et al.* (2003). The local stellar density was quantified from the minimum volume required to contain the ten nearest neighbours of the stars. In contrast, the global stellar number-density was calculated from the volume required to contain half of the total number of stars. In these simulations the initial free-fall time for the cloud was 1.9×10^5 yr.

Bonnell *et al.* (2003) found that the local number-densities increased rapidly (hierarchical model), once the first stars formed and fell towards each other in their local sub-cluster, where the stars occupied only a small fraction of the total volume of the star-forming region (monolithic model). Additionally, Bonnell *et al.* (2003) found that the local number-density then decreased after reaching a maximum during the sub-clustering phase. This decrease was due to the ejection of stars from the sub-clusters through dynamical interactions, and due to the kinetic heating during the merging of sub-clusters, which produced a single centrally condensed cluster (Bonnell *et al.*, 2003).

Prior to the comparison with the theoretical models it has been demonstrated that, for the sample of stars found to be in clusters by CEREAL, the B stars can be considered to be the most massive stars within their clusters, and that the B stars are located in the centre of their clusters, like the HAeBe star HD200775, which is located in the Cep Flare molecular cloud (Zucker *et al.*, 2020). This meant, that the B stars could have been formed by the monolithic collapse scenario. Although, it should be noted that the existence of truly isolated massive stars is one of the caveats of this model (Beltrán, 2018, references therein).

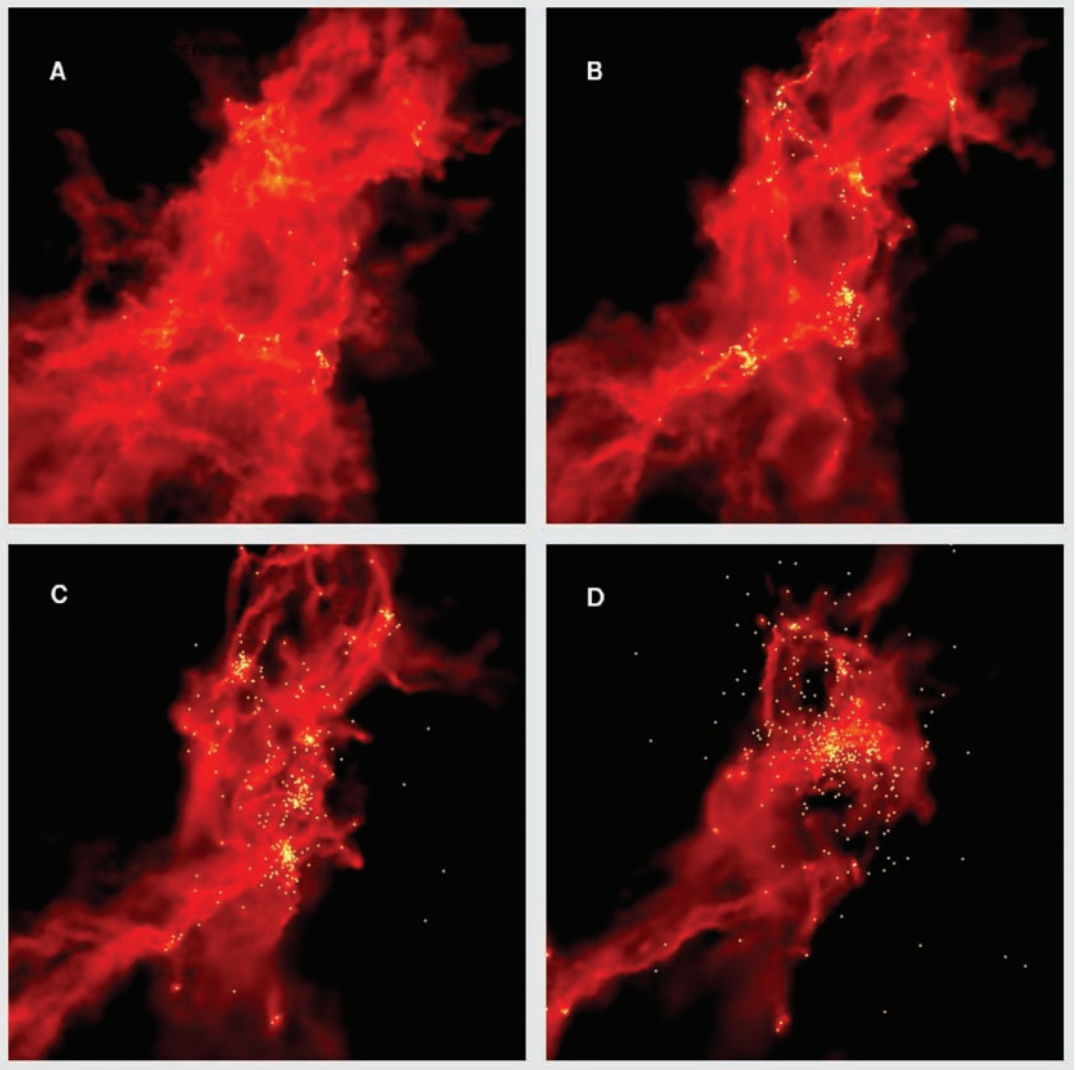


Figure 5.8: Stellar cluster formation through the hierarchical fragmentation of a turbulent molecular cloud. The white dots represent the stars and each panel shows a different stage of the cluster formation, where the initial free-fall time for the cloud was 1.9×10^5 yr. *Panel A:* The turbulence causes shocks to form in the molecular cloud, dissipating kinetic energy and producing filamentary structures, which fragment to form dense cores and individual stars. *Panel B:* formation of subclusters. *Panel C:* the subclusters evolve by accreting more stars and gas, ejecting stars, and by merging with other subclusters; and, *Panel D:* shows a fully formed centrally condensed cluster with small substructures. Image taken from Bonnell *et al.* (2003)

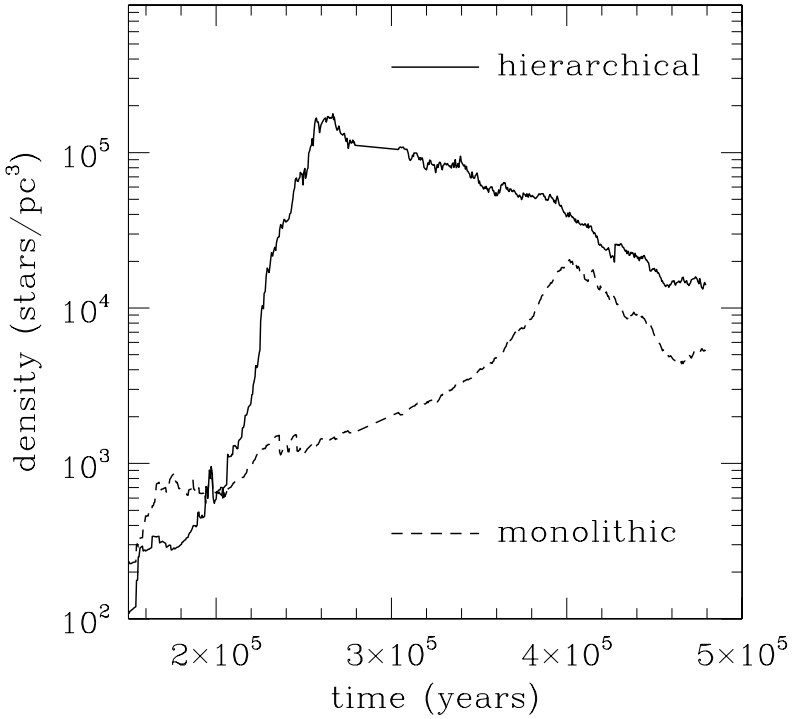


Figure 5.9: Representation of the theoretical models. This figure shows how the local and global stellar densities evolve as as a function of time, taking intro consideration the initial free-fall time of the cloud (1.9x10^5 yr). The local density is compared to the hierarchical model (solid line) and the global density to the monolithic model (dashed line). The formation of sub-clusters can explain the rapid rise of the stellar density in the hierarchical model compared to monolithic model. The difference between the models decreases with time when the substructures are erased to produce a single cluster. Figure taken from Bonnell *et al.* (2003).

However, the results obtained from CEREAL for the B stars showed that these stars are surrounded by several low mass companions that share similar properties (parallax and proper motions) with the HAeBe star. This phenomenon could be explained by the hierarchical model which relies on the fact that nearly all massive stars are formed in stellar clusters (Rivilla *et al.*, 2013). Another way to explain the hierarchical model is by the result obtained by Bonnell *et al.* (2003); in their simulations the cluster was formed by the fragmentation of the cloud into several sub-clusters which could then grow and merge to form a final large cluster. This can be observed in the clusters found by CEREAL (§3.2.4) around the HAeBe stars V590Mon, HBC217 and HBC222, which are part of the

cluster NGC2264 that is itself embedded in the Mon OB1 cloud complex (Buckner *et al.*, 2020; Peretto *et al.*, 2006). Additional observational evidence for the hierarchical model is shown by Testi *et al.* (2000) in their study of the Serpens cluster-forming core. They showed that the star formation in the cloud appeared localised in separate sub-clusters, which then have a stellar density much higher than the mean for the entire Serpens cluster (Testi *et al.*, 2000), this was also observed by Wright & Parker (2019) in their study of the NGC6530. They showed that NGC6530 provides the first dynamical evidence that star clusters can form by hierarchical mergers between sub-clusters due to it containing many massive stars that formed in a pre-collapse sub-structured distribution or could possibly have accreted additional mass during the collapse process.

For comparison of the results obtained here and the model results, figure 5.10 represents the "real" stellar density as a function of the age of the B stars that were found to be in a cluster by CEREAL. This result is the same as the one shown in figure 5.7 but now shows the trend for all B stars, that were classified as yes and maybe. Figure 5.10 also includes a highlighted area which represents the same time window, between 1.6x10^5 to 5x10^5 yrs, that is depicted in figure 5.9 for the theoretical models.

As could be seen in figure 5.7 it is again apparent that there appear to be two regions

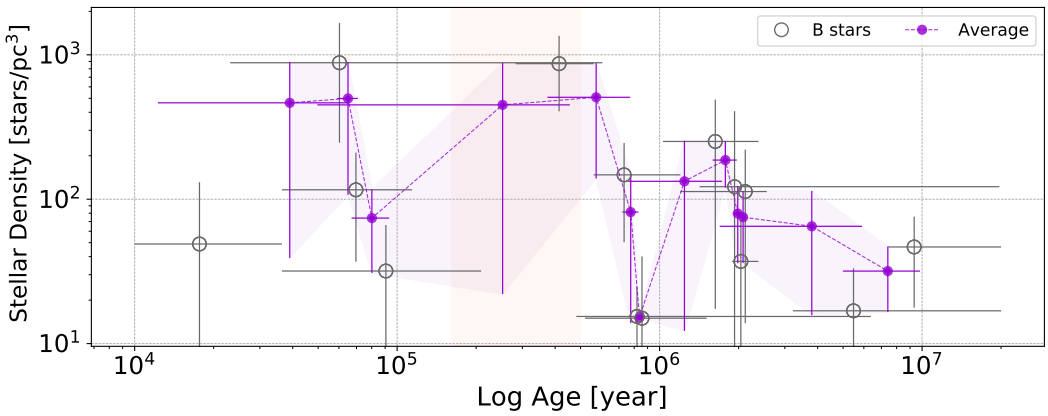


Figure 5.10: Real Stellar densities vs Age. The open grey circles represent the B stars found to be in a cluster by CEREAL and the purple filled circles represent the median rolling average (for every 2 points) of those stars. The purple shadows represent the median error of the rolling average. The light orange area represents the time scale depicted in figure 5.9.

of differing behaviour within the age range observed; at early times (<7x10^5 yrs) there are clusters that have higher stellar densities (generally between 1.0x10^2 stars/pc^3 and 1.0x10^3 stars/pc^3) than those at older ages (>7x10^5 yrs) that appear to generally have stellar densities between 1x10^1 and 2x10^2 stars/pc^3). For the clusters around stars younger than 7x10^5 yrs there is no clear dependence of stellar density with age with the density rising then falling then rising again both when considering the rolling average and the individual clusters. This can also be seen for the stars older than 7x10^5 yrs where there is again no clear trend of stellar density with age.

The main difference between the stellar densities for stars at older and younger ages here is the high densities (8.8x10^2 and 8.7x10^2 stars/pc^3) found around AS310 and V361Cep. Additionally, the low densities of the clusters found around Hen3-949 and HD37371 (1.5x10^1 stars/pc^3) prevent the assignment of a general trend of the stellar density for clusters older than 7x10^5 yrs to decrease with age. Without the inclusion of AS310, V361Cep, Hen3-949 and HD37371 there would be a general trend for stellar density to increase with age up to 1.5x10^6 yrs and then for stellar density to decrease with age. Considering the whole age range after 6x10^4 yrs there is an overall trend for the stellar densities of the cluster to decrease with age.

From this figure it cannot be concluded that the clusters around the B stars appear to follow either of the hierarchical or monolithic formation paths. This is because there is no clear trend in stellar density, to increase and then decrease, between 1.6x10^5 to 5x10^5 yrs, as was shown in the analysis made by Bonnell *et al.* (2003, see figure 5.9). The change in stellar density that occurs in this region is abrupt and results in the transition between the region of higher stellar density clusters younger than 7x10^5 yrs and the region of clusters of lower stellar density older than 7x10^5 yrs. However, more generally there is an observable trend of stellar density decreasing with time over the whole age range in figure 5.10 between 6x10^4 to 1x10^7 yrs. There is an initial region with the rolling average of the stellar density generally above 4x10^2 stars/pc^3 and then a fall after 7x10^5 yrs up to 1x10^7 yrs. The general trend for stellar density to decrease with time can be considered to be in agreement with the long time behaviour predicted by both the hierarchical and monolithic models beyond 4x10^5 yrs. Figure 5.10 is only however, a 'snapshot' of a specific period in time of the formation of the clusters around the B stars. To have a fairer comparison with the models it is potentially important to take into

consideration the initial free-fall time for the cloud (1.9x10^5 yr) used by Bonnell *et al.* (2003) in their simulations.

Figure 5.11 shows the "real" stellar density and age of the B stars found to be in a cluster by CEREAL. The age of the stars was increased by 1.9x10^5 yr, the nominal free-fall time from Bonnell *et al.* (2003). An area which represents the same timescale as was shown in figure 5.9, between 1.6x10^5 and 5x10^5 yr, has again been shown in this figure in the same manner as in figure 5.10. After the inclusion of the free-fall time the behaviour observed for the stellar densities with time still does not appear to be in perfect agreement with the behaviour expected to be present from the analysis made by Bonnell *et al.* (2003, see figure 5.9).

The average stellar density appears to be relatively flat between 2x10^5 and 8x10^5 yr and falls to a plateau between 9 x10^5 and 3x10^6 yr. For clusters older than 3x10^6 yr there is now a general trend for the stellar density to decrease with age. For comparison the general trend expected from the work of Bonnell *et al.* (2003) was for the stellar density to increase after $\sim$2.0x10^5 yr and then decrease after $\sim$4.0 x10^5 yr for the monolithic model, or $\sim$2.5x10^5 yr for the hierarchical model.

There trend of stellar densities with age observed within the snapshot window here do not match those expected for either the monolithic or hierarchical models. Compared to the magnitude of the increase in stellar density in the hierarchical model from $<10^3$ to $>10^5$ stars/pc^3, followed by a decrease to 10^4 stars/pc^3 or from $\sim$3x10^2 to $>$1x10^4 stars/pc^3 before falling to $<$5x10^3 stars/pc^3 in the monolithic model, the changes in stellar density observed here from $\sim$9x10^2 to $\sim$3x10^1 stars/pc^3 were smaller. In addition, for the observations here the stellar density is still relatively flat (around 0.3-8.9x10^2 stars/pc^3) during the 'snapshot' of time shared with the work of Bonnell *et al.* (2003) when compared to the changes in stellar densities predicted by both models. With the rise then decrease then rise again in stellar densities observed here not being predicted by either the monolithic model or the hierarchical model in Bonnell *et al.* (2003).

The stellar density observed here does however generally fall from higher densities at earlier times to lower densities at later times ($>$8x10^5 yr) to a new lower level (0.1-3.0x10^2 stars/pc^3). The differences between the models and the observed behaviour could be due to differences in how the ages of the clusters are defined, this is because the general trend for a decrease in stellar density is in agreement with the prediction of both the hierarchical and monolithic models at larger ages. The high stellar densities ($>$1x10^2

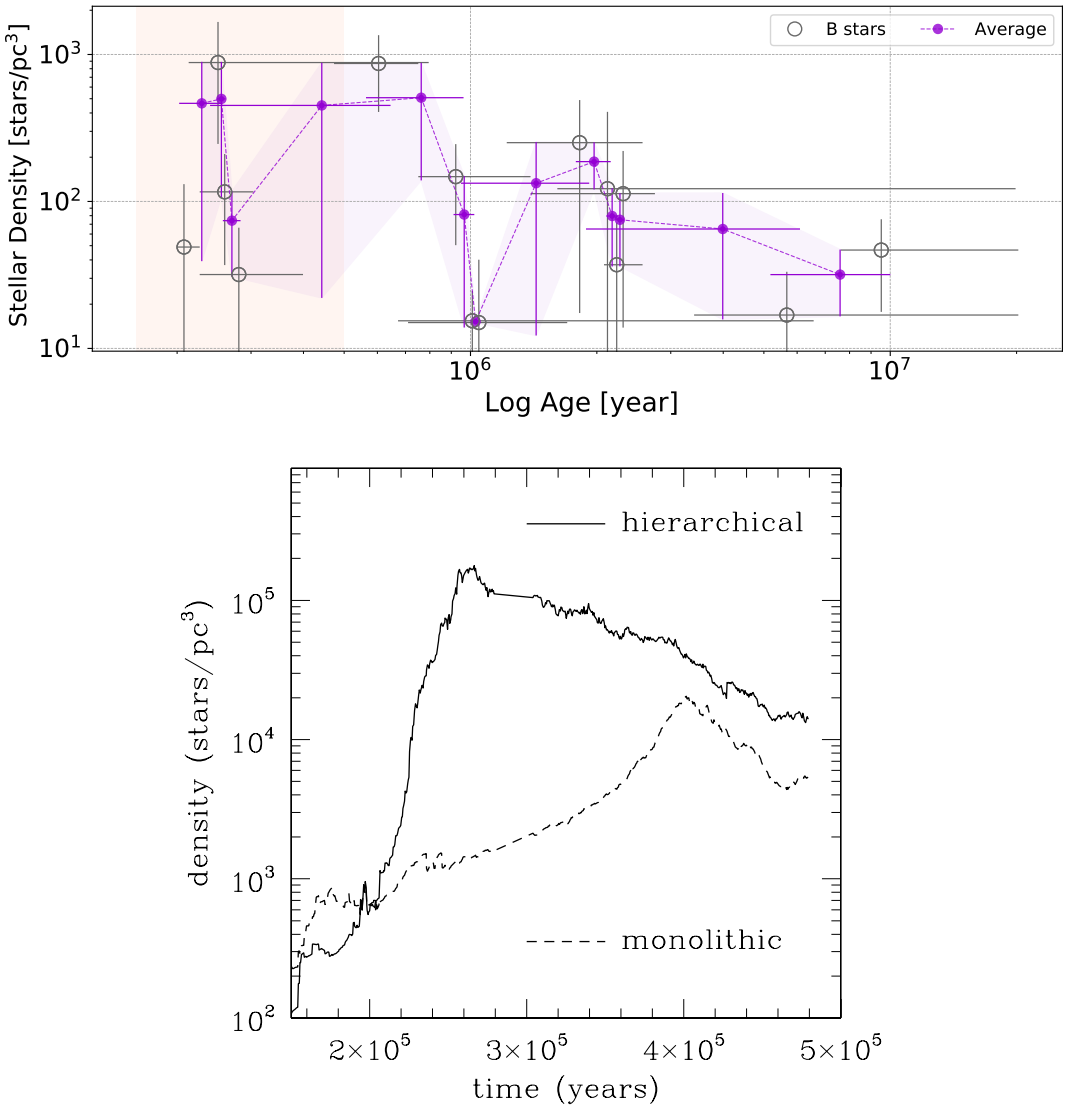


Figure 5.11: Real Stellar densities vs Age, considering the initial free-fall time for the cloud. *Top plot:* Real Stellar densities vs Age, considering the initial free-fall time for the cloud used by Bonnell *et al.* (1.9x10^5 yr 2003). This figure has the same descriptions as used in figure 5.10. *Bottom plot:*This figure shows how the local and global stellar densities evolve as as a function of time, taking intro consideration the initial free-fall time of the cloud (1.9x10^5 yr). This figure has the same descriptions as figure 5.9.

stars/pc^3) at early times (<8x10^5 yr) falling to lower densities at longer times (~3x10^1 stars/pc^3) appears in reasonable agreement with the predictions of both the hierarchical and monolithic model; however, the timescales and magnitude of the changes still differ.

An additional potential cause for this discrepancy could be in the selection of the cluster size as discussed previously (see section §2.2.1 and appendix B), the cluster radii used here are larger than those used by Hillenbrand *et al.* (1995, 0.15pc) and Testi *et al.* (1997, 1998, 1999, 0.2pc) and this leads to lower stellar densities despite including more of the potential cluster members in the calculations. It should also be noted that under 16% of the objects within the clusters in this analysis were directly observed and found by CEREAL with the remainder coming from the IMF corrections.

The simulation from Bonnell *et al.* (2003) showed a rapid and large increase in the stellar density as function of time, from $\sim 10^2$ stars/pc^3 to $\sim 10^5$ stars/pc^3 for the hierarchical model; and from $\sim 10^2$ stars/pc^3 to $\sim 10^4$ stars/pc^3 for the monolithic model. These large changes in the stellar density are not observed for the clusters found around the B stars by CEREAL, which appear to be relatively constant around $\sim$1x10^2 stars/pc^3 to $\sim$1x10^3 stars/pc^3 before 8.0x10^5 yr, followed by a decrease to under 5x10^1 stars/pc^3 by 5x10^6 yr. If the clusters observed here were all further shifted in age then the decrease in stellar densities of the clusters with ages could be in agreement with the late time behaviour of either the hierarchical or the monolithic models in Bonnell *et al.* (2003). There are large uncertainties on the ages (taken from Vioque *et al.*, 2018) of the clusters observed here which make direct assignment of the trend in stellar density with age complex, in particular where the changes occur over a small window of time 1-5x10^5 yr when several of the clusters have uncertainties in their age approximately 1x10^6 yr. For example, AS310 the cluster with the highest stellar density (therefore the peak in the observed distribution) could have an age between 2.1x10^5 and 7.9x10^5 yr (when corrected for free fall) which is a larger window of time than the snapshot of time included in Bonnell *et al.* (2003).

From the comparison of the sample of B stars in clusters with the stellar formation models it could not be concluded that these stars had been formed by either of the models explained previously. However, the 15 B stars can be seen to follow certain aspects of both models, where there are both massive stars formed in the centre of the clusters (monolithic collapse) and these are surrounded by several low mass companions (hierarchical model); but these observations do not seem to be enough evidence alone to explain how those clusters were formed.

Further analysis in this area would be of particular value in understanding the formation of clusters around intermediate mass stars, like the HAeBe stars. This would then

also provide more information and understanding about the formation of clusters around massive stars.

5.3 Conclusion

From this comparison, it was not possible to draw a definite conclusion on how massive stars and their clusters form, due to the results obtained for the B stars showing that the clusters do not appear to follow exactly either the hierarchical or the monolithic models; although, some of the expectations from both models are observed.

Both models are independent ways to explain massive star formation, but it is possible that a scenario where both theories worked together might occur, whenever the physical conditions in the molecular clouds allowed them to work (Peretto *et al.*, 2006; Rivilla *et al.*, 2013; Smilgys & Bonnell, 2017, references therein). It might be the case that, a combination of the two models is required to explain the formation of massive stars and their clusters.

However, this comparison does not give an explanation for the A-type stars that where found to be in clusters by `CEREAL`, which were not considered in this chapter. They could also follow one of the two star formation theories, but there was not enough evidence yet to explain them.

Further investigations should be made to test whether it is possible to combine both models to explain the formation of massive stars in clusters. A starting point could be the 26 B stars that were found to be in a cluster by `CEREAL` that have not yet been associated with a known star formation region. In addition, a comparison with other numerical simulations that use Gaia data in their analysis should be considered. For example, Arnold *et al.* (2022) investigated whether Moran's I statistic can be used to distinguish between the monolithic and hierarchical models.

Chapter 6

Conclusion

'My mission in life is not merely to survive, but to **thrive** and to do so with some passion, some compassion, some humour and some style'

— Maya Angelou.

The work presented in this thesis has focused on investigating the presence of clusters around known intermediate-mass pre-main sequence Herbig Ae/Be stars using the detailed astrometric data offered by the Gaia satellite. This chapter provides a summary of the findings within this thesis, along with suggestions of the future work that could be carried out following on from the work contained in this thesis.

Herbig Ae/Be stars are pre-main sequence stars of intermediate-mass between 2-10 $M_{\odot}$; Herbig Ae/Be stars represent the most massive objects to experience an optically visible pre-main sequence phase, bridging the transition between low- and high-mass stars (Gomez *et al.*, 1993; Mendigutía *et al.*, 2012; Testi *et al.*, 1997). The sample of 269 Herbig Ae/Be stars studied in this work were the result of a cross-match between the Chen *et al.* (2016) and Vioque *et al.* (2018) catalogues.

The presence of clusters around these Herbig Ae/Be stars was assessed using Gaia DR2 (Gaia Collaboration *et al.*, 2016b, 2018b) , which provided a catalogue of high-precision parallaxes and proper motions for over 1 billion sources. For this work, a tool was developed to evaluate the presence of clusters around the Herbig Ae/Be stars.

Chapter 2: ClustER detEction ALgorithm: CEREAL

Chapter 2 introduced the new algorithm created in this work with the goal of detecting and analysing the presence of clusters around intermediate mass stars using data from Gaia. In this chapter the method used to gather the data from the Gaia archive (§ 2.1.1) was given and a detailed description of the process with which the algorithm evaluated the presence of cluster around the targets (§ 2.1.2) was provided.

To test the ability of CEREAL to assess the presence of clusters around Herbig Ae/Be stars, an analysis around known clusters NGC6475, Blanco1, IC2602 and αPer, was performed first (§ 2.2). These known clusters were previously studied by Gaia Collaboration *et al.* (2017) and van Leeuwen (2009). The results obtained with CEREAL for the known clusters agreed very well with the analysis carried out by Gaia Collaboration *et al.* (2017) and van Leeuwen (2009) for the same sources. This demonstrated that CEREAL was a powerful algorithm which can take large data sets as an input, without the need for making a pre-selection of the sample; and then retrieve objects which share similar astrometric parameters; like parallax and proper motion, with the known target. CEREAL also gives the user total control over the selection process for the sample, with which they can then obtain the results.

Additionally, a quantitative analysis (§2.2.1) was carried out, where the density profiles of the sample of known clusters and Herbig Ae/Be stars were studied with the aim to find a value of the cluster radius, which was used to compare with the theoretical models in Chapter 5. This quantitative analysis also had the result of reducing the number of objects that were then included in the comparison. Several methods were tested for their ability to define a cluster radius (where the density rose 3σ above the background density); however, none of these methods were entirely successful in evaluating the cluster radii. It was not possible here to easily use the density profiles alone (without additional parameters such as parallax and proper motions) to statistically derived the cluster radius.

From the methods tested the rolling window method provided a series of cluster radii for each star that varied in position dependent on the window size. The cluster radius was defined as the first point to rise 3σ above the background density when moving from the outside of the density profile towards the central star. The dependence of the assigned cluster radii on the window sizes was evaluated for each star. The cluster radii used in

the later analysis were taken from the smallest window size for which a consistent cluster radius was seen (within the errors).

The cluster radii used in this work (0.21 - $\sim$3 pc) were generally larger than those utilised previously by Hillenbrand *et al.* (1995, 0.15 pc) and Testi *et al.* (1997, 1998, 1999, 0.2 pc). For the sample of Herbig Ae/Be stars considered here the cluster radii were between 0.21 and 1.07 pc with an average of 0.62 pc these larger cluster radii will contain most of the possible cluster members when compared to those utilised by Hillenbrand *et al.* (1995, 0.15 pc) and Testi *et al.* (1997, 1998, 1999, 0.2 pc).

Chapter 3: Presence of clusters around Herbig Ae/Be stars

In Chapter 3 the results obtained from the assessment of the presence of clusters around the sample of known HAeBe stars made by `CEREAL` using Gaia DR2 was presented. The HAeBe stars sample was obtained from the cross-match between the Chen *et al.* (2016) and Vioque *et al.* (2018) catalogues.

The analysis of 269 known HAeBe stars with `CEREAL` has shown the capability of `CEREAL` at finding clusters around stars using Gaia DR2 data. Although, the classification of this sample was not a trivial work to do, because there was not an agreed well defined formula on which to categorize the presence of clusters around HAeBe stars, $76(28\%)$ of the stars were either found to be in or to potentially be in a cluster. The number of stars found to be in a cluster might appear small but this still represents a big improvement on the 15 stars found previously by Testi *et al.* (1999, § 3.4), to be in either large or small clusters. These 76 stars also represent the first catalogue of HAeBe stars found to be in clusters by applying a single tool over the same data homogeneously (§ 3.2.4).

Chapter 3 evaluates the ability of `CEREAL` at finding clusters by testing it in a comparison with other available algorithms, which can also be used to evaluate the presence of overdensities in large samples, like Gaia DR2. The algorithms used in this comparison were *DBSCAN* (Ester *et al.*, 1996), *OPTICS* (Ankerst *et al.*, 1999) and *HDBSCAN* (Campello *et al.*, 2015; McInnes *et al.*, 2017), which are a sub-type of machine learning algorithms dedicated to the study of overdensities. The analysis using the clustering algorithms followed the procedure described in Cánovas *et al.* (2019).

Each independent technique analysed a sample of 255 HAeBe stars, following this, the results of each of the clustering algorithms was combined and then compared with

CEREAL. The results obtained from this comparison showed that each method found similar numbers of stars in clusters by using the same data, but with different quality controls and techniques. The use of different quality controls might have led to the differences in results. Where for example, the different methods could find the same clusters, but not the same member stars within the clusters (§ 3.3.1) which happened on 28 occasions; or where a certain algorithm could not find a cluster because in the application of the many different quality control techniques to extract the sources before the analysis (§ 3.3.2) all the potential cluster members were removed, which happened on 4 occasions.

Even when CEREAL was run with the DCAs conditions, it still was possible to find a similar number of clusters (29 yes, 22 maybe from 51 stars previously assigned to the gold and silver samples). From the 3 DCA's, *HDBSCAN* was the only that found the same number of clusters (51 stars) than CEREAL, with DBSCAN and OPTICS finding fewer (42 and 46 stars, respectively; see appendix G).

Although, there were differences between the results of the assessments made with the DCAs and CEREAL, it was possible following the comparison, through the combination of these results, to define two separate samples (gold and silver sample) which included all those stars that were found to be or possibly be in a cluster by each independent method. In addition, the results of this comparison clearly demonstrated the reliability of CEREAL, and showed the importance of the fact that CEREAL gave more freedom to the user for the evaluation of the presence of clusters. From the work in this chapter it was clear that CEREAL was shown to be a highly effective method to find clusters, of any size, from any sample.

In Chapter 3 a re-analysis of the sample of 43 HAeBe stars from Testi *et al.* (1999) was performed using CEREAL (§ 3.4). In general, CEREAL found similar results, for the sample covering the spectral types from O9 to A7, to those reported by Testi *et al.* (1999). The results of CEREAL mostly agreed with the spectral type limit defined by Testi *et al.* (1999), that stars earlier than B7 could possibly have a cluster around them and that stars later than B7 were not associated with a cluster. However, an important difference in the re-analysis was that CEREAL was able to find clusters around 4 A-type stars. These 4 stars were included in the sample from Testi *et al.* (1999), and these stars were considered in their analysis to be either in a small aggregate or to be background stars (see figure 3.14). The difference between the assessment of the A-type stars made by Testi *et al.* (1999) and by CEREAL might be due to the selection of a larger radius around the central

sample (by CEREAL) or it could be because the A stars were not the most massive objects in their clusters, which means they were not at the centre of their cluster. This analysis performed with CEREAL has proved that it was possible to find more clusters around the same objects in the sample from Testi *et al.* (1999), by using the data from Gaia DR2 and large cone search around the HAeBe stars, further demonstrating the value of CEREAL for searching for clusters around HAeBe stars.

Chapter 4: Cluster properties of the Herbig Ae/Be stars

The work described in chapter 4 investigated the influence that certain parameters, such as the parallax and G magnitude, could have had on the cluster classification found for each of the HAeBe stars, specifically the B and A stars, using CEREAL. This chapter also investigated if there was a difference between the clustering fractions of the B and A type stars; this investigation also considered whether the spectral type of the HAeBe star had any influence on its position within its cluster.

The analysis of the relationship between the spectral type of the targets and their parameters, such as parallax (§ 4.1) and G magnitude (§4.2), showed that the objects that were not found to be in clusters were distributed over the full range of parallax and G magnitude. In addition, the fraction of stars found not to be in clusters appeared to be constant, within their error bars, across the whole range of these parameters.

This contrasted with the stars that were classified by CEREAL to be in, or potentially be in a cluster; these stars appeared to be preferentially distributed within specific ranges of both parallax (0 to 7 mas) and G magnitude (6 to 18 mag). The apparent preferential ranges were then studied in greater detail to interpret if this observation was due to the presence of biases within the sample; the effect of these potential biases could then be evaluated by producing sub samples where only the objects within specific ranges of parallax (1 to 6 mas), G magnitude (7 to 17.5) and the combination of both (§4.3) these selections were included. These particular ranges in the parameters were chosen through a careful consideration of the nature of the B-type and A-type stars and the limitations and the quality of the Gaia DR2 data.

For parallax, the B stars in general tended to be located at larger distances than the A stars. For HAeBe stars that had G magnitudes brighter than 6 mag it was not possible to find clusters around them. This limitation could have been due to the nature of bright

stars, that they were so bright that they obscured their neighbours; therefore, it was not possible to observe any low mass companions for these bright objects even if they existed.

After these selections were made, statistical analysis on the parallax and G magnitude distributions of the B and A stars (that were found to be in, or to not be in clusters) was performed using a KS test. The results of this analysis showed that the distribution of the stars found to be in and found not to be in clusters appeared to be drawn from similar distributions and could therefore be compared to each other; this result was found for the whole sample and the samples of B and A stars.

In the analysis of the clustering fractions by spectral type that then followed, the fraction of the B and A type stars that were found to be in clusters appeared to be evenly distributed across all the sub spectral types, within their error bars. This analysis did not find any difference between the fraction of B and A stars that were found to be in, or not in a cluster and they appeared to be similar (§4.4).

Further analysis in chapter 4 then studied in more detail the possible effect that the spectral type, even the sub spectral type, could have on the position where the HAeBe star was within its cluster (§4.4.1). This analysis concluded that the B stars in clusters tend to be at the centre of their clusters. Being in the center of their clusters could then be interpreted as meaning that the B stars were likely to be the most massive objects in their clusters. In contrast to the results observed for the B stars classified to be in clusters by `CEREAL`, the A stars tended to not be in the centre of the clusters. This was interpreted as meaning that, in general, the A stars classified as yes were part of a sub cluster. This conclusion was further supported by the fact that most of the A stars in the sample, that were found to be in clusters, were located in either known star forming regions or known clusters. For example, a comparison of the A type stars found to be in clusters by `CEREAL` with SIMBAD showed that 6% of these were found to belong to a known cluster.

Further to the comparison of the B and A type stars, this analysis was then extended to consider the sub spectral types within these. This sub spectral type analysis showed that for B stars across all their sub spectral types the B stars tended to be at the centre of their clusters (§4.4.2). When this analysis was carried out for the A stars, again there seemed to be no sub spectral type dependence with the A stars across all sub spectral types in generally being found to not be in the centre of their clusters; the conclusion from this was that for all sub spectral types A stars were part of a sub cluster.

Comparing the results of this sub spectral type analysis with that of Testi *et al.* (1997) showed significant disagreement. In this work clusters were found around B stars over the whole range of sub spectral types, not just for objects of type B7 and earlier as was found by Testi *et al.* (1997). Unlike the A stars, which were found to be in sub clusters rather than being the centre of their own clusters, the result that clusters were found around late type B stars was not caused by these not being the centre of their own clusters. This work instead found an almost uniform distribution of the fraction of stars found to be in clusters around B types stars across all sub spectral types, and these B stars were additionally found to be at the centre of their own clusters. This result could potentially be due to the larger search radius utilised in this work, compared to the work of Testi *et al.* (1997), along with the more detailed astrometric data from Gaia DR2 available during the current work that was not available to Testi *et al.* (1997).

In addition, `CEREAL` found clusters around A type stars where Testi *et al.* (1997) did not find any. This result could be due to a combination of `CEREAL` searching larger areas around the HAeBe stars, along with the fact that the A stars found to be in clusters in this work, were not found to be at the centre of their own clusters, for all sub spectral types. The work of Testi *et al.* (1997) assumed that HAeBe stars would be at the centre of their own clusters and this combined with a narrow search radius may have prevented them observing the clusters, that the A stars were found to be part of here.

Chapter 5: Cluster formation around B stars: observations versus theory

In Chapter 5, a comparison between the theoretical models, which simulated how stars were formed within stellar clusters, and the cluster found around the HAeBe stars by `CEREAL` was carried out. The aim of this comparison was to evaluate whether the clusters around the massive stars, in this sample, appeared to follow either of these models.

Prior to the comparison with the models, a small pre-selection was carried out, using the condition that the B stars to be considered had to be located at the centre of their clusters (§4.4 and 4.4.2). The rolling window method was used to assign a cluster radius for all the HAeBe stars considered in this analysis (§ 2.2.1 and §5.2.1). These radii are expected to be sufficiently large to include most of the objects within the clusters. At a larger radius, the source counts are dominated by Galactic background/foreground stars; and, at smaller radii the effects of cloud extinction can be seen, as stellar surface densities

decrease with decreasing radial distance, but still the source counts begin to increase with decreasing radial distance due to the presence of small stellar clusters near the centre (Hillenbrand *et al.*, 1995).

The number of stars in each cluster was then calculated from the number of objects found within this radius. To allow for comparison with the models the number of objects in these clusters needed to be converted into their stellar densities. When the stellar density for each cluster was calculated it was important to consider the impact of the minimum mass limit observed for each of the clusters. This minimum mass limit leads to a significant underestimation of the true stellar densities of the clusters, due to the observational inability to detect very small and faint objects. These missing objects were corrected for using the IMF to evaluate what the stellar density of the complete cluster would be. These corrected stellar densities could then be compared to the theoretical models of massive star formation, the stellar densities calculated with the IMF correction had an average of 194 stars/pc^3 compared to only 30 stars/pc^3 stars prior to this correction; this showed that under 16% of the low mass companions assigned to the HAeBe stars in clusters were observed. The numerical simulation to which the results of this analysis were compared were carried out by Bonnell *et al.* (2003). These simulations followed the fragmentation of a turbulent molecular cloud, and then the subsequent formation and early evolution of a stellar cluster (§5.2.7).

The comparison of the stellar densities of the clusters found around the B stars by CEREAL with the theoretical models (§ 5.2.7) did not provide any definite conclusion about the formation of massive stars and their clusters. This was because the B stars in this sample did not appear to follow either the hierarchical or the monolithic models. The star formation simulations from Bonnell *et al.* (2003) show a rapid increase in the stellar density as a function of age (from $\sim$10^2 stars/pc^3 to $\sim$10^5 stars/pc^3) and monolithic model (from $\sim$10^2 stars/pc^3 to $\sim$10^4 stars/pc^3) which was not observed for the clusters found around the B stars by CEREAL. The stellar densities found by CEREAL appeared to be relatively flat at early ages ($<$7.5x10^5 yr) before beginning to fall at later ages ($>$2.0x10^6 yr). For the stellar densities found by CEREAL at early times there was a larger variance in stellar densities (3.1x10^1 - 8.8x10^2 stars/pc^3; $<$7.5x10^5 yr) than at later times (1.7x10^1 - 1.1x10^2 stars/pc^3; $>$2.0x10^6 yr). The general trend for stellar density to fall with age after appears in line with that observed by both the monolithic (after 4.0x10^5 yr) and hierarchical (after 2.5x10^5 yr) models in Bonnell *et al.* (2003). This lack of perfect agreement with the

models still appears to be there once a correction for the free-fall time included in the models was applied to the ages of the results found by CEREAL.

Although, there is not a perfect agreement with either of the models, some of the expectations from both models are observed in the analysis of this sample of B stars. The observation of some of the predictions from both models may open up the possibility that some aspects of both models may have been present in the evolution of the clusters observed in this work. It might be that by considering a combination of both of the models it will be possible to explain the formation of massive stars and their clusters (Rivilla *et al.*, 2013; Smilgys & Bonnell, 2017, references therein).

6.1 Future work

This section presents ideas for future work that could be carried out, to further the work described in this thesis; this section will also mention potential improvements to the methods described earlier, that could further increase the impact of this methodology in any future work.

Chapter 2: ClustER detEction ALgorithm: CEREAL

CEREAL was created as a tool to assess the presence of clusters around HAeBe stars. During this thesis the ability of CEREAL was tested, by first analysing a sample of known clusters (§ 2.2) and then a sample of known HAeBe stars (§3). The results of both these studies demonstrated that CEREAL was a good tool for analysing the presence of clusters as it produced similar results for the same samples when compared with several other studies such as Cánovas *et al.* (2019); Gaia Collaboration *et al.* (2017); van Leeuwen (2009).

However, there are a few improvements to the algorithm that could further increase the power of CEREAL as a tool for detecting clusters. For example if a definition for what a cluster is and what the indicators might be observed when a cluster is present are pre-defined, prior to using CEREAL to perform the cluster analysis, differences in classifications (by different users) could potentially be avoided.

The introduction of this concept could make this algorithm more accessible to all of the scientific community (and it would also require less prior knowledge to use it) and this

could potentially allow it to be used more easily for analysis around other types of stars; whilst simultaneously, reducing variances in the results that CEREAL would produce, due to different user's individual selections.

A further improvement to CEREAL that could be considered would be the inclusion of a fitting function to the density profiles in the analysis. This function could then be used as a first estimation of the possible cluster radius.

Chapter 3: Presence of clusters around Herbig Ae/Be stars

The classification system (*yes*, *maybe* or *no*) was created for CEREAL; when this classification system was used, in the analysis of a sample of 269 HAeBe stars, 41 (15%) stars were found to be in clusters (classified as *Yes*), 35 (13%) stars were found to be potentially in a cluster (classified as *Maybe*) and 193 (72%) stars were found to not be in clusters (classified as *No*).

No further analysis was made for the 193 (72%) stars classified as *No* by CEREAL (appendix E), because that analysis would have been beyond the scope of the present work. A further study of the stars not found to be in clusters could represent an important extension to the work included in this thesis. A detailed analysis focused on these objects could be carried out to determine if these HAeBe stars are indeed not in clusters, are isolated HAeBe stars, or if they are in a cluster that CEREAL was not sensitive to either due to the nature of the data available at the time, or the nature of the HAeBe stars themselves.

The sample of 144 stars in common to not be in a cluster by both CEREAL and the DCAs would provide a good starting point for further analysis of this kind in combination with newer Gaia data releases.

The comparison of the cluster classifications found by CEREAL and the DCAs produced two catalogues of HAeBe stars found to be in clusters (Gold and Silver). The gold catalogue contained all those objects found by both and the DCAs to be in clusters, the silver catalogue contained those objects that were a mix of yes and maybes. The gold and silver catalogues contain 28 and 23 stars, respectively, these stars would make an excellent staring point in the future study of cluster members around HAeBe stars.

To increase the statistics on the fraction of HAeBe stars found to be in clusters, would require analysis of a larger sample of stars. This increased sample could be obtained

from Vioque *et al.* (2020), which found 8470 new pre-main sequence candidates of which at least 1361 sources were potentially new Herbig Ae/Be candidates according to their position in the Hertzsprung-Russell diagram (see figure 1.5). In addition, the use of more accurate data, which will likely be available through future Gaia data releases, will allow improved analysis of the clustering properties of Herbig Ae/Be stars with CEREAL.

Chapter 4: Cluster properties of the Herbig Ae/Be stars

The analysis performed with CEREAL has produced the first catalogue of Herbig Ae/Be stars to be classified as being part of a cluster by using a homogeneous data set and methodology. However, a full characterization of the cluster members within the 76 clusters around the Herbig Ae/Be stars was not carried out.

In order to fully characterize these cluster members it would be first necessary to carry out a homogeneous spectroscopic study for both the HAeBe stars and their possible cluster members; this was beyond the scope of the work carried out here, but would represent a valuable research area for further study.

Therefore, a multi-wavelength study of a large samples of objects would improve the knowledge of them and the areas they form in (Rivilla *et al.*, 2013). A multi-wavelength study could also help understanding, from an observational perspective, of how these clusters are formed and provide insights that are currently missing about the formation of massive stars. In addition, a multi-wavelength study would help in the characterization of these clusters, by using a combination of their photometric data to fit isochrones to calculate the age and mass of the sample and also this would allow for the estimation of their extinction.

An example of isochrone fitting is shown in figure 6.1. This shows the colour-magnitude distribution of two known clusters and two Herbig Ae/Be stars classified by CEREAL and a test fitting the isochrones from MIST (Dotter, 2016) and PARSEC (Bressan *et al.*, 2012; Marigo *et al.*, 2017) to these.

Additionally, a sample of the HAeBe stars, which include HD245185, MWC1080 and UYOri, found in clusters have been observed with the spectrograph Boller & Chivens located in the San Pedro Martir National Astronomical Observatory in Mexico. This provided low resolution spectra (R ∼1000) from approximately 3900 to 7100 $\mathring{A}$. This

represents the beginning of the characterization of the clusters found for HAeBe with CEREAL that is required and this represents an important area for future work.

Chapter 5: Cluster formation around B stars: observations versus theory

More studies should be made around Herbig stars to have a better understanding of the formation of those objects and their clusters. A starting point could be the 26 B stars that were found to be in a cluster by CEREAL that have not yet been associated with a known star formation region.

Additionally, further investigations should be made to test whether it is possible to combine both models to explain the formation of massive stars in clusters (Peretto *et al.*, 2006; Rivilla *et al.*, 2013). A recent model that could be used for this analysis is the the numerical simulation proposed by Arnold *et al.* (2022).

Further work is required to bridge the gap between the results observed here and the current theoretical predictions of massive star formation.

6.2 Closing remarks

In conclusion, this thesis presents an assessment of the presence of clusters around a sample of previously known Herbig Ae/Be stars using astrometric data from Gaia DR2.

The analysis was carried out with CEREAL, which is a novel clustering algorithm developed in this work. This algorithm was able to find clusters around 76 Herbig Ae/Be stars, which increases the number of Herbig Ae/Be stars associated with clusters from the previous studies (Testi *et al.*, 1999). These clusters represent the first homogeneous classification made for this type of stars using a single dataset. Also, comparison of CEREAL with other clustering algorithms and previous studies has demonstrated that CEREAL is a simple, but powerful, algorithm that can find clusters, of any size, from any sample. CEREAL, along with several examples of its use for cluster classification, is available from `https://github.com/yumiry/CEREAL`.

Additionally, the analysis made in this thesis has shown that it is possible to find clusters around all the sub-spectral type of B stars, which differs from the results found by Testi *et al.* (1997, 1998, 1999) which found that clusters around HAeBe stars were limited to the sub spectral types range from B7 and earlier. Another important difference

between the work here and that of Testi *et al.* (1997, 1998, 1999) is that CEREAL found clusters around 4 A-type stars.

More detailed analysis of these observations and further work with the theoretical models is needed to fully explain the formation of intermediate mass stars and their clusters.

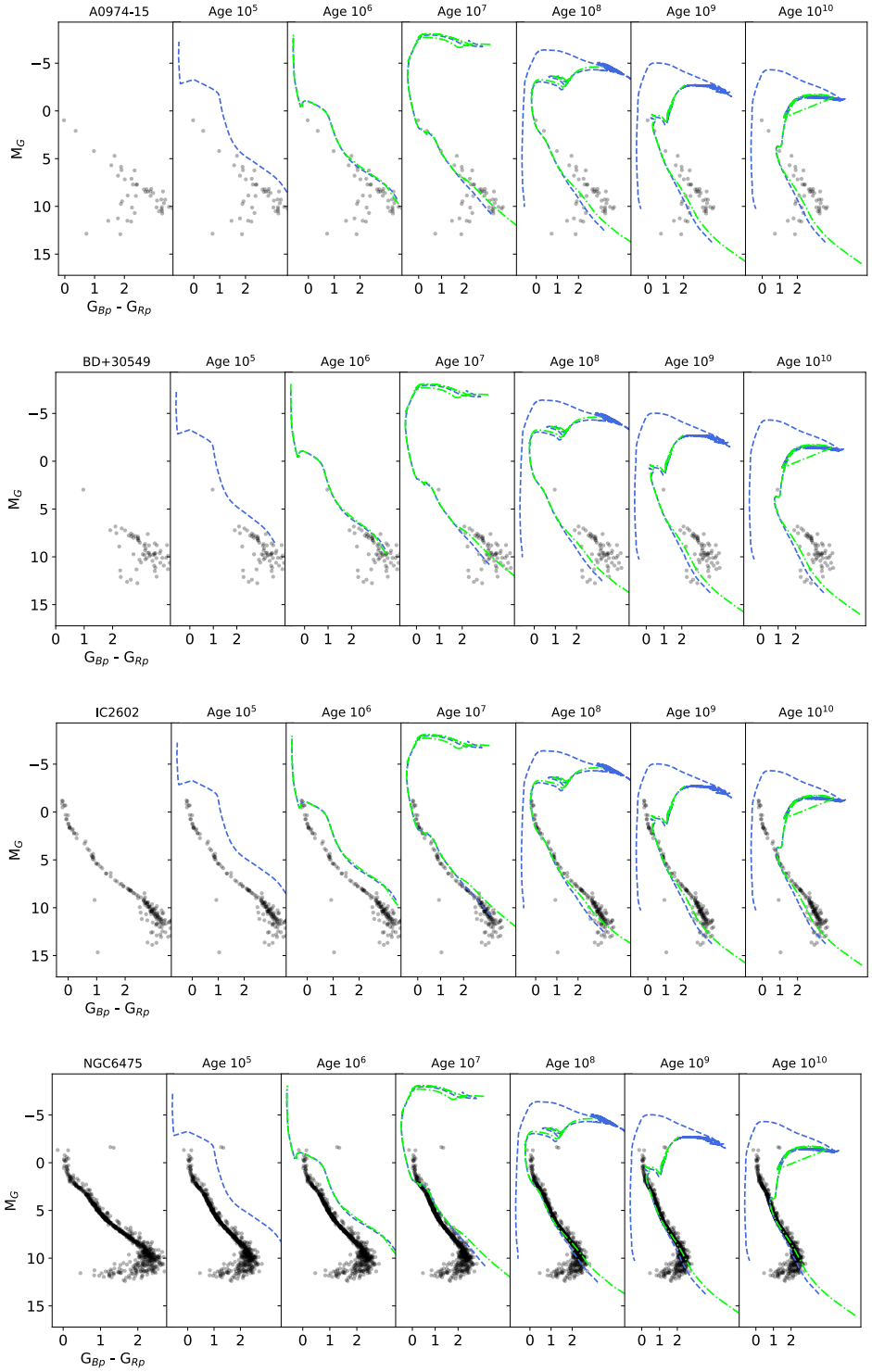


Figure 6.1: Colour-Magnitude distribution of a sample of known clusters and Herbig Ae/Be stars. The black filled points represent the stars selected by CEREAL and the light blue and green dashed lines represent the different isochrones taken from MIST (Dotter, 2016) and PARSEC (Bressan *et al.*, 2012; Marigo *et al.*, 2017).

Appendix A

CEREAL flow diagram

A schematic representation of the step-by step process on how CEREAL finds clusters around the objects of interest using Gaia data describe in section 2.1.

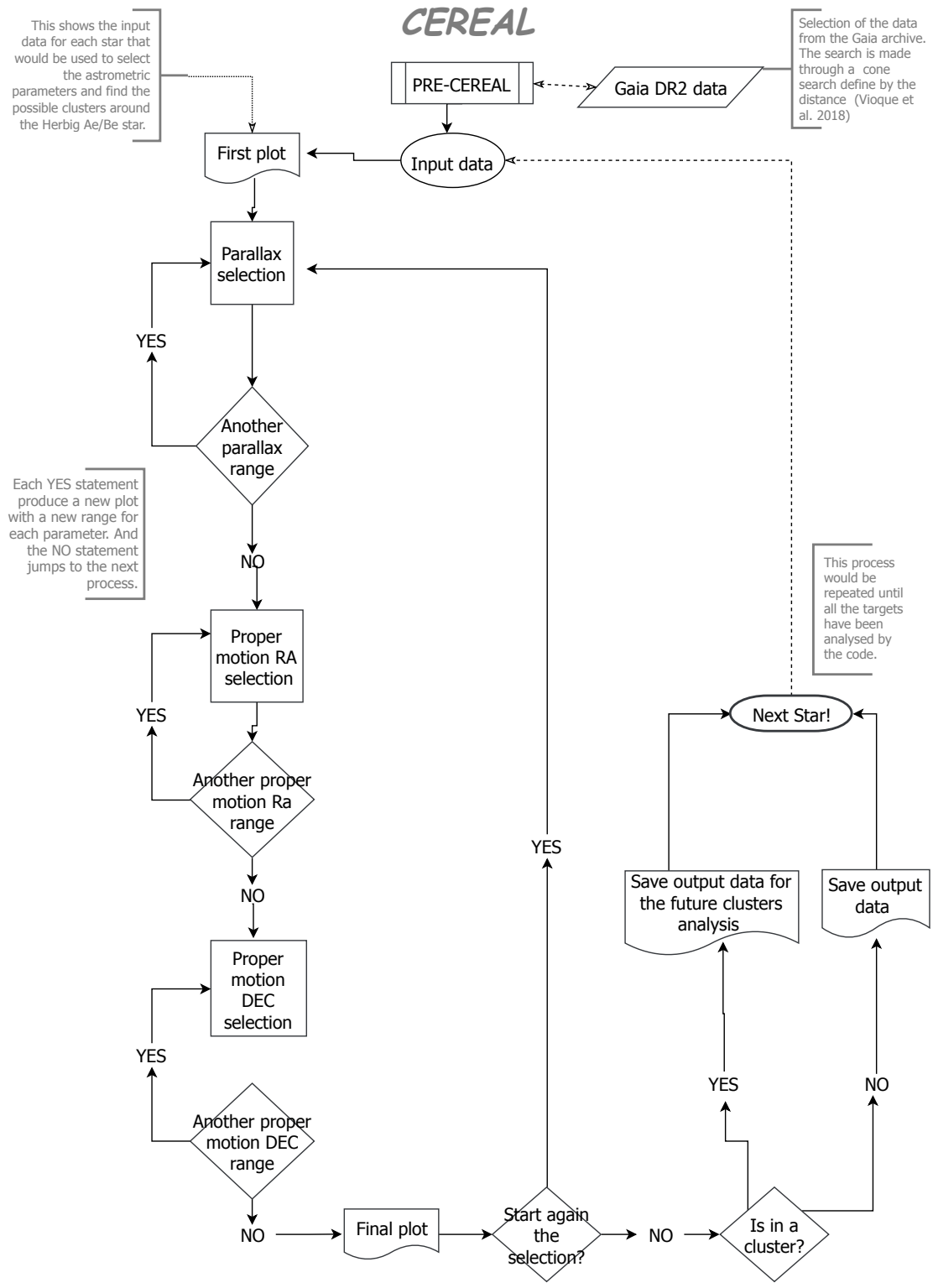


Figure A.1: CEREAL Flow diagram. Step-by-step the process of the code described in § 2.1.

Appendix B

Quantitative analysis

This appendix contains the quantitative analysis introduced in section §2.2.1 for the other known clusters studied in chapter 2 and a sample of Herbig Ae/Be stars.

3σ method

As mentioned in section §2.2.1 finding a particular value for the cluster radius for the known clusters and the Herbig Ae/Be stars was very difficult; it was not possible to find only one cluster aperture that contained densities greater than the background at the 3σ level. Figure B.1 and table B.1 show quantitative analysis for the known cluster Blanco 1 where the stellar density is significantly greater than the background density between the distance of 1.16 and 3.87 pc, with approximately 34 possible cluster members observed. A further example is also shown in figure B.2 and table B.2 where, the stellar density is significantly greater than the background density between the distance of 1.49 and 3.97 pc, with approximately 35 possible cluster members observed.

The Herbig Ae/Be stars used in this analysis were taken from the catalogue of clusters around Herbig Ae/Be stars presented in tables 3.1 and 3.2. The objects selected from these tables were the B stars classified by `CEREAL` as yes (Y) and maybe (M).

As described in section §2.2.1, this study focused only on the use of the density profiles to find the radius of the possible clusters around the Herbig Ae/Be stars. The same procedure utilised in section §2.2.1 was applied to the study of the Herbig Ae/Be stars previous classified as *yes* and *maybe* and this resulted in a reduction in the number

of stars found with respect to those found in section §3.2.4. This reduction in the number

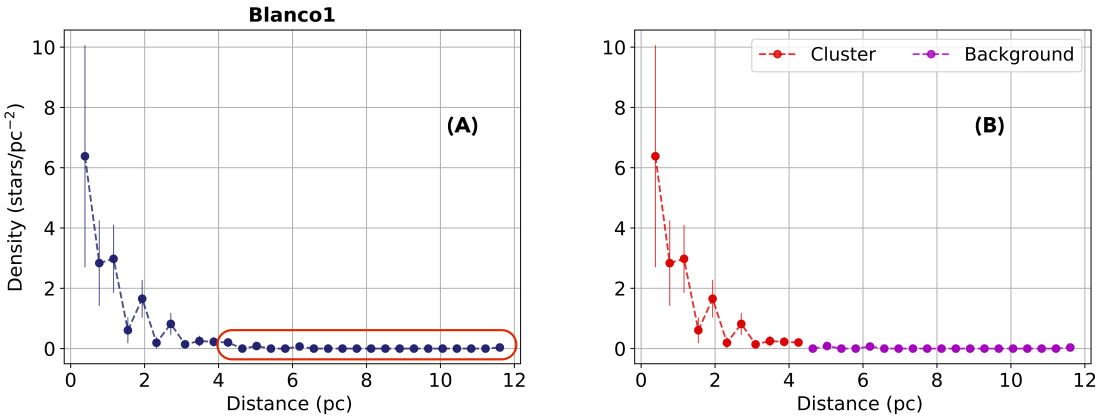


Figure B.1: Background selection with Gaia DR1 data for the known cluster Blanco 1. Panel (A) shows the density profile obtained by CEREAL previously (dark blue filled circles) where the possible location of background is shown in the red box. This area, which starts at $\sim$ 6pc, was used in the first evaluation of the background. Panel (B) is the same density profile where the anticipated location of the cluster (light red filled circles) and background (purple filled circles) are shown.

Table B.1: Cluster aperture and densities for the known cluster Blanco 1 with Gaia DR1 data

Area	Cluster aperture pc^2	Distance pc	Number of Stars	Density stars/pc^{-2}
(1)	(2)	(3)	(4)	(5)
Background	366.28	4.26	3	0.01$\pm$ 0.004
Cluster	0.47	0.39	3	6.37 $\pm$ 3.68
	1.88	0.77	7	3.71 $\pm$ 1.41
	4.23	1.16	14	3.3 $\pm$ 0.88
	7.52	1.55	16	2.12 $\pm$ 0.53
	11.75	1.93	23	1.95 $\pm$ 0.41
	16.93	2.32	24	1.41 $\pm$ 0.29
	23.04	2.71	29	1.25 $\pm$ 0.23
	30.09	3.09	30	0.99 $\pm$ 0.18
	38.09	3.48	32	0.83 $\pm$ 0.15
	47.02	3.87	34	0.71 $\pm$ 0.12

Columns (3) represent the distance from the centre of the cluster.

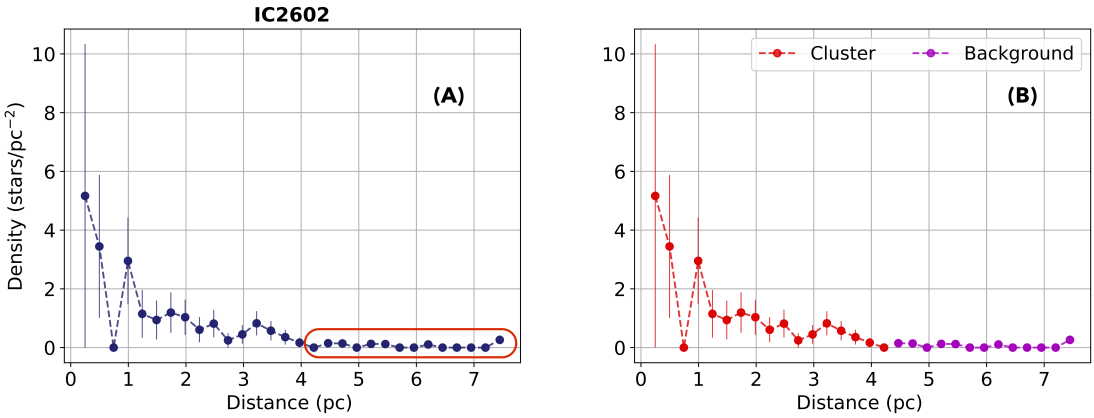


Figure B.2: Background selection with Gaia DR1 data for the known cluster IC2602. See description of figure B.1.

Table B.2: Cluster aperture and densities for the known cluster IC2602 with Gaia DR1 data

Area	Cluster aperture pc^2	Distance pc	Number of Stars	Density stars/pc^{-2}
(1)	(2)	(3)	(4)	(5)
Background	118.29	4.22	8	0.07 ± 0.02
Cluster	0.19	0.25	1	5.1 ± 5.17
	0.77	0.5	3	3.81 ± 2.24
	1.74	0.74	3	1.65 ± 0.99
	3.1	0.99	7	2.19 ± 0.85
	4.84	1.24	9	1.79 ± 0.62
	6.97	1.49	11	1.51 ± 0.48
	9.49	1.74	14	1.41 ± 0.39
	12.39	1.99	17	1.3 ± 0.33
	15.68	2.23	19	1.14 ± 0.28
	19.36	2.48	22	1.07 ± 0.24
	23.42	2.73	23	0.91 ± 0.2
	27.88	2.98	25	0.83 ± 0.18
	32.72	3.23	29	0.82 ± 0.16
	37.94	3.48	32	0.78 ± 0.15
	43.56	3.72	34	0.71 ± 0.13
	49.56	3.97	35	0.64 ± 0.12

Columns (3) represent the distance from the centre of the cluster.

of stars was due to many of the objects analysed following section §2.2.1 did not meet the condition of having a density that was considered to be sufficiently high, when compared to the density of the background.

The analysis described in section §2.2.1 shows how difficult it was to determine a radius for the known clusters, which in general have more stars within them than the ones found in the clusters around the Herbig Ae/Be stars. It was also found during the analysis of the HAeBe stars that they had multiple cluster apertures which met the condition of being statistically significantly higher than the background density (3σ level). An example of the analysis carried out in this section is show in figure B.3 and table B.3 for the Herbig Ae/Be star V361Cep and figure B.4 and table B.4 for the Herbig Ae/Be star ILCep. The result for V361Cep was that all the cluster apertures between 0.16 and 0.99 pc had densities statistically greater than the background; with the result for ILCep being, that all cluster apertures larger than 0.35 pc were 3σ over the background level.

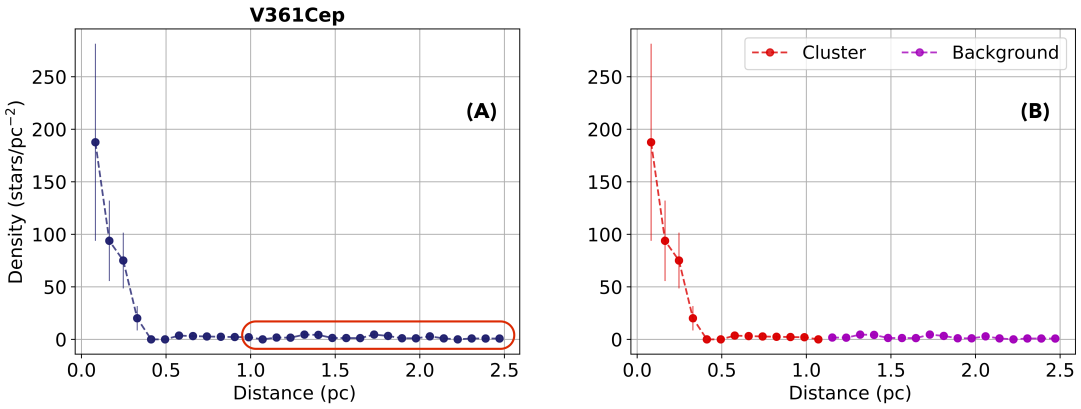


Figure B.3: Background selection with Gaia DR2 data for the Herbig Ae/Be star V361Cep. Panel (A) shows the density profile obtained by CEREAL previously (dark blue filled circles) where the possible location of the background is shown in the red box. This area, which starts $\sim$ 1pc, was used as a first approximation for the calculation of the background. Panel (B) shows the same density profile where the estimated location of the cluster (light red filled circles) and background (purple filled circles) are shown.

Table B.3: Cluster aperture and densities for the Herbig Ae/Be star V361Cep with Gaia DR2 data

Area	Cluster aperture pc^2	Distance pc	Number of Stars	Density $stars/pc^{-2}$
(1)	(2)	(3)	(4)	(5)
Background	15.58	1.07	27	1.73 ± 0.33
Cluster	0.02	0.08	4	187.62 ± 93.81
	0.09	0.16	10	117.26 ± 37.08
	0.19	0.25	18	93.81 ± 22.11
	0.34	0.33	21	61.56 ± 13.43
	0.53	0.41	21	39.4 ± 8.6
	0.77	0.49	21	27.36 ± 5.97
	1.04	0.58	22	21.06 ± 4.49
	1.36	0.66	23	16.86 ± 3.51
	1.73	0.74	24	13.9 ± 2.84
	2.13	0.82	25	11.73 ± 2.35
	2.58	0.91	26	10.08 ± 1.98
	3.07	0.99	27	8.79 ± 1.69

Columns (3) represent the distance from the centre of the cluster.

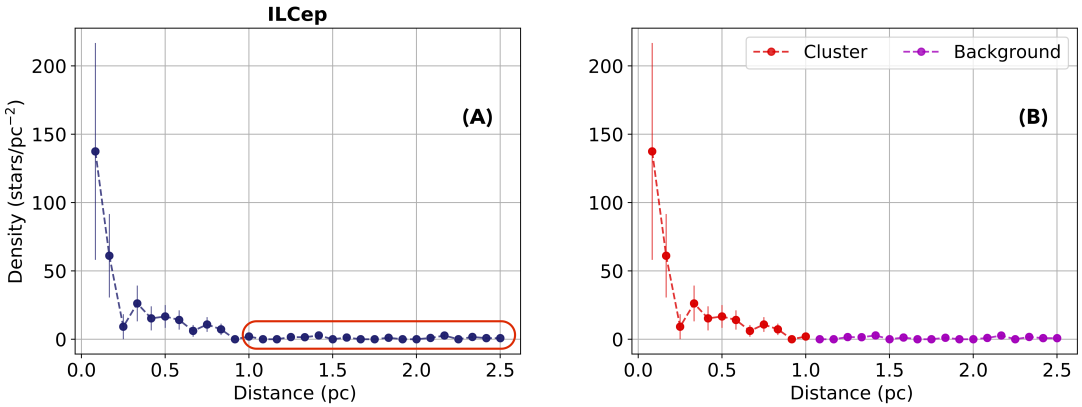


Figure B.4: Background selection with Gaia DR2 data for the Herbig Ae/Be star ILCep. See description of figure B.3

Rolling window method

Following this, the same rolling window analysis that was carried out for the known clusters (in section §2.2.1) was carried out for the sample of HAeBe stars here. In the rolling window analysis the sum of the objects and ring areas in the original density

Table B.4: Cluster aperture and densities for the Herbig Ae/Be star ILCep with Gaia DR2 data

Area	Cluster aperture pc^2	Distance pc	Number of Stars	Density stars/pc^{-2}
(1)	(2)	(3)	(4)	(5)
Background	16.5	1.0	14	0.85 ± 0.23
Cluster	0.02	0.08	3	136.57 ± 79.34
	0.09	0.17	7	79.31 ± 30.3
	0.2	0.25	8	39.87 ± 14.4
	0.35	0.33	12	33.51 ± 9.92
	0.55	0.42	15	26.63 ± 7.1
	0.79	0.5	19	23.33 ± 5.55
	1.07	0.58	23	20.65 ± 4.48
	1.4	0.67	25	17.04 ± 3.58
	1.77	0.75	29	15.55 ± 3.05
	2.18	0.83	32	13.81 ± 2.59
	2.64	0.92	32	11.27 ± 2.14

Columns (3) represent the distance from the centre of the cluster.

profiles that fell within the windows were added and used to calculate new densities for the window region; this window was then moved across the density profile to produce a new density profile. The window is defined by a window size (between 1 and 10) that refers to the number of the ring areas from the original density profiles that are evaluated within the window. For example with a window size of 5 the rolling window starts with the sum of the first five ring areas in the original density profiles then the window is moved along the original density profile away from the start, adding in the next ring area and removing the last ring area in the window. The use of larger window sizes (a window size of 1 reproduces the original density profiles) provided larger ring areas and higher counts reducing the statistical uncertainties in the densities. However, larger window sizes do reduce the resolution of the resultant density profiles.

Windows of size 1 to 10 were passed over the data and the location of all the point to rise 2 and 3σ over the background density are presented in table B.5; the values highlighted show the clusters radius select for each object. Figure B.5 shows an example of the new density profiles generated for the HAeBe stars V361Cep and ILCep; the location of all the points that rise 2 and 3 sigma over the background density is given by a red full point and open black squares, respectively, when these occurred.

Table B.5: Windows of size 1 to 10 passed over sample of Known clusters and Herbig Ae/Be stars.

Star	CEREAL Class	Distance pc Window 1	Sigma Level	Count Num Stars	Distance pc Window 2	Sigma Level	Count Num Stars	Distance pc Window 3	Sigma Level	Count Num Stars	Distance pc Window 4	Sigma Level	Count Num Stars	Distance pc Window 5	Sigma Level	Count Num Stars
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
A0974-15	M															
AS310	M				**0.39 ± 0.04**	**2.32**	**6**	0.43 ± 0.09	2.22	6	0.47 ± 0.13	2.34	7	0.52 ± 0.17	2.22	7
AlphaPer	KC							**2.67 ± 0.33**	**3.47**	**20**	2.84 ± 0.50	3.47	23	3.34 ± 0.67	3.11	25
BD+30549	Y										0.21 ± 0.13	3.12	10	0.26 ± 0.17	3.10	10
Blanco1	KC				**0.97 ± 0.19**	**3.23**	**11**	2.32 ± 0.39	3.28	13	2.51 ± 0.58	3.28	14	2.71 ± 0.77	3.42	16
GSC5360-1033	M													**0.21 ± 0.14**	**2.09**	**5**
HD36982	M	0.68 ± 0.00	2.10	6	0.73 ± 0.04	2.38	9	0.77 ± 0.09	2.23	10	**0.64 ± 0.13**	**3.04**	**15**	0.60 ± 0.17	3.21	17
HD37371	M															
HD46060	M				1.32 ± 0.04	2.04	8	0.85 ± 0.09	2.29	9	0.90 ± 0.13	2.42	11	1.19 ± 0.17	2.01	12
HUCMa	M	1.05 ± 0.00	2.64	11	1.09 ± 0.04	2.31	13	1.13 ± 0.08	2.18	16	1.17 ± 0.12	2.22	20	1.13 ± 0.16	2.10	22
Hen3-949	Y										**0.89 ± 0.13**	**3.04**	**13**	0.94 ± 0.17	3.18	15
IC2602	KC				1.12 ± 0.12	2.18	6	2.23 ± 0.25	2.09	8	3.35 ± 0.37	2.04	11	**1.49 ± 0.50**	**3.14**	**14**
ILCep	Y							0.42 ± 0.08	3.19	12	**0.63 ± 0.13**	**3.07**	**13**	0.75 ± 0.17	3.11	15
LKHa338	M	0.92 ± 0.00	2.08	6	1.13 ± 0.04	2.24	9	1.09 ± 0.08	2.52	12	**1.05 ± 0.13**	**3.26**	**18**	1.01 ± 0.17	3.35	20
LkHa324	Y										**0.47 ± 0.13**	**3.14**	**12**	0.51 ± 0.17	3.02	12
MWC137	M							1.03 ± 0.08	2.27	9	1.07 ± 0.12	2.21	10	1.10 ± 0.16	2.52	13
NGC6475	KC	2.61 ± 0.00	2.55	8	**2.80 ± 0.19**	**3.13**	**13**	2.99 ± 0.37	3.00	14	2.80 ± 0.56	3.14	16	3.36 ± 0.75	3.16	19
V361Cep	Y				**0.29 ± 0.04**	**3.19**	**11**	0.33 ± 0.08	3.25	12	0.37 ± 0.12	3.13	12	0.41 ± 0.16	3.31	14
V373Cep	Y							**0.29 ± 0.07**	**3.13**	**11**	0.33 ± 0.11	3.04	11	0.36 ± 0.15	3.09	12

- Column 2: Y stars classified as yes and M stars classified as maybe, both by CEREAL; KC = KnownCluster.

- **Bold values** represent the cluster radius used for the analysis in this work.

Table B.6: Continued table B.5

Star	CEREAL Class	Distance pc Window 6	Sigma Level	Count Num Stars	Distance pc Window 7	Sigma Level	Count Num Stars	Distance pc Window 8	Sigma Level	Count Num Stars	Distance pc Window 9	Sigma Level	Count Num Stars	Distance pc Window 10	Sigma Level	Count Num Stars
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
A0974-15	M				**0.34 ± 0.26**	**2.27**	**7**	0.39 ± 0.30	2.15	7	0.43 ± 0.34	2.02	7	0.47 ± 0.39	2.10	8
AS310	M	0.56 ± 0.22	2.09	7	0.60 ± 0.26	2.17	8	0.65 ± 0.30	2.02	8	0.52 ± 0.34	2.32	3	0.56 ± 0.39	2.18	9
AlphaPer	KC	3.51 ± 0.83	3.43	31	3.67 ± 1.00	2.95	31	3.51 ± 1.17	3.59	38	4.01 ± 1.34	3.19	41	4.17 ± 1.50	3.04	44
BD+30549	Y	**0.73 ± 0.21**	**3.09**	**11**	0.86 ± 0.26	3.01	11	0.81 ± 0.30	3.30	13	0.86 ± 0.34	3.24	13	0.90 ± 0.39	3.18	13
Blanco1	KC	2.90 ± 0.97	3.54	18	3.09 ± 1.16	3.63	20	3.29 ± 1.35	3.45	20	3.48 ± 1.55	3.39	21	3.68 ± 1.74	3.18	21
GSC5360-1033	M	0.25 ± 0.18	2.03	5												
HD36982	M	0.73 ± 0.21	3.23	20	0.94 ± 0.26	3.03	23	0.90 ± 0.30	3.29	26	1.03 ± 0.34	3.16	29	1.07 ± 0.38	3.08	31
HD37371	M													**0.67 ± 0.36**	**2.15**	**8**
HD46060	M	1.15 ± 0.21	2.19	14	1.11 ± 0.26	2.54	17	**1.07 ± 0.30**	**3.02**	**21**	0.60 ± 0.34	3.07	17	0.64 ± 0.38	3.16	19
HUCMa	M	1.09 ± 0.20	2.19	25	**0.81 ± 0.24**	**3.16**	**29**	0.77 ± 0.28	3.24	31	0.81 ± 0.32	3.07	33	0.85 ± 0.36	3.00	36
Hen3-949	Y	0.98 ± 0.21	3.15	16	1.02 ± 0.26	3.12	17	1.07 ± 0.30	3.08	18	1.11 ± 0.34	3.03	19	0.98 ± 0.38	3.06	19
IC2602	KC	1.61 ± 0.62	3.27	16	1.74 ± 0.74	3.51	19	2.61 ± 0.87	3.16	21	2.73 ± 0.99	3.19	23	2.85 ± 1.12	3.07	24
ILCep	Y	0.79 ± 0.21	3.06	16	0.84 ± 0.25	3.01	17	0.79 ± 0.29	3.21	19	0.92 ± 0.33	3.01	20	0.88 ± 0.38	3.21	22
LKHa338	M	1.05 ± 0.21	3.07	20	1.09 ± 0.25	3.06	22	1.13 ± 0.29	3.05	24	1.17 ± 0.34	3.02	26	1.13 ± 0.38	3.28	29
LkHa324	Y	0.55 ± 0.21	3.04	13	0.60 ± 0.26	3.05	14	0.72 ± 0.30	3.10	16	0.77 ± 0.34	3.08	17	0.72 ± 0.38	3.32	19
MWC137	M	1.14 ± 0.20	2.25	13	1.18 ± 0.24	2.52	16	1.22 ± 0.28	2.43	17	1.26 ± 0.32	2.15	17	**0.75 ± 0.36**	**3.16**	**19**
NGC6475	KC	3.55 ± 0.93	3.14	21	3.36 ± 1.12	3.27	23	3.92 ± 1.31	3.04	25	3.73 ± 1.49	3.18	27	3.55 ± 1.68	3.46	30
V361Cep	Y	0.45 ± 0.21	3.32	15	0.49 ± 0.25	3.31	16	0.54 ± 0.29	3.43	18	0.58 ± 0.33	3.26	18	0.62 ± 0.37	3.21	19
V373Cep	Y	0.33 ± 0.18	3.22	13	0.36 ± 0.22	3.10	13	0.47 ± 0.25	3.00	14	0.51 ± 0.29	3.01	15	0.47 ± 0.33	3.14	16

- Column 2: Y stars classified as yes and M stars classified as maybe, both by CEREAL; KC = KnownCluster

- **Bold values** represent the cluster radius used for the analysis in this work.

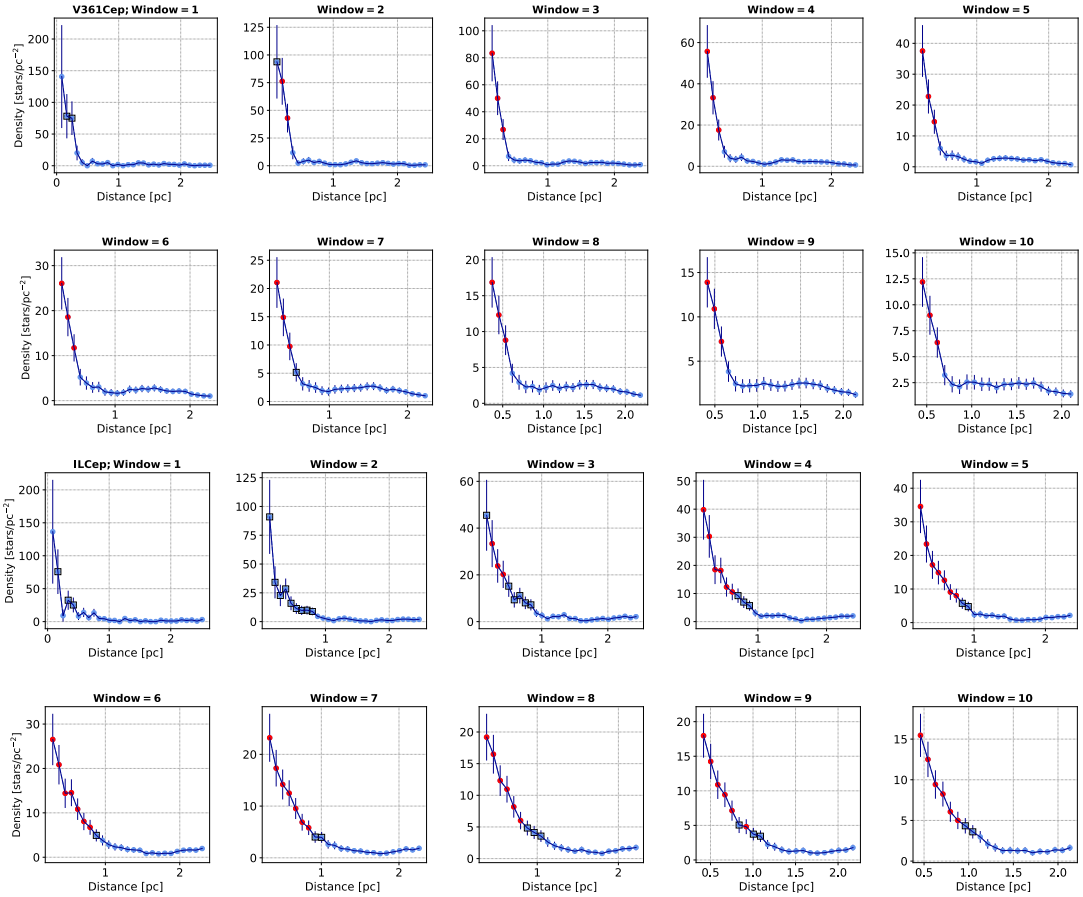


Figure B.5: Density profile created with different window of sizes for the Herbig Ae/Be stars V361Cep (top panel) and ILCep (bottom panel). Each panel show the density profile created with the different window of sizes from 1 to 10. The red full points and open black squares show the location of the all the points that rise 3 and 2σ over the background, respectively.

For the stars found as yes in general for window sizes over 4 the new density profile had regions that rose 3σ over the background level. For the stars found as maybe the majority had regions that rose 3σ over the background level for window sizes greater than 7. Although for some stars found as maybe (e.g HD37371) even with window sizes of 10 there were no points found that rose 3σ over the background density with this method. For most of the HAeBe stars evaluated here the location of the first point to rise 3σ above the background was dependent on the window size used.

Evaluation of the window size dependence of the cluster radii assigned here was carried out in the same manner as was described in section §2.2.1 for the known clusters.

Where the change in cluster radius position with window size did not lead to a significant change in the cluster radius within their combined errors, the radius assigned from the smallest window size to generate points that rose 3σ over the background was used to assign the cluster radius. When step changes in the cluster radius with increased window size led to a change in the cluster radius that was greater than the combined errors (see figure B.6) the cluster radius was assigned using the smallest window size after the change to a new consistent cluster radius had been reached. For example, for Herbig Ae/Be stars BD+3549 for window sizes 4 and 5 find a small cluster radius associated with the initial region closest to the central star with larger window sizes (6 and larger) finding a region of density at a larger region that also then meets the criteria of being 3σ over the background. Therefore, for BD+3549 the window size of 6 was chosen to assign the cluster radius.

The cluster radii assigned using this rolling window method vary greatly in size and range between 0.2pc and ∼3pc. The lower limit for the cluster radii found here agree with previous values from the literature that have been used to study clusters around HAeBe stars by Hillenbrand *et al.* (1995, 0.15pc) and Testi *et al.* (1997, 1998, 1999, 0.2pc). In these works the values for the cluster radius were chosen such that they were considered

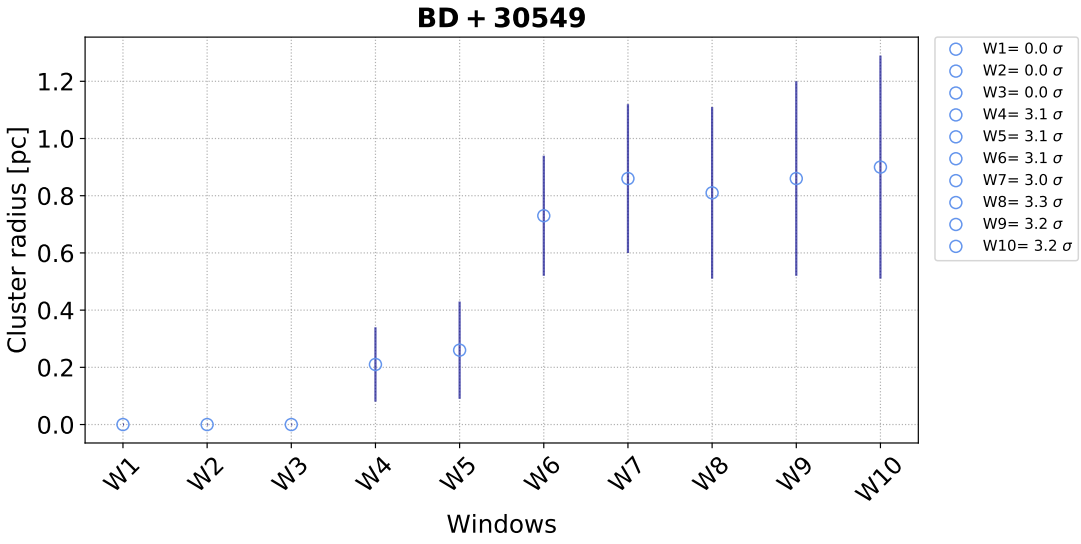


Figure B.6: Cluster radius assigned for the different window sizes for the Herbig Ae/Be stars BD+30549. The figure show the different window sizes from 1 to 10 versus the cluster radius find for each window size.

to be large enough to capture all the important objects around the main star allowing the study of both poor and rich clusters. For the known clusters studied here, the upper limit of the cluster radius found here were comparable to the values found by Gomez *et al.* (1993) which found stars within 0.5 - 1pc (Testi *et al.*, 1997, and their references). This range of cluster radii should allow for the inclusion of most of the central objects within the target clusters.

It is important to consider the impact that the size of the cluster radius values will have on the subsequent calculations, where a larger cluster radius is used, it means that more objects will be included in the clusters than would have been included had a radius as the ones reported in the literature been chosen. The inclusion of a larger number of objects in the clusters will not necessarily lead to an increase in the assigned stellar densities. This is because increasing the cluster radius increases the cluster volume by the cube of this, for example increasing the cluster radius by a factor of 2.5 increases the volume by a factor of 15.6. Therefore, it is likely that despite additional objects being captured using the larger radii found in this analysis, the stellar density assigned to the cluster here will be reduced (chapter §5). Table B.7 shows the number of stars found for the sample of HAeBe stars using the cluster radius assigned in this work and those used by Hillenbrand *et al.* (1995, 0.15pc) and Testi *et al.* (1997, 1998, 1999, 0.2pc).

CEREAL also re-analysed the sample of the Herbig Ae/Be stars that were found to have a sufficiently high density compared to the density of the background and compared these with the original results found previously for the same stars. The classification made by CEREAL assigned a flag of yes (Y) or maybe (M) if the star was considered to be in a cluster or potentially in a cluster, respectively. The comparison show that the main difference between both methods was in the number of stars found by each analysis. In this comparison, CEREAL originally found 14 stars classified as yes and 1 star as maybe and the quantitative analysis found 6 stars classified as yes and 9 stars as maybe.

Table B.7: Number of stars found by each cluster radio

Number of Stars		This Work	Testi *et al.* (1997)	Hillenbrand *et al.* (1995)
Cluster Radio	This Work		0.2 pc	0.15 pc
A0974-15	0.344	3	2	2
AS310	0.387	8	3	3
BD+30549	0.729	15	6	5
GSC5360-1033	0.213	4	4	2
HD36982	0.641	18	3	2
HD37371	0.672	5	1	1
HD46060	1.067	24	5	4
HUCMa	0.808	21	1	1
Hen3-949	0.895	16	1	1
ILCep	0.627	27	9	6
LKHa338	1.048	27	3	3
LkHa324	0.468	15	5	4
MWC137	0.75	11	1	0
V361Cep	0.288	16	8	4
V373Cep	0.291	4	1	0

Appendix C

Stellar parameters of the Herbig Ae/Be stars

This appendix contains the stellar parameters for the sample of 269 HAeBe stars introduced in section 3.1.

Tables C.1 to C.11 contain the following parameters of the HAeBe stars: right ascension, declination, spectral type reported by Chen *et al.* (2016, flag C) and Vioque *et al.* (2018, flag V), parallax, proper motions in RA and DEC, G magnitude, B_p-R_p colour, distance reported by Vioque *et al.* (2018), flag showing whether it was classified by The *et al.* (1994) as a HAeBe star, and flags showing whether the object was in the Chen *et al.* (2016, C) or Vioque *et al.* (2018, V) catalogues.

The stars with distance values reported in this table were those that had data from Gaia DR2. These distances were not obtained by inverting the parallax since this conversion is not trivial (Bailer-Jones, 2015; Vioque *et al.*, 2018).

Instead, Vioque *et al.* (2018) followed the method of Luri *et al.* (2018) on how to manage the parallax from Gaia, and estimated the distance by applying a simpler exponentially decreasing prior. For more information about how to estimate the distance using Gaia DR2 data please see Vioque *et al.* (2018, references therein).

In addition, tables C.12 to C.17 show more information about the sample of HAeBe stars, such as their luminosity class which was obtained from SIMBAD (flag S), Fairlamb *et al.* (2015, flag F), Wichittanakom *et al.* (2020, flag W), their mass and age reported by

Vioque *et al.* (2018) and the search radius angle used to select the data from the Gaia archive, described in § 2.1.1.

Vioque *et al.* (2018) calculated the mass and age of the HAeBE stars by assigning the two closest isochrone points in a HRD. The value was then estimated by assigning the average value weighted by the distance to these points. This used hundred PARSEC isochrones with solar metalicity (Bressan *et al.*, 2012; Marigo *et al.*, 2017) from 0.01 to 20 Myr. Additionally, the uncertainties of the mass and age were derived from the error bars in their HRD, keeping a minimum error of 5% (Vioque *et al.*, 2018).

Table C.1: Stars parameters of the Herbig Ae/Be stars

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
A0974-15	05:06:55.50	-03:21:13.0	B3	2.51 ± 0.04	1.05 ± 0.07	-1.16 ± 0.05	12.7	1.67	$397.6^{+8.9}_{-9.6}$		CV
ABAur	04:55:45.90	+30:33:04.0	A0	6.14 ± 0.06	3.93 ± 0.1	-24.11 ± 0.07	7.08	0.28	$162.9^{+2.4}_{-2.6}$	Yes	CV
AKSco	16:54:44.80	-36:53:19.0	F5	7.11 ± 0.06	-8.4 ± 0.12	-29.27 ± 0.08	9.28	0.95	$140.6^{+2.0}_{-2.1}$		CV
AS310	18:33:21.20	-04:58:06.0	B1	0.39 ± 0.05	-0.18 ± 0.07	-2.87 ± 0.07	12.39	1.45	$2108.4^{+235.6}_{-354.5}$	Yes	CV
AS470	21:36:14.20	+57:21:31.0	A0	0.14 ± 0.03	-2.22 ± 0.06	-1.25 ± 0.05	12.17	1.35	$4039.6^{+439.7}_{-616.9}$		CV
AS477	21:52:34.10	+47:13:44.0	A0	1.29 ± 0.03	-3.16 ± 0.05	-2.92 ± 0.04	10.0	0.68	$773.0^{+26.6}_{-30.2}$	Yes	CV
BD+30549	03:29:19.80	+31:24:57.0	B8	3.38 ± 0.08	6.82 ± 0.11	-9.45 ± 0.08	10.33	0.97	$295.4^{+11.1}_{-12.8}$		CV
BD+413731	20:24:15.71	+42:18:01.3	B3	0.99 ± 0.04	-2.53 ± 0.06	-5.88 ± 0.05	9.84	0.17	$1002.8^{+52.7}_{-64.4}$		V
BFOri	05:37:13.30	-06:35:01.0	A7	2.57 ± 0.05	0.88 ± 0.1	0.46 ± 0.1	10.36	0.56	$388.8^{+12.3}_{-13.8}$	Yes	CV
BHCep	22:01:42.90	+69:44:36.0	F5	2.98 ± 0.02	9.65 ± 0.03	-1.6 ± 0.03	11.17	0.95	$335.1^{+3.3}_{-3.4}$	Yes	CV
BOCep	22:16:54.10	+70:03:45.0	F4	2.67 ± 0.02	5.04 ± 0.04	3.28 ± 0.04	11.4	0.84	$374.5^{+4.5}_{-4.7}$		CV
BPPsc	23:22:24.70	-02:13:41.0	G9	2.79 ± 0.39	45.4 ± 0.76	-24.69 ± 0.58	11.28	1.36	$348.9^{+54.5}_{-114.5}$		CV
COOri	05:27:38.30	+11:25:39.0	G5	2.47 ± 0.04	2.23 ± 0.1	-2.21 ± 0.07	11.61	1.59	$404.0^{+10.7}_{-11.8}$		CV
CPM25	06:23:56.30	+14:30:28.0	B2	0.29 ± 0.07	-0.9 ± 0.11	-0.97 ± 0.09	15.1	1.65	$2128.9^{+293.6}_{-467.1}$		CV
CQTau	05:35:58.50	+24:44:54.0	F5	6.13 ± 0.08	2.56 ± 0.13	-26.03 ± 0.11	10.45	1.14	$163.1^{+3.5}_{-3.7}$	Yes	CV
DGCir	15:03:23.80	-63:22:58.0	B5	1.19 ± 0.04	-2.26 ± 0.05	-3.75 ± 0.07	12.71	1.98	$832.9^{+43.1}_{-52.5}$	Yes	CV
DKCha	12:53:17.20	-77:07:11.0	A0	4.1 ± 0.37	-18.38 ± 0.6	-3.58 ± 0.53	16.87	4.09	$242.9^{+28.2}_{-47.3}$	Yes	CV
DWCMa	07:19:35.90	-17:39:18.0	B2	0.28 ± 0.04	-1.28 ± 0.06	1.84 ± 0.07	12.92	2.32	$2552.6^{+306.0}_{-460.7}$		CV
GSC1829-0331	04:30:50.29	+23:00:08.5	F1	8.05 ± 0.16	11.85 ± 0.33	-15.68 ± 0.24	14.99	3.75			C
GSC1876-0892	06:07:15.40	+29:57:55.0	B2	0.15 ± 0.05	-0.33 ± 0.09	-0.83 ± 0.07	13.78	2.1	$3001.1^{+388.5}_{-579.6}$		CV
GSC3975-0579	21:38:08.50	+57:26:48.0	A2	1.06 ± 0.03	-2.58 ± 0.06	-4.74 ± 0.05	11.49	0.51	$942.2^{+39.5}_{-46.1}$		CV
GSC4805-1306	06:58:44.36	-03:41:10.0	O9				15.02	1.92			C
GSC4823-0146	07:08:38.80	-04:19:05.0	B0	0.12 ± 0.05	-1.05 ± 0.08	-0.17 ± 0.07	12.03	1.06	$2887.9^{+397.1}_{-599.3}$		CV
GSC5360-1033	05:57:49.50	-14:05:34.0	B5	1.65 ± 0.03	-2.36 ± 0.05	1.66 ± 0.06	13.82	1.61	$605.2^{+19.2}_{-21.6}$		CV
GSC5379-0359	06:49:58.60	-07:38:52.0	B9	0.75 ± 0.02	-1.22 ± 0.04	0.42 ± 0.03	13.24	1.12	$1315.6^{+53.4}_{-62.1}$		CV
GSC5987-1399	07:31:48.80	-19:27:34.0	B1		-7.4 ± 3.8	6.1 ± 4.0	13.3	1.05			C

Table C.2: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
GSC5988-2257	07:41:41.00	-20:00:13.0	B3	-4.66 ± 0.86	-5.14 ± 1.13	3.11 ± 1.11	16.0	1.72	$984.4^{+267.1}_{-522.2}$		CV
GSC5990-0021	06:48:41.70	-16:48:06.0	B9	0.88 ± 0.02	-3.44 ± 0.03	5.08 ± 0.03	13.12	0.6	$1130.0^{+37.2}_{-41.9}$		CV
GSC6542-2339	07:24:36.98	-24:34:47.4	O9	1.12 ± 0.11	-2.75 ± 0.16	3.71 ± 0.19	14.58	2.31	$846.5^{+97.5}_{-157.1}$		CV
GSC6546-3156	07:24:17.60	-26:16:05.0	A0	0.7 ± 0.02	-2.08 ± 0.03	2.14 ± 0.04	14.38	1.13	$1408.8^{+68.4}_{-82.1}$		CV
GSC8143-1225	07:59:11.60	-50:22:47.0	F3	2.59 ± 0.02	-5.13 ± 0.04	8.78 ± 0.05	12.83	1.27	$386.3^{+5.1}_{-5.3}$		CV
GSC8581-2002	08:44:23.60	-59:56:58.0	A5	1.79 ± 0.02	-8.11 ± 0.05	6.5 ± 0.05	10.72	0.38	$557.6^{+11.5}_{-12.3}$		CV
GSC8645-1401	12:17:47.50	-59:43:59.0	F2	0.54 ± 0.03	-6.39 ± 0.04	0.28 ± 0.04	12.84	1.67	$1776.7^{+129.6}_{-171.2}$		CV
GSC8993-0397	13:03:21.50	-62:13:26.0	B3	0.27 ± 0.03	-6.98 ± 0.04	-1.0 ± 0.04	12.74	0.79	$2953.4^{+301.5}_{-430.7}$		CV
GSC8994-3902	13:19:04.00	-62:34:10.0	B2	0.35 ± 0.04	-4.3 ± 0.04	-1.82 ± 0.04	10.68	0.45	$2393.7^{+250.6}_{-366.1}$	Yes	CV
GUCMa	07:01:49.51	-11:18:03.3	B2	2.25 ± 2.04	-5.08 ± 2.23	4.59 ± 1.93	6.77	0.19			C
HBC1	00:07:02.60	+65:38:38.0	A5	0.16 ± 0.52	-7.6 ± 0.89	-10.82 ± 0.64	15.98	2.74	$760.3^{+191.9}_{-435.7}$	Yes	CV
HBC217	06:40:42.20	+09:33:37.0	G0	1.43 ± 0.04	-1.57 ± 0.07	-3.68 ± 0.06	11.85	0.79	$695.6^{+29.6}_{-34.7}$		CV
HBC222	06:40:51.20	+09:44:46.0	F8	1.41 ± 0.04	-2.47 ± 0.06	-3.44 ± 0.06	11.89	0.82	$706.3^{+26.8}_{-30.8}$		CV
HBC324	00:07:30.70	+65:39:52.0	A5	-0.27 ± 0.47	1.56 ± 0.74	-6.59 ± 0.6	15.35	1.47	$895.1^{+222.2}_{-468.5}$	Yes	CV
HBC334	02:16:30.71	+55:23:00.1	B3	0.53 ± 0.03	-0.81 ± 0.06	-0.93 ± 0.07	14.53	1.3	$1774.4^{+137.4}_{-184.7}$	Yes	CV
HBC442	05:34:14.20	-05:36:54.0	F8	2.59 ± 0.04	0.61 ± 0.06	-0.28 ± 0.07	10.07	0.79	$385.7^{+9.1}_{-10.0}$		CV
HBC694	20:24:29.54	+42:14:02.0	A5	0.9 ± 0.43	-1.77 ± 0.76	-7.11 ± 0.62	17.22	1.64	$673.9^{+153.5}_{-366.9}$	Yes	V
HBC7	00:43:25.34	+61:38:23.3	B2	0.34 ± 0.02	-1.55 ± 0.03	-1.4 ± 0.03	13.02	1.8	$2757.3^{+194.4}_{-252.4}$		V
HBC705	20:51:02.70	+43:49:31.0	B0	0.46 ± 0.03	-0.8 ± 0.04	-2.63 ± 0.04	13.48	2.23	$2065.2^{+155.6}_{-207.0}$	Yes	CV
HBC717	20:52:06.10	+44:17:16.0	F6	0.49 ± 0.12	-3.7 ± 0.21	-2.86 ± 0.21	12.9	1.87	$1394.9^{+220.9}_{-390.9}$		CV
HD100453	11:33:05.50	-54:19:29.0	A9	9.6 ± 0.04	-36.93 ± 0.06	-6.08 ± 0.05	7.72	0.42	$104.2^{+0.7}_{-0.7}$		CV
HD100546	11:33:25.30	-70:11:41.0	B9	9.09 ± 0.05	-38.59 ± 0.08	-0.24 ± 0.08	6.66	0.06	$110.0^{+1.0}_{-1.0}$	Yes	CV
HD101412	11:39:44.40	-60:10:28.0	A0	2.43 ± 0.03	-3.77 ± 0.04	-4.14 ± 0.05	9.21	0.3	$411.3^{+7.6}_{-8.1}$	Yes	CV
HD104237	12:00:04.90	-78:11:35.0	A4	9.23 ± 0.06	-39.31 ± 0.11	-6.21 ± 0.08	6.54	0.39	$108.4^{+1.1}_{-1.1}$	Yes	CV
HD114981	13:14:40.70	-38:39:06.0	B5	1.41 ± 0.06	-12.42 ± 0.09	8.51 ± 0.09	7.08	-0.03	$704.8^{+44.4}_{-56.7}$		CV
HD130437	14:50:50.20	-60:17:10.0	B8	0.57 ± 0.04	-3.79 ± 0.04	-2.29 ± 0.06	9.54	1.19	$1653.2^{+131.8}_{-179.1}$	Yes	CV

Table C.3: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
HD132947	15:04:56.10	-63:07:53.0	A0	2.62 ± 0.06	-6.92 ± 0.07	-9.66 ± 0.1	8.89	0.09	$381.6^{+12.8}_{-14.5}$	Yes	CV
HD135344	15:15:48.42	-37:09:16.4	A0	7.37 ± 0.08	-19.11 ± 0.11	-23.14 ± 0.08	8.51	0.73	$142.1^{+2.9}_{-3.1}$		V
HD135344B	15:15:48.40	-37:09:16.0	F8	7.37 ± 0.08	-19.11 ± 0.11	-23.14 ± 0.08	8.51	0.73	$135.8^{+2.3}_{-2.4}$		CV
HD139614	15:40:46.40	-42:29:54.0	A8	7.42 ± 0.05	-18.0 ± 0.12	-25.34 ± 0.08	8.21	0.35	$134.7^{+1.6}_{-1.6}$		CV
HD141569	15:49:57.70	-03:55:17.0	A0	9.04 ± 0.04	-17.63 ± 0.08	-18.95 ± 0.07	7.09	0.16	$110.6^{+0.9}_{-0.9}$	Yes	CV
HD141926	15:54:21.80	-55:19:44.0	B2	0.72 ± 0.04	-3.11 ± 0.06	-3.29 ± 0.05	8.68	0.92	$1338.2^{+108.3}_{-148.4}$		CV
HD142527	15:56:41.90	-42:19:24.0	F6	6.36 ± 0.05	-11.3 ± 0.09	-26.34 ± 0.06	8.09	1.01	$157.3^{+1.9}_{-2.0}$		CV
HD142666	15:56:40.00	-22:01:40.0	A8	6.74 ± 0.05	-13.05 ± 0.11	-22.16 ± 0.06	8.57	0.77	$148.3^{+1.9}_{-2.0}$	Yes	CV
HD143006	15:58:36.90	-22:57:16.0	G6	6.02 ± 0.15	-10.88 ± 0.11	-20.96 ± 0.06	9.89	1.1	$166.1^{+6.2}_{-7.1}$		CV
HD144432	16:06:57.90	-27:43:10.0	A8	6.44 ± 0.06	-12.59 ± 0.1	-24.4 ± 0.05	8.1	0.55	$155.4^{+2.2}_{-2.4}$	Yes	CV
HD149914	16:38:28.60	-18:13:14.0	B9	6.3 ± 0.08	-13.91 ± 0.16	-20.38 ± 0.06	6.62	0.51	$158.8^{+3.2}_{-3.5}$		CV
HD150193	16:40:17.90	-23:53:45.0	A0	6.63 ± 0.07	-4.91 ± 0.13	-19.09 ± 0.07	8.65	0.92	$150.8^{+2.5}_{-2.7}$	Yes	CV
HD155448	17:12:58.80	-32:14:34.0	B9	1.02 ± 0.06	1.99 ± 0.1	-1.61 ± 0.07	8.72	0.12	$953.9^{+73.8}_{-100.2}$		CV
HD158643	17:31:24.95	-23:57:45.9	A0	8.15 ± 0.3	1.6 ± 0.49	-24.45 ± 0.37	4.71	0.06	$122.8^{+6.7}_{-8.2}$		CV
HD163296	17:56:21.30	-21:57:22.0	A0	9.85 ± 0.11	-7.61 ± 0.15	-39.42 ± 0.12	6.81	0.2	$101.5^{+1.9}_{-2.0}$	Yes	CV
HD169142	18:24:29.80	-29:46:50.0	B9	8.77 ± 0.06	-2.32 ± 0.11	-37.8 ± 0.09	8.05	0.42	$114.0^{+1.3}_{-1.4}$		CV
HD17081	02:44:07.34	-13:51:31.6	B7	9.38 ± 0.37	-10.61 ± 0.74	-25.41 ± 0.74	4.17	-0.16	$106.7^{+6.3}_{-7.9}$		CV
HD174571	18:50:47.20	+08:42:10.0	B2	0.88 ± 0.06	-1.17 ± 0.1	-1.64 ± 0.09	8.73	0.79	$1095.0^{+94.9}_{-133.8}$		CV
HD176386	19:01:38.90	-36:53:27.0	A0	6.28 ± 0.06	4.16 ± 0.11	-28.5 ± 0.1	7.27	0.27	$159.2^{+2.5}_{-2.6}$		CV
HD179218	19:11:11.30	+15:47:15.0	A0	3.76 ± 0.05	5.18 ± 0.07	-20.82 ± 0.06	7.35	0.15	$266.0^{+5.2}_{-5.6}$	Yes	CV
HD199603	20:58:41.78	-14:28:59.5	A9	11.2 ± 0.1	-54.99 ± 0.16	-18.43 ± 0.12	5.9	0.36	$89.3^{+1.3}_{-1.4}$		CV
HD200775	21:01:36.90	+68:09:48.0	B2	2.77 ± 0.04	8.34 ± 0.08	-1.57 ± 0.08	7.16	0.73	$360.8^{+9.1}_{-10.0}$	Yes	CV
HD203024	21:16:03.05	+68:54:52.1	A5		10.5 ± 1.4	-2.2 ± 1.4	8.82	0.27			C
HD235495	21:21:27.50	+50:59:48.0	A0	1.91 ± 0.03	1.95 ± 0.06	-1.73 ± 0.07	9.59	0.44	$523.9^{+13.9}_{-15.3}$		CV
HD244314	05:30:19.00	+11:20:20.0	A1	2.31 ± 0.06	2.11 ± 0.1	-2.13 ± 0.08	10.02	0.34	$431.7^{+16.8}_{-19.4}$		CV
HD244604	05:31:57.30	+11:17:41.0	A0	2.37 ± 0.06	0.73 ± 0.09	-1.66 ± 0.07	9.37	0.3	$420.6^{+16.6}_{-19.3}$	Yes	CV

Table C.4: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
HD245185	05:35:09.60	+10:01:51.0	A0	2.32 ± 0.11	0.34 ± 0.14	-1.93 ± 0.1	9.84	0.18	$428.6^{+28.6}_{-37.2}$	Yes	CV
HD245906	05:39:30.50	+26:19:55.0	A6	0.67 ± 0.48	0.97 ± 0.81	-3.57 ± 0.72	10.41	0.79	$687.5^{+165.8}_{-395.6}$		CV
HD249879	05:58:55.80	+16:39:57.0	B8	1.48 ± 0.08	-0.84 ± 0.13	-6.93 ± 0.11	10.63	0.06	$669.0^{+49.3}_{-65.9}$		CV
HD250550	06:02:00.00	+16:30:57.0	B9	1.4 ± 0.1	-0.59 ± 0.13	-5.15 ± 0.1	9.48	0.26	$697.1^{+64.3}_{-93.6}$	Yes	CV
HD259431	06:33:05.20	+10:19:20.0	B6	1.38 ± 0.05	-2.32 ± 0.08	-3.24 ± 0.08	8.68	0.69	$720.9^{+36.6}_{-44.4}$	Yes	CV
HD287823	05:24:08.00	+02:27:47.0	A0	2.78 ± 0.05	1.16 ± 0.09	-0.27 ± 0.07	9.69	0.24	$358.9^{+10.7}_{-12.0}$		CV
HD288012	05:33:04.80	+02:28:10.0	A0	2.52 ± 0.06	1.95 ± 0.1	0.33 ± 0.09	9.65	0.08	$395.8^{+13.5}_{-15.3}$		CV
HD290380	05:23:31.00	-01:04:24.0	F0	2.82 ± 0.05	1.27 ± 0.09	-0.52 ± 0.07	10.28	0.73	$354.3^{+9.5}_{-10.4}$		CV
HD290409	05:27:05.50	+00:25:08.0	B9	2.19 ± 0.08	0.64 ± 0.14	1.21 ± 0.12	9.95	0.17	$454.9^{+25.1}_{-31.0}$		CV
HD290500	05:29:48.10	-00:23:43.0	A2	2.28 ± 0.07	2.0 ± 0.1	0.91 ± 0.08	10.99	0.48	$438.3^{+19.9}_{-23.6}$		CV
HD290764	05:38:05.30	-01:15:22.0	F0	2.51 ± 0.06	-0.28 ± 0.1	-0.68 ± 0.09	9.76	0.5	$397.9^{+15.4}_{-17.8}$		CV
HD290770	05:37:02.40	-01:37:21.0	B9	2.5 ± 0.07	-0.57 ± 0.1	-0.72 ± 0.09	9.21	0.1	$399.1^{+17.8}_{-21.0}$		CV
HD305298	10:33:05.00	-60:19:51.0	O8	0.13 ± 0.03	-4.52 ± 0.05	2.02 ± 0.05	10.78	0.39	$4039.5^{+442.9}_{-628.2}$		CV
HD313571	18:01:07.20	-22:15:04.0	B3	0.62 ± 0.08	-0.05 ± 0.09	-2.39 ± 0.07	9.67	0.9	$1397.0^{+174.4}_{-282.9}$		V
HD31648	04:58:46.30	+29:50:37.0	A5	6.18 ± 0.08	4.79 ± 0.13	-25.35 ± 0.07	7.65	0.29	$161.8^{+3.2}_{-3.4}$	Yes	CV
HD319896	17:31:05.80	-35:08:29.0	B4	0.72 ± 0.06	0.26 ± 0.08	-0.51 ± 0.07	10.66	1.04	$1295.2^{+134.0}_{-201.7}$		CV
HD323771	17:34:04.60	-39:23:41.0	B5	0.91 ± 0.06	1.4 ± 0.16	-3.21 ± 0.11	10.98	0.61	$1067.9^{+86.3}_{-118.7}$		CV
HD34282	05:16:00.50	-09:48:35.0	A0	3.21 ± 0.05	0.31 ± 0.08	-0.99 ± 0.08	9.87	0.28	$311.5^{+7.2}_{-7.9}$	Yes	CV
HD344261	19:21:53.54	+21:31:50.8	F2	3.33 ± 0.04	24.12 ± 0.04	11.23 ± 0.05	10.52	0.72	$300.5^{+5.2}_{-5.6}$		V
HD34700	05:19:41.40	+05:38:43.0	G0	2.8 ± 0.05	0.69 ± 0.09	1.44 ± 0.06	9.01	0.76	$356.4^{+9.6}_{-10.6}$		CV
HD35187	05:24:01.20	+24:57:37.0	A2	6.13 ± 0.1	4.39 ± 0.3	-25.83 ± 0.19	8.4	0.35	$163.0^{+4.2}_{-4.6}$		CV
HD35929	05:27:42.80	-08:19:39.0	A5	2.58 ± 0.05	-1.28 ± 0.1	-4.9 ± 0.08	8.0	0.61	$387.4^{+12.0}_{-13.5}$	Yes	CV
HD36112	05:30:27.50	+25:19:57.0	A8	6.24 ± 0.07	3.59 ± 0.11	-26.75 ± 0.08	8.2	0.42	$160.3^{+2.8}_{-2.9}$	Yes	CV
HD36408	05:32:14.10	+17:03:29.0	B7	2.29 ± 0.1	-5.29 ± 0.18	-9.71 ± 0.13	6.05	0.03	$435.1^{+27.1}_{-34.5}$		CV
HD36917	05:34:47.00	-05:34:15.0	A0	2.1 ± 0.07	1.67 ± 0.13	-2.33 ± 0.11	7.94	0.39	$474.0^{+24.1}_{-29.3}$		CV
HD36982	05:35:09.84	-05:27:53.2	B1.5	2.45 ± 0.06	1.48 ± 0.12	1.61 ± 0.1	8.33	0.31	$408.0^{+16.3}_{-18.9}$		V

Table C.5: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
HD37357	05:37:47.10	-06:42:30.0	A0	1.27 ± 0.3	0.59 ± 0.6	0.2 ± 0.55	8.85	0.2	$649.6^{+123.6}_{-278.3}$	Yes	CV
HD37371	05:38:09.90	-00:11:02.0	B9	2.43 ± 0.06	-0.0 ± 0.09	-0.7 ± 0.08	8.77	0.11	$411.3^{+16.0}_{-18.5}$		CV
HD37411	05:38:14.51	-05:25:13.3	B9		-1.6 ± 1.4	-0.8 ± 1.4	9.79	0.28			C
HD37490	05:39:11.10	+04:07:17.0	B3	3.16 ± 0.36	0.04 ± 0.66	0.23 ± 0.58	4.44	-0.02	$312.0^{+42.6}_{-79.9}$	Yes	CV
HD37806	05:41:02.30	-02:43:01.0	B9	2.34 ± 0.06	4.37 ± 0.09	-1.32 ± 0.09	7.86	0.18	$427.6^{+16.4}_{-18.9}$	Yes	CV
HD38087	05:43:00.60	-02:18:45.0	B5	2.95 ± 0.1	-0.79 ± 0.14	-2.46 ± 0.14	8.21	0.32	$338.1^{+17.6}_{-21.5}$		CV
HD38120	05:43:11.90	-04:59:50.0	A0	2.46 ± 0.08	0.39 ± 0.1	-2.05 ± 0.1	9.03	0.13	$405.0^{+19.8}_{-23.9}$		CV
HD39014	05:44:46.40	-65:44:08.0	A7	22.68 ± 0.3	-28.83 ± 0.73	7.08 ± 0.74	4.23	0.33	$44.1^{+0.9}_{-1.0}$		CV
HD41511	06:04:59.10	-16:29:04.0	A2	4.5 ± 0.17	-5.95 ± 0.39	-3.71 ± 0.4	4.61	0.93	$222.2^{+12.6}_{-15.7}$		CV
HD45677	06:28:17.42	-13:03:11.1	B2	1.6 ± 0.06	1.32 ± 0.09	1.71 ± 0.09	7.27	0.22	$620.5^{+33.2}_{-40.7}$		V
HD46060	06:30:49.80	-09:39:15.0	B8	1.05 ± 0.06	-3.63 ± 0.1	0.08 ± 0.09	8.76	0.45	$932.9^{+70.8}_{-95.5}$		CV
HD50083	06:51:45.80	+05:05:04.0	B3	0.9 ± 0.05	-0.41 ± 0.11	-1.79 ± 0.1	6.82	0.21	$1089.8^{+77.1}_{-101.6}$		CV
HD50138	06:51:33.40	-06:57:59.0	B8	2.63 ± 0.06	-3.68 ± 0.1	3.86 ± 0.1	6.52	0.17	$379.9^{+14.3}_{-16.4}$		CV
HD53367	07:04:25.50	-10:27:16.0	B0	7.77 ± 0.79	7.19 ± 1.38	-7.04 ± 1.21	6.99	0.75	$129.7^{+16.7}_{-30.4}$	Yes	CV
HD56895B	07:18:31.80	-11:11:34.0	F2	6.05 ± 0.05	1.38 ± 0.07	5.05 ± 0.06	8.31	0.53	$165.3^{+2.1}_{-2.2}$		CV
HD58647	07:25:56.10	-14:10:44.0	B9	3.14 ± 0.04	-4.96 ± 0.06	-3.27 ± 0.05	6.76	0.15	$318.5^{+6.8}_{-7.4}$		CV
HD59319	07:28:36.80	-21:57:49.0	B8	1.49 ± 0.05	-3.82 ± 0.06	-0.4 ± 0.07	8.29	-0.08	$668.4^{+32.8}_{-39.5}$		V
HD68695	08:11:44.60	-44:05:09.0	A0	2.53 ± 0.04	-6.38 ± 0.06	8.82 ± 0.06	9.8	0.17	$395.7^{+9.1}_{-9.9}$		CV
HD72106	08:29:34.89	-38:36:21.0	A0	0.03 ± 0.83	-8.08 ± 1.39	5.12 ± 1.27	8.78	0.02	$597.2^{+169.2}_{-426.4}$		CV
HD76534	08:55:08.70	-43:28:00.0	B2	1.09 ± 0.04	-5.68 ± 0.06	5.61 ± 0.06	8.28	0.26	$910.6^{+45.5}_{-55.0}$	Yes	CV
HD85567	09:50:28.50	-60:58:03.0	B2	0.97 ± 0.03	-8.46 ± 0.05	5.77 ± 0.05	8.43	0.42	$1023.0^{+44.8}_{-52.8}$	Yes	CV
HD87403	10:02:51.40	-59:16:55.0	A1	0.47 ± 0.04	-6.88 ± 0.08	3.95 ± 0.08	9.22	0.13	$1905.5^{+192.5}_{-281.9}$		CV
HD87643	10:04:30.27	-58:39:52.0	B3	0.01 ± 0.12	-7.84 ± 0.19	8.54 ± 0.18	8.74	1.41	$2014.7^{+346.3}_{-569.0}$		V
HD94509	10:53:27.20	-58:25:24.0	B8	0.51 ± 0.03	-7.66 ± 0.06	2.56 ± 0.06	9.13	0.08	$1833.7^{+151.7}_{-208.1}$	Yes	CV
HD95881	11:01:57.60	-71:30:48.0	A0	0.84 ± 0.03	-8.99 ± 0.06	1.22 ± 0.06	8.18	0.33	$1168.3^{+65.9}_{-81.6}$	Yes	CV
HD96042	11:03:40.50	-59:25:59.0	B1	0.22 ± 0.03	-7.01 ± 0.05	2.18 ± 0.05	8.42	0.22	$3100.2^{+350.3}_{-509.7}$		CV

Table C.6: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
HD9672	01:34:37.88	-15:40:34.9	A1	17.52 ± 0.1	94.21 ± 0.19	-3.17 ± 0.11	5.59	0.07	$57.1^{+0.5}_{-0.5}$		V
HD97048	11:08:03.20	-77:39:17.0	A0	5.41 ± 0.04	-22.44 ± 0.06	1.3 ± 0.06	8.34	0.61	$184.8^{+2.1}_{-2.2}$	Yes	CV
HD98922	11:22:31.70	-53:22:11.0	B9	1.45 ± 0.03	-8.66 ± 0.05	0.82 ± 0.04	6.71	0.18	$688.8^{+24.6}_{-28.1}$	Yes	CV
HKOri	05:31:28.05	+12:09:10.1	A2		-2.2 ± 2.2	1.1 ± 2.1	11.29	0.84			C
HR5999	16:08:34.30	-39:06:19.0	A7	6.21 ± 0.07	-8.95 ± 0.15	-23.01 ± 0.08	7.87	0.82	$161.1^{+2.9}_{-3.1}$	Yes	CV
HTCMa	07:02:42.50	-11:26:12.0	A0	0.88 ± 0.04	-3.88 ± 0.07	1.41 ± 0.07	11.74	0.77	$1120.8^{+66.2}_{-82.9}$	Yes	CV
HUCMa	07:04:06.70	-11:26:08.0	B8	0.83 ± 0.04	-4.43 ± 0.07	1.58 ± 0.07	11.66	0.52	$1174.2^{+77.3}_{-99.6}$	Yes	CV
Hen2-80	12:22:23.20	-63:17:17.0	B5	0.71 ± 0.38	0.83 ± 0.71	1.12 ± 0.56	12.78	1.49	$753.5^{+169.7}_{-387.2}$		CV
Hen3-1121	15:58:09.62	-53:51:18.4	B0.5	0.31 ± 0.1	-2.72 ± 0.14	-2.46 ± 0.12	12.68	1.57	$1749.9^{+275.1}_{-464.6}$		V
Hen3-1121S	15:58:09.66	-53:51:35.3	B0	0.55 ± 0.17	-2.31 ± 0.27	-4.01 ± 0.24	11.08	1.21	$1149.4^{+205.6}_{-392.0}$		V
Hen3-1191	16:27:15.10	-48:39:26.0	B0	0.49 ± 0.07	5.62 ± 0.08	2.52 ± 0.06	13.08	1.66	$1661.5^{+211.7}_{-340.3}$	Yes	CV
Hen3-1475	17:45:14.18	-17:56:46.9	B7				12.51	1.78			C
Hen3-373	10:10:00.30	-57:02:07.0	B2	0.42 ± 0.06	-5.29 ± 0.11	2.91 ± 0.11	12.72	2.37	$1925.5^{+232.6}_{-361.6}$		CV
Hen3-823	12:48:42.40	-59:54:35.0	B3	0.8 ± 0.07	-6.97 ± 0.08	-0.59 ± 0.09	10.24	0.41	$1166.1^{+119.1}_{-179.0}$		V
Hen3-847	13:01:17.80	-48:53:19.0	B5	1.03 ± 0.23	-8.53 ± 0.35	5.41 ± 0.28	10.76	0.34	$784.8^{+140.2}_{-291.8}$	Yes	CV
Hen3-938	13:52:42.90	-63:32:49.0	O9	0.12 ± 0.03	-9.09 ± 0.04	-2.44 ± 0.05	12.37	2.12	$3851.6^{+449.3}_{-639.0}$		CV
Hen3-949	13:57:44.00	-39:58:47.0	B3	1.55 ± 0.04	-9.47 ± 0.06	-0.66 ± 0.06	10.23	0.47	$642.5^{+27.7}_{-32.5}$		CV
ILCep	22:53:15.60	+62:08:45.0	B2	1.24 ± 0.03	-0.76 ± 0.04	-2.01 ± 0.04	8.99	1.11	$805.2^{+27.3}_{-30.9}$	Yes	CV
IPPer	03:40:47.00	+32:31:54.0	A6	3.25 ± 0.08	6.73 ± 0.14	-9.71 ± 0.08	10.29	0.54	$307.5^{+11.3}_{-13.0}$	Yes	CV
KKOph	17:10:08.10	-27:15:19.0	A5	4.52 ± 0.14	1.89 ± 0.25	-20.44 ± 0.15	12.11	1.21	$221.1^{+10.4}_{-12.4}$	Yes	CV
LKHa260	18:19:09.40	-13:50:41.0	B6	0.77 ± 0.06	0.4 ± 0.07	-1.48 ± 0.06	14.38	1.42	$1234.5^{+112.2}_{-160.6}$		CV
LKHa338	06:10:47.13	-06:12:50.6	B9	1.12 ± 0.04	-3.09 ± 0.07	0.5 ± 0.07	14.18	2.01	$884.6^{+50.3}_{-62.5}$		CV
LkHa208	06:07:49.50	+18:39:26.0	A7	1.49 ± 0.2	-0.89 ± 0.37	-7.28 ± 0.33	11.37	0.74	$627.2^{+90.6}_{-170.2}$	Yes	CV
LkHa215	06:32:41.80	+10:09:34.0	B6	1.39 ± 0.04	-1.58 ± 0.09	-4.2 ± 0.07	10.51	1.15	$713.1^{+34.0}_{-40.8}$	Yes	CV
LkHa257	21:54:18.70	+47:12:09.0	B5	1.26 ± 0.02	-2.69 ± 0.03	-2.4 ± 0.03	13.06	1.07	$793.8^{+16.4}_{-17.6}$	Yes	CV
LkHa259	23:58:41.60	+66:26:13.0	A9	1.32 ± 0.03	0.61 ± 0.06	-1.87 ± 0.05	14.26	2.64	$755.5^{+29.2}_{-33.7}$	Yes	CV

Table C.7: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
LkHa324	21:03:54.20	+50:15:10.0	B9	1.65 ± 0.02	0.97 ± 0.05	-3.54 ± 0.05	12.22	1.62	$605.3^{+14.3}_{-15.5}$	Yes	CV
LkHa339	06:10:57.80	-06:14:40.0	A1	1.16 ± 0.03	-3.79 ± 0.04	0.39 ± 0.04	13.27	1.35	$857.1^{+29.5}_{-33.4}$	Yes	CV
MQCas	00:09:37.50	+58:13:10.0	A0	1.22 ± 0.09	8.04 ± 0.11	-2.86 ± 0.09	13.49	1.01	$797.2^{+75.3}_{-110.5}$		CV
MWC1021	20:29:26.90	+41:40:44.0	B0	0.51 ± 0.07	-3.12 ± 0.11	-4.25 ± 0.1	10.7	3.32	$1643.6^{+201.9}_{-320.5}$		CV
MWC1080	23:17:25.60	+60:50:43.0	B0	0.73 ± 0.09	-3.75 ± 0.16	-2.55 ± 0.16	10.87	2.19	$1203.0^{+154.5}_{-257.1}$	Yes	CV
MWC137	06:18:45.50	+15:16:52.0	B0	0.11 ± 0.05	-0.26 ± 0.09	-0.41 ± 0.07	11.17	1.86	$2907.4^{+401.1}_{-604.8}$	Yes	CV
MWC297	18:27:39.50	-03:49:52.0	B1	2.66 ± 0.09	1.45 ± 0.14	-8.32 ± 0.14	10.31	2.97	$374.9^{+18.0}_{-21.6}$	Yes	CV
MWC314	19:21:33.97	+14:52:56.8	B3	0.19 ± 0.04	-2.33 ± 0.06	-4.82 ± 0.06	8.94	2.17	$2977.0^{+367.7}_{-545.8}$		CV
MWC342	20:23:03.60	+39:29:50.0	B0	0.54 ± 0.02	-3.05 ± 0.04	-5.8 ± 0.05	9.72	1.88	$1808.3^{+111.5}_{-140.8}$		CV
MWC593	17:49:10.20	-24:14:21.0	B4	0.68 ± 0.06	1.79 ± 0.13	-0.52 ± 0.1	9.65	1.07	$1341.4^{+144.0}_{-219.9}$		V
MWC623	19:56:31.50	+31:06:20.0	B0	0.17 ± 0.04	-2.83 ± 0.05	-5.23 ± 0.06	9.93	1.99	$3279.8^{+388.5}_{-565.8}$		CV
MWC655	22:38:31.80	+55:50:05.0	B1	0.42 ± 0.03	-2.39 ± 0.05	-1.49 ± 0.05	9.19	0.7	$2170.0^{+182.0}_{-249.6}$		CV
MWC657	22:42:41.80	+60:24:00.0	B0	0.26 ± 0.02	-3.2 ± 0.04	-1.83 ± 0.04	11.85	1.99	$3164.2^{+286.7}_{-395.6}$		CV
MWC778	05:50:13.90	+23:52:17.7	B1				12.95	1.5			C
MWC878	17:24:44.70	-38:43:51.0	B1	0.49 ± 0.05	0.33 ± 0.08	-1.18 ± 0.06	10.24	1.22	$1773.8^{+197.8}_{-301.4}$		CV
MWC930	18:26:25.20	-07:13:18.0	B0	-0.16 ± 0.09	-2.95 ± 0.16	-6.63 ± 0.15	9.57	3.5	$2585.8^{+422.5}_{-652.0}$		CV
MWC953	18:43:28.40	-03:46:17.0	B2	0.49 ± 0.06	-0.69 ± 0.08	-1.78 ± 0.07	10.55	1.51	$1721.3^{+200.3}_{-310.6}$		V
NSV2968	06:26:53.90	-10:15:35.0	B0	0.94 ± 0.06	-3.26 ± 0.1	0.78 ± 0.09	13.93	2.13	$1026.3^{+88.1}_{-123.9}$		CV
NVOri	05:35:31.40	-05:33:09.0	F6	2.58 ± 0.06	1.55 ± 0.11	0.42 ± 0.09	9.69	0.61	$386.5^{+13.3}_{-15.0}$		CV
NXPup	07:19:28.30	-44:35:11.0	A0	-9.84 ± 0.65	-24.31 ± 1.38	6.76 ± 1.25	10.24	0.93	$1672.5^{+378.4}_{-625.3}$	Yes	CV
PDS002	01:17:43.50	-52:33:31.0	F3	2.41 ± 0.03	-21.33 ± 0.04	-2.39 ± 0.04	10.78	0.54	$414.6^{+7.3}_{-7.8}$		CV
PDS004	03:39:00.60	+29:41:46.0	A0	2.5 ± 0.05	3.29 ± 0.08	-4.85 ± 0.05	10.58	0.55	$399.0^{+12.4}_{-13.8}$		CV
PDS021	06:02:14.90	-10:00:59.0	B5	1.12 ± 0.04	-1.75 ± 0.06	1.79 ± 0.06	10.27	0.52	$886.8^{+43.8}_{-52.8}$		CV
PDS022	06:03:37.10	-14:53:03.0	A0	1.11 ± 0.1	-0.43 ± 0.17	1.6 ± 0.17	10.82	0.21	$854.6^{+92.6}_{-144.7}$		CV
PDS025	06:54:27.90	-25:02:16.0	A3	1.31 ± 0.02	-3.47 ± 0.02	3.63 ± 0.03	13.26	1.12	$763.5^{+17.6}_{-19.1}$		CV
PDS123	05:50:54.77	+20:14:47.7	B0	0.65 ± 0.03	0.62 ± 0.06	-0.96 ± 0.05	12.89	1.1	$1487.1^{+107.2}_{-141.5}$		CV

Table C.8: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
PDS124	06:06:58.50	-05:55:07.0	A0	1.16 ± 0.05	-1.4 ± 0.07	0.63 ± 0.09	12.34	0.8	$852.8^{+51.2}_{-64.5}$		CV
PDS126	06:13:37.30	-06:25:02.0	A7	1.17 ± 0.04	-3.77 ± 0.06	0.56 ± 0.06	11.65	0.75	$846.6^{+42.2}_{-51.0}$		CV
PDS129	06:31:03.60	+10:01:13.0	F5	1.45 ± 0.07	-2.31 ± 0.09	-4.9 ± 0.07	11.93	0.99	$683.8^{+46.4}_{-60.5}$		CV
PDS133	07:25:04.90	-25:45:49.0	B6	0.67 ± 0.02	-2.68 ± 0.03	3.14 ± 0.04	14.44	1.42	$1474.6^{+75.8}_{-91.9}$		CV
PDS134	07:32:26.60	-21:55:36.0	B6	0.33 ± 0.03	-2.0 ± 0.04	1.57 ± 0.05	11.93	0.63	$2551.9^{+255.2}_{-366.4}$		CV
PDS138	11:53:13.20	-62:05:21.0	B0	0.04 ± 0.03	-6.2 ± 0.05	1.97 ± 0.04	12.61	1.94	$4630.9^{+538.5}_{-745.4}$		CV
PDS144S	15:49:15.31	-26:00:55.1	A5	6.69 ± 0.12	-15.21 ± 0.18	-25.3 ± 0.11	13.16	0.88	$149.6^{+4.2}_{-4.6}$		CV
PDS211	06:10:17.30	+29:25:17.0	B9	0.92 ± 0.03	-0.07 ± 0.05	-2.14 ± 0.04	13.42	1.27	$1073.8^{+53.7}_{-64.9}$		CV
PDS229	06:55:40.00	-03:09:50.0	A0	-0.52 ± 0.53	-2.43 ± 0.92	0.53 ± 0.94	13.32	0.99	$879.9^{+224.5}_{-474.2}$		CV
PDS24	06:48:41.68	-16:48:05.5	B9	0.88 ± 0.02	-3.44 ± 0.03	5.08 ± 0.04	13.12	0.6			V
PDS277	08:23:11.80	-39:07:01.0	F3	2.89 ± 0.04	-7.04 ± 0.05	10.14 ± 0.05	9.87	0.58	$345.8^{+6.8}_{-7.3}$		CV
PDS286	09:06:00.00	-47:18:58.0	B0	0.52 ± 0.03	-3.44 ± 0.06	7.11 ± 0.06	11.15	2.4	$1817.9^{+151.1}_{-207.6}$		CV
PDS290	09:26:11.10	-52:42:27.0	B7	1.15 ± 0.03	-11.75 ± 0.05	18.24 ± 0.04	14.88	1.2	$869.1^{+31.4}_{-35.9}$		CV
PDS297	09:42:40.30	-56:15:34.0	A7	0.61 ± 0.03	-6.46 ± 0.05	4.88 ± 0.05	11.96	0.5	$1586.5^{+108.4}_{-140.8}$		CV
PDS322	10:52:08.68	-56:12:06.8	B3	-1.3 ± 0.27	-5.95 ± 0.44	5.89 ± 0.5	11.99	0.35	$1732.0^{+362.6}_{-606.7}$		CV
PDS324	10:57:24.30	-62:53:13.0	B1	0.31 ± 0.02	-6.17 ± 0.04	2.37 ± 0.03	14.22	1.43	$2879.8^{+241.1}_{-327.3}$		CV
PDS33	08:48:45.70	-40:48:21.0	A0	1.05 ± 0.03	-6.14 ± 0.04	4.68 ± 0.04	12.28	0.43	$951.1^{+37.1}_{-42.8}$		CV
PDS34	08:49:58.50	-45:53:06.0	B2	0.46 ± 0.02	-4.86 ± 0.03	5.62 ± 0.03	13.69	1.28	$2132.9^{+111.8}_{-135.9}$		CV
PDS344	11:40:32.80	-64:32:06.0	B5	0.4 ± 0.02	-6.96 ± 0.03	0.28 ± 0.03	13.12	0.49	$2439.5^{+139.7}_{-173.0}$		CV
PDS364	13:20:03.60	-62:23:54.0	B2	-0.39 ± 0.12	-4.47 ± 0.16	-2.18 ± 0.15	13.33	0.84	$2434.7^{+423.1}_{-658.6}$		CV
PDS371	13:47:31.40	-36:39:50.0	O9	9.87 ± 0.21	-34.32 ± 0.46	-25.59 ± 0.32	15.62	2.54	$101.4^{+3.4}_{-3.8}$		CV
PDS389	15:14:47.05	-62:16:59.8	A3	1.24 ± 0.03	-2.85 ± 0.04	-4.44 ± 0.05	13.19	2.37	$804.4^{+32.4}_{-37.7}$		CV
PDS394	15:35:17.13	-61:59:04.2	F0				13.36	1.21			C
PDS406	16:05:03.91	-39:45:03.8	A5				14.14	0.91			C
PDS415N	16:18:37.22	-24:05:18.5	F0	6.93 ± 0.13	-9.0 ± 0.31	-21.3 ± 0.19	11.42	1.29	$144.2^{+4.2}_{-4.7}$		CV
PDS431	16:54:59.20	-43:21:50.0	A0	0.53 ± 0.03	1.74 ± 0.06	-0.64 ± 0.04	13.26	0.89	$1811.5^{+122.1}_{-157.8}$		CV

Table C.9: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
PDS453	17:20:56.10	-26:03:31.0	F2	5.7 ± 0.57	-9.57 ± 0.87	-28.87 ± 0.68	13.11	1.15	$175.9^{+22.2}_{-39.7}$		CV
PDS469	17:50:58.10	-14:16:12.0	A0	0.86 ± 0.08	1.08 ± 0.12	-0.61 ± 0.09	12.89	0.98	$1075.6^{+120.0}_{-188.2}$		CV
PDS477	18:00:30.30	-16:47:26.0	B1	-0.19 ± 0.1	-1.16 ± 0.18	-0.69 ± 0.14	13.65	1.91	$2471.9^{+413.8}_{-644.7}$		CV
PDS520	18:30:06.20	+00:42:34.0	F3	2.6 ± 0.19	4.44 ± 0.21	-8.68 ± 0.17	14.54	2.82	$380.4^{+37.0}_{-55.6}$		CV
PDS530	18:41:34.40	+08:08:21.0	F0	0.27 ± 0.17	-1.16 ± 0.25	-6.37 ± 0.24	14.04	0.84	$1391.4^{+257.9}_{-469.9}$		CV
PDS543	18:48:00.70	+02:54:17.0	B1	0.64 ± 0.06	-0.19 ± 0.11	-2.2 ± 0.11	11.23	2.61	$1413.2^{+155.2}_{-238.6}$		CV
PDS551	18:55:22.98	+04:04:35.1	O9	1.99 ± 0.12	1.48 ± 0.2	-3.97 ± 0.19	16.31	2.5	$496.1^{+41.6}_{-58.4}$		CV
PDS581	19:36:18.90	+29:32:50.0	B0	0.96 ± 0.38	-2.02 ± 0.61	-7.28 ± 0.63	13.52	1.13	$687.9^{+150.7}_{-354.3}$		CV
PVCep	20:45:53.98	+67:57:38.6	A5	2.91 ± 0.06	8.23 ± 0.13	-1.98 ± 0.11	13.84	3.28	$343.4^{+10.9}_{-12.2}$	Yes	CV
PXVul	19:26:40.30	+23:53:51.0	F3	1.58 ± 0.07	0.64 ± 0.1	-6.12 ± 0.1	11.25	1.33	$627.4^{+39.1}_{-49.8}$		CV
RCrA	19:01:53.69	-36:57:08.6	A5	10.54 ± 0.7	1.58 ± 1.2	-30.84 ± 1.19	12.79	2.35	$95.4^{+8.7}_{-12.9}$	Yes	CV
RMon	06:39:09.95	+08:44:09.6	B8				12.77	1.29			C
RRTau	05:39:30.50	+26:22:27.0	A0	1.28 ± 0.05	-1.16 ± 0.09	-4.2 ± 0.07	12.35	1.27	$773.4^{+40.5}_{-49.5}$	Yes	CV
RYOri	05:32:09.90	-02:49:47.0	F6	2.71 ± 0.04	1.42 ± 0.07	-0.63 ± 0.06	11.39	1.09	$368.5^{+8.3}_{-9.0}$		CV
SAO185668	17:43:55.60	-22:05:45.0	B3	0.6 ± 0.06	2.02 ± 0.11	-0.06 ± 0.08	9.45	0.64	$1481.6^{+163.7}_{-251.6}$		CV
SAO220669	08:55:45.90	-44:25:14.0	B4	1.07 ± 0.03	-10.73 ± 0.05	6.67 ± 0.06	8.69	0.9	$932.1^{+40.3}_{-47.4}$		CV
SVCep	22:21:33.20	+73:40:27.0	A0	2.9 ± 0.02	5.21 ± 0.04	1.59 ± 0.04	10.35	0.61	$344.3^{+3.8}_{-4.0}$	Yes	CV
TCrA	19:01:58.79	-36:57:50.3	F0		5.3 ± 4.7	-24.9 ± 4.7	13.33	1.54			C
TOri	05:35:50.50	-05:28:35.0	A3	2.45 ± 0.04	2.14 ± 0.08	0.59 ± 0.07	11.44	1.12	$407.5^{+11.4}_{-12.6}$	Yes	CV
TYCrA	19:01:40.80	-36:52:34.0	B9	7.33 ± 0.15	2.15 ± 0.24	-32.72 ± 0.23	9.22	1.11	$136.5^{+4.2}_{-4.7}$	Yes	CV
UXOri	05:04:30.00	-03:47:14.0	A2	3.08 ± 0.05	0.79 ± 0.09	-3.92 ± 0.06	10.4	0.75	$324.9^{+8.4}_{-9.3}$	Yes	CV
UYOri	05:32:00.30	-04:55:54.0	B9	2.81 ± 0.08	1.74 ± 0.18	0.88 ± 0.11	12.4	0.36	$355.3^{+15.6}_{-18.4}$		CV
V1012Ori	05:11:36.55	-02:22:48.5	A3	2.59 ± 0.05	1.85 ± 0.11	-0.65 ± 0.09	12.22	0.84	$386.4^{+11.8}_{-13.2}$	Yes	CV
V1295Aql	20:03:02.50	+05:44:17.0	A0	1.12 ± 0.07	0.45 ± 0.1	-9.25 ± 0.06	7.78	0.23	$870.9^{+70.0}_{-96.3}$	Yes	CV
V1478Cyg	20:32:45.50	+40:39:37.0	B0	0.62 ± 0.1	-3.01 ± 0.19	-4.11 ± 0.2	11.03	2.96	$1296.7^{+186.8}_{-323.8}$		CV
V1493Cyg	20:52:04.60	+44:37:30.0	A2	0.83 ± 0.09	-2.05 ± 0.13	-2.96 ± 0.13	13.87	2.87	$1105.0^{+130.3}_{-209.5}$	Yes	CV

Table C.10: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
V1685Cyg	20:20:28.20	+41:21:51.0	B3	1.09 ± 0.03	-2.4 ± 0.04	-5.77 ± 0.05	10.23	1.41	$910.2^{+39.1}_{-45.9}$	Yes	CV
V1686Cyg	20:20:29.40	+41:21:28.0	A4	0.78 ± 0.12	-1.67 ± 0.44	-6.98 ± 0.72	13.21	2.18	$1078.8^{+155.9}_{-277.3}$	Yes	CV
V1787Ori	05:38:09.30	-06:49:17.0	A5	2.56 ± 0.05	0.24 ± 0.08	-0.54 ± 0.09	13.02	2.28	$391.0^{+12.6}_{-14.2}$	Yes	CV
V1818Ori	05:53:42.60	-10:24:01.0	B7	1.41 ± 0.09	-1.42 ± 0.16	1.42 ± 0.21	10.58	2.11	$695.0^{+58.7}_{-82.4}$		CV
V1977Cyg	20:47:37.50	+43:47:25.0	B8	1.16 ± 0.03	-1.67 ± 0.04	-1.73 ± 0.05	10.73	1.08	$859.8^{+33.5}_{-38.7}$	Yes	CV
V2019Cyg	20:48:04.80	+43:47:26.0	B2	1.18 ± 0.03	-1.3 ± 0.05	-1.99 ± 0.04	11.01	1.07	$842.6^{+31.4}_{-36.0}$	Yes	CV
V346Ori	05:24:42.80	+01:43:48.0	A8	2.73 ± 0.04	1.08 ± 0.07	-0.94 ± 0.06	10.08	0.37	$366.4^{+9.1}_{-9.9}$	Yes	CV
V350Ori	05:40:11.80	-09:42:11.0	A1	2.55 ± 0.12	1.23 ± 0.21	0.14 ± 0.18	11.73	0.9	$390.3^{+26.6}_{-34.7}$	Yes	CV
V351Ori	05:44:18.80	+00:08:40.0	A7	2.92 ± 0.05	2.32 ± 0.08	-0.77 ± 0.08	8.89	0.55	$341.8^{+8.3}_{-9.1}$		CV
V361Cep	21:42:50.20	+66:06:35.0	B2	1.12 ± 0.03	-1.9 ± 0.05	-3.75 ± 0.05	9.92	0.76	$892.7^{+31.0}_{-35.2}$	Yes	CV
V373Cep	21:43:06.80	+66:06:54.0	B5	1.08 ± 0.02	-1.54 ± 0.04	-3.24 ± 0.04	12.2	1.56	$922.1^{+29.4}_{-33.0}$	Yes	CV
V374Cep	23:05:07.50	+62:15:36.0	B4	1.14 ± 0.03	-1.35 ± 0.05	-2.22 ± 0.04	10.21	1.25	$872.2^{+34.6}_{-40.0}$	Yes	CV
V375Lac	22:34:40.99	+40:40:04.3	A7		-15.7 ± 5.4	9.9 ± 5.4	14.14	1.38			C
V376Cas	00:11:26.56	+58:50:03.8	B5				16.43	2.33			C
V380Ori	05:36:25.40	-06:42:58.0	A1	2.04 ± 0.16	0.02 ± 0.29	-2.42 ± 0.25	10.52	1.52	$481.7^{+49.2}_{-75.6}$	Yes	CV
V388Vel	08:42:17.30	-40:44:10.0	A1	0.2 ± 0.06	-6.33 ± 0.1	3.73 ± 0.1	15.39	2.43	$2466.9^{+344.6}_{-535.9}$	Yes	CV
V431Sct	18:29:25.70	-06:04:37.0	B1	0.63 ± 0.07	-2.07 ± 0.11	-5.64 ± 0.1	11.22	1.5	$1403.8^{+161.1}_{-252.2}$	Yes	CV
V586Ori	05:36:59.25	-06:09:16.3	A2	2.75 ± 0.1	2.38 ± 0.17	-1.56 ± 0.16	9.64	0.25	$362.7^{+19.3}_{-23.7}$	Yes	CV
V590Mon	06:40:44.60	+09:48:02.0	B7	1.14 ± 0.13	-1.55 ± 0.23	-3.5 ± 0.21	12.75	0.36	$818.4^{+102.6}_{-173.3}$	Yes	CV
V594Cas	00:43:18.30	+61:54:40.0	B8	1.76 ± 0.03	3.37 ± 0.04	-2.08 ± 0.04	10.26	1.04	$569.2^{+14.2}_{-15.6}$	Yes	CV
V599Ori	05:38:58.60	-07:16:46.0	A8	2.44 ± 0.04	0.13 ± 0.08	-0.59 ± 0.1	12.76	2.32	$410.2^{+11.2}_{-12.3}$	Yes	CV
V633Cas	00:11:26.50	+58:49:29.0	B9		2.7 ± 3.3	-8.2 ± 4.6	14.2	1.69			C
V645Cyg	21:39:58.30	+50:14:21.0	O7	0.53 ± 0.4	-6.98 ± 0.73	-0.73 ± 0.67	12.21	2.47	$786.9^{+182.0}_{-408.7}$	Yes	CV
V669Cep	22:26:38.70	+61:13:32.0	B5	0.99 ± 0.07	-1.37 ± 0.12	-2.4 ± 0.12	11.94	1.35	$977.6^{+85.2}_{-120.6}$		CV
V718Sco	16:13:11.60	-22:29:07.0	A8	6.56 ± 0.08	-8.6 ± 0.12	-23.64 ± 0.08	8.73	0.69	$152.5^{+3.0}_{-3.2}$		CV
V883Ori	05:38:18.10	-07:02:26.0	B0	3.71 ± 0.44	3.34 ± 0.82	-2.04 ± 0.71	16.66	5.09			C

Table C.11: Table C.1 continued

Star	RA	DEC	SPT	ϖ	μ_{α^*}	μ_δ	G	B_p-R_p	Distance	The *et al.* (1994) flag	Catalogue flag
	h:m:s	deg:m:s		mas	mas yr^{-1}	mas yr^{-1}	mag		pc		
V892Tau	04:18:40.60	+28:19:16.0	A0	8.52 ± 0.12	6.62 ± 0.24	-29.87 ± 0.14	13.48	3.26	$117.5^{+2.5}_{-2.7}$	Yes	CV
V921Sco	16:59:06.80	-42:42:08.0	B0	0.59 ± 0.05	1.02 ± 0.13	-2.89 ± 0.09	10.31	2.18	$1545.6^{+160.3}_{-239.6}$	Yes	CV
VVSer	18:28:47.90	+00:08:40.0	B6	2.38 ± 0.05	3.47 ± 0.07	-8.61 ± 0.07	11.75	1.7	$419.7^{+12.6}_{-14.0}$	Yes	CV
VXCas	00:31:30.70	+61:58:51.0	A0	1.86 ± 0.04	3.41 ± 0.05	-2.5 ± 0.05	11.45	0.52	$536.6^{+16.4}_{-18.3}$	Yes	CV
VYMon	06:31:06.90	+10:26:05.0	B5	-1.94 ± 0.38	7.7 ± 0.56	-6.72 ± 0.51	13.86	2.67	$1467.2^{+332.4}_{-579.5}$	Yes	CV
WRAY15-1435	16:13:06.70	-50:23:20.1	B1	0.47 ± 0.05	-3.15 ± 0.1	-2.58 ± 0.06	12.36	1.61	$1858.3^{+200.1}_{-300.1}$		V
WWVul	19:25:58.80	+21:12:31.0	A3	1.98 ± 0.04	0.84 ± 0.05	-6.47 ± 0.05	10.42	0.59	$503.5^{+14.5}_{-16.2}$	Yes	CV
XYPere+w	03:49:36.35	+38:58:55.4	A2	2.17 ± 0.09	8.61 ± 0.2	-7.82 ± 0.14	9.46	0.71	$459.1^{+27.4}_{-34.6}$	Yes	CV
ZCMa	07:03:43.20	-11:33:06.0	B5	4.3 ± 0.89	-3.47 ± 1.43	5.04 ± 1.38	9.15	1.64	$229.7^{+46.8}_{-146.0}$	Yes	CV

Table C.12: Other stellar parameters of the Herbig Ae/Be stars

Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg	Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A0974-15	B3e	F	$2.71^{+0.36}_{-0.23}$	$1.93^{+17.77}_{-0.51}$	0.36	GSC5988-2257	B3	S			0.146
ABAur	A0Ve	S	$2.15^{+0.36}_{-0.21}$	$4.05^{+1.43}_{-1.49}$	0.879	GSC5990-0021	B9e	S	$1.95^{+0.1}_{-0.1}$	$10.1^{+9.9}_{-3.72}$	0.127
AKSco	F5V	F	$1.4^{+0.07}_{-0.07}$	$8.38^{+1.72}_{-0.42}$	1.019	GSC6542-2339	O9	S			0.169
AS310	B1e	S	$11.89^{+4.76}_{-3.4}$	$0.06^{+0.54}_{-0.04}$	0.068	GSC6546-3156	A0	S	$2.04^{+0.29}_{-0.14}$	$5.68^{+14.32}_{-1.94}$	0.102
AS470	A:	S	$7.05^{+2.83}_{-1.77}$	$0.1^{+0.2}_{-0.07}$	0.035	GSC8143-1225	F3V	S	$1.27^{+0.06}_{-0.07}$	$17.73^{+2.28}_{-3.02}$	0.371
AS477	A0.5IIIer	S	$3.33^{+1.13}_{-0.41}$	$1.25^{+0.64}_{-0.73}$	0.185	GSC8581-2002	A5V	F	$1.88^{+0.09}_{-0.09}$	$7.87^{+12.13}_{-0.98}$	0.257
BD+30549	B8:p	S	$2.28^{+0.37}_{-0.19}$	$5.48^{+14.52}_{-2.26}$	0.485	GSC8645-1401	F2	S	$4.04^{+0.66}_{-0.56}$	$0.56^{+0.33}_{-0.23}$	0.081
BD+413731	B3n	S	$4.98^{+1.24}_{-0.69}$	$0.52^{+2.96}_{-0.26}$	0.143	GSC8993-0397	B3	S	$5.0^{+1.0}_{-0.7}$	$0.6^{+3.76}_{-0.28}$	0.048
BFOri	A7III	F	$1.81^{+0.09}_{-0.09}$	$6.38^{+0.32}_{-0.46}$	0.368	GSC8994-3902	B2V	S	$8.42^{+3.94}_{-1.99}$	$0.12^{+0.44}_{-0.08}$	0.06
BHCep	F5IIIe	S	$1.37^{+0.15}_{-0.1}$	$10.57^{+3.03}_{-3.14}$	0.427	GUCMa	B2Vne	F			1.0
BOCep	F5Ve	S	$1.22^{+0.06}_{-0.06}$	$17.1^{+0.86}_{-2.4}$	0.382	HBC1	F4?	S			0.188
BPPsc	G9IIIe	S	$1.9^{+0.5}_{-0.26}$	$1.66^{+1.56}_{-1.01}$	0.411	HBC217	G0	S	$1.51^{+0.16}_{-0.08}$	$7.29^{+0.73}_{-1.81}$	0.206
COOri	F7Ve	S	$2.6^{+0.2}_{-0.24}$	$1.76^{+0.62}_{-0.36}$	0.355	HBC222	F8V	S	$1.53^{+0.14}_{-0.08}$	$7.16^{+0.86}_{-1.68}$	0.203
CPM25	B2	S	$5.17^{+2.25}_{-1.22}$	$0.65^{+5.73}_{-0.43}$	0.067	HBC324	A5e	S	$1.5^{+0.29}_{-0.08}$	$13.22^{+6.78}_{-7.3}$	0.16
CQTau	F5IVe	S	$1.47^{+0.19}_{-0.11}$	$8.9^{+2.8}_{-2.52}$	0.878	HBC334	B3	S	$3.71^{+0.49}_{-0.19}$	$2.09^{+4.29}_{-1.06}$	0.081
DGCir	Ae/B[e]	F	$2.3^{+0.6}_{-0.64}$	$4.18^{+15.82}_{-2.55}$	0.172	HBC442	F8	S	$1.8^{+0.25}_{-0.09}$	$4.46^{+0.25}_{-1.47}$	0.371
DKCha	A0_sh	F	$1.37^{+0.07}_{-0.07}$	$17.16^{+2.84}_{-3.56}$	0.59	HBC694	A5V	S			0.213
DWCMa	B2?e	F	$12.24^{+5.53}_{-3.44}$	$0.04^{+0.07}_{-0.03}$	0.056	HBC7	B2e	S	$9.55^{+4.12}_{-2.09}$	$0.09^{+0.11}_{-0.06}$	0.052
GSC1829-0331	F1	S			1.0	HBC705	B2e	S	$8.45^{+3.61}_{-1.82}$	$0.12^{+0.21}_{-0.08}$	0.069
GSC1876-0892	B2	S	$9.42^{+5.22}_{-2.41}$	$0.09^{+0.13}_{-0.06}$	0.048	HBC717	F6III	S	$3.37^{+1.01}_{-0.49}$	$0.88^{+0.52}_{-0.5}$	0.103
GSC3975-0579	A2	S	$2.06^{+0.37}_{-0.12}$	$4.37^{+0.71}_{-1.6}$	0.152	HD100453	A9Ve	F	$1.25^{+0.06}_{-0.06}$	$6.53^{+0.45}_{-0.49}$	1.375
GSC4805-1306	O9?e	S			1.0	HD100546	A0VaekB8_lB	F	$2.06^{+0.1}_{-0.12}$	$5.48^{+1.41}_{-0.77}$	1.302
GSC4823-0146	B0	F	$11.07^{+2.34}_{-1.31}$	$0.08^{+0.04}_{-0.03}$	0.05	HD101412	A3VaekA0mA0_lB	F	$2.1^{+0.1}_{-0.1}$	$4.37^{+0.22}_{-0.32}$	0.348
GSC5360-1033	B5?	S			0.237	HD104237	A0_sh	S	$1.85^{+0.09}_{-0.09}$	$5.48^{+0.27}_{-0.4}$	1.321
GSC5379-0359	B9	F	$2.33^{+0.12}_{-0.12}$	$3.48^{+0.27}_{-0.26}$	0.109	HD114981	B5V	F	$6.09^{+0.59}_{-0.34}$	$0.28^{+0.05}_{-0.07}$	0.203
GSC5987-1399	Be	S			1.0	HD130437	B1V	S	$13.44^{+4.56}_{-3.75}$	$0.05^{+0.08}_{-0.03}$	0.087

Table C.13: Table C.12 continued

Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg	Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HD130437	B1V	S	$13.44^{+4.56}_{-3.75}$	$0.05^{+0.08}_{-0.03}$	0.087	HD244604	A0Vesh	F	$1.98^{+0.1}_{-0.1}$	$4.89^{+0.24}_{-0.52}$	0.341
HD132947	B9V	F	$2.22^{+0.11}_{-0.11}$	$4.05^{+0.32}_{-0.2}$	0.375	HD245185	A0Vae	F	$1.92^{+0.18}_{-0.1}$	$7.64^{+12.36}_{-2.56}$	0.334
HD135344	A0V	S	$1.74^{+0.09}_{-0.09}$	$5.75^{+0.29}_{-0.67}$	1.0	HD245906	A6	S	$2.39^{+1.09}_{-0.55}$	$2.77^{+2.71}_{-1.82}$	0.208
HD135344B	F8V	F	$1.43^{+0.07}_{-0.07}$	$8.93^{+0.45}_{-0.91}$	1.055	HD249879	B8?e	S	$2.25^{+0.49}_{-0.16}$	$5.08v^{+14.92}_{-2.31}$	0.214
HD139614	A9VekA5mA5(_lB)	F	$1.48^{+0.07}_{-0.07}$	$14.49^{+1.4}_{-3.6}$	1.063	HD250550	B9e	F	$2.6^{+0.3}_{-0.14}$	$2.56^{+0.43}_{-0.67}$	0.205
HD141569	A0Ve	W	$1.86^{+0.09}_{-0.09}$	$8.62^{+11.38}_{-1.19}$	1.295	HD259431	B6ep	S	$5.16^{+1.84}_{-1.28}$	$0.42^{+0.53}_{-0.28}$	0.199
HD141926	B2IIIn	F	$19.51^{+2.44}_{-2.25}$	$0.02^{+0.01}_{-0.01}$	0.107	HD287823	A0	F	$1.7^{+0.09}_{-0.09}$	$7.43^{+0.37}_{-0.37}$	0.399
HD142527	F6III	F	$1.61^{+0.12}_{-0.08}$	$6.63^{+0.33}_{-1.55}$	0.91	HD288012	A2	S	$2.22^{+0.39}_{-0.13}$	$3.75^{+0.62}_{-1.37}$	0.362
HD142666	A8Ve	W	$1.49^{+0.08}_{-0.08}$	$9.33^{+0.77}_{-0.47}$	0.966	HD290380	keF6IVeb	S	$1.53^{+0.16}_{-0.08}$	$7.21^{+0.81}_{-1.73}$	0.404
HD143006	G5IVe	S	$1.56^{+0.1}_{-0.14}$	$3.75^{+1.73}_{-0.76}$	0.862	HD290409	B8.5Ve	F	$1.9^{+0.18}_{-0.09}$	$7.43^{+12.57}_{-1.95}$	0.315
HD144432	A9/F0V	F	$1.39^{+0.07}_{-0.07}$	$4.98^{+0.25}_{-0.55}$	0.922	HD290500	A2	F	$1.38^{+0.08}_{-0.07}$	$10.4^{+9.3}_{-3.27}$	0.327
HD149914	B9.5IV	S	$2.97^{+0.39}_{-0.29}$	$1.74^{+0.64}_{-0.54}$	0.902	HD290764	F0V	F	$1.69^{+0.13}_{-0.09}$	$6.89^{+0.54}_{-1.41}$	0.36
HD150193	B9.5Ve/A2IVe	W	$1.89^{+0.09}_{-0.09}$	$5.48^{+0.44}_{-0.27}$	0.95	HD290770	B9	F	$2.22^{+0.11}_{-0.11}$	$4.59^{+0.49}_{-0.54}$	0.359
HD155448	B1V	S	$4.82^{+0.92}_{-0.98}$	$0.44^{+0.45}_{-0.2}$	0.15	HD305298	O8	F	$17.72^{+2.13}_{-2.04}$	$0.04^{+0.31}_{-0.01}$	0.035
HD158643	A0V	S	$3.35^{+0.79}_{-0.22}$	$1.22^{+0.29}_{-0.57}$	1.166	HD313571	B3Ve	S	$7.26^{+2.94}_{-1.23}$	$0.17^{+0.14}_{-0.11}$	0.103
HD163296	A1Vep	W	$1.83^{+0.09}_{-0.09}$	$7.6^{+1.05}_{-1.22}$	1.411	HD31648	A5Vep	S	$1.78^{+0.13}_{-0.09}$	$6.2^{+0.31}_{-1.12}$	0.885
HD169142	F1VekA3mA3_lB?	S	$2.0^{+0.13}_{-0.13}$	$8.98^{+11.02}_{-3.9}$	1.257	HD319896	B4	S	$5.9^{+1.22}_{-0.85}$	$0.3^{+0.18}_{-0.13}$	0.111
HD17081	B7IV	S	$3.94^{+0.81}_{-0.47}$	$0.89^{+0.4}_{-0.41}$	1.342	HD323771	B5Vp	S	$3.84^{+0.36}_{-0.38}$	$1.08^{+0.55}_{-0.26}$	0.134
HD174571	B2	S	$12.85^{+6.44}_{-3.2}$	$0.04^{+0.06}_{-0.03}$	0.131	HD34282	B9.5V	F	$1.45^{+0.07}_{-0.07}$	$6.54^{+2.41}_{-0.63}$	0.46
HD176386	B9V	S	$2.3^{+0.14}_{-0.3}$	$4.05^{+15.95}_{-0.57}$	0.9	HD344261	F2	S	$1.34^{+0.07}_{-0.07}$	$12.14^{+3.76}_{-2.04}$	0.477
HD179218	A0IVe	W	$2.98^{+0.18}_{-0.3}$	$1.66^{+0.54}_{-0.26}$	0.539	HD34700	G0IV	S	$2.66^{+0.32}_{-0.13}$	$1.4^{+0.23}_{-0.44}$	0.402
HD199603	A9V	S	$2.03^{+0.1}_{-0.1}$	$4.05^{+0.2}_{-0.2}$	1.605	HD35187	A2e+A7	S	$2.1^{+0.25}_{-0.25}$	$4.9^{+15.1}_{-1.68}$	0.879
HD200775	B2Ve	S	$5.34^{+1.33}_{-0.47}$	$0.41^{+0.15}_{-0.2}$	0.397	HD244604	A0Vesh	F	$1.98^{+0.1}_{-0.1}$	$4.89^{+0.24}_{-0.52}$	0.341
HD203024	A5V	W			1.0	HD245185	A0Vae	F	$1.92^{+0.18}_{-0.1}$	$7.64^{+12.36}_{-2.56}$	0.334
HD235495	A0	S	$2.06^{+0.29}_{-0.1}$	$4.71^{+0.77}_{-1.49}$	0.273	HD245906	A6	S	$2.39^{+1.09}_{-0.55}$	$2.77^{+2.71}_{-1.82}$	0.208
HD244314	A1V	F	$1.69^{+0.09}_{-0.09}$	$7.43^{+0.37}_{-0.54}$	0.332	HD249879	B8?e	S	$2.25^{+0.49}_{-0.16}$	$5.08^{+14.92}_{-2.31}$	0.214

Table C.14: Table C.12 continued

Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg	Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HD250550	B9e	F	$2.6^{+0.3}_{-0.14}$	$2.56^{+0.43}_{-0.67}$	0.205	HD37490	B3Ve	S	$8.58^{+3.93}_{-1.65}$	$0.1^{+0.11}_{-0.07}$	0.459
HD259431	B6ep	S	$5.16^{+1.84}_{-1.28}$	$0.42^{+0.53}_{-0.28}$	0.199	HD37806	B9/9.5II/III	S	$3.11^{+0.55}_{-0.33}$	$1.56^{+0.64}_{-0.6}$	0.335
HD287823	A0	F	$1.7^{+0.09}_{-0.09}$	$7.43^{+0.37}_{-0.37}$	0.399	HD38087	B3II	S	$3.21^{+0.79}_{-0.38}$	$1.75^{+9.15}_{-0.64}$	0.424
HD288012	A2	S	$2.22^{+0.39}_{-0.13}$	$3.75^{+0.62}_{-1.37}$	0.362	HD38120	B9Vnne	S	$2.37^{+0.43}_{-0.24}$	$3.48^{+13.62}_{-1.44}$	0.354
HD290380	keF6IVeb	S	$1.53^{+0.16}_{-0.08}$	$7.21^{+0.81}_{-1.73}$	0.404	HD39014	A7V	S	$2.0^{+0.23}_{-0.1}$	$4.37^{+0.22}_{-1.15}$	3.248
HD290409	B8.5Ve	F	$1.9^{+0.18}_{-0.09}$	$7.43^{+12.57}_{-1.95}$	0.315	HD41511	A3+K/M	S	$5.79^{+1.27}_{-0.76}$	$0.22^{+0.14}_{-0.1}$	0.645
HD290500	A2	F	$1.38^{+0.08}_{-0.07}$	$10.4^{+9.3}_{-3.27}$	0.327	HD45677	B2IV/V	S	$4.72^{+1.19}_{-0.38}$	$0.6^{+3.76}_{-0.3}$	0.231
HD290764	F0V	F	$1.69^{+0.13}_{-0.09}$	$6.89^{+0.54}_{-1.41}$	0.36	HD46060	B2II	S	$9.64^{+3.37}_{-2.4}$	$0.09^{+0.12}_{-0.05}$	0.154
HD290770	B9	F	$2.22^{+0.11}_{-0.11}$	$4.59^{+0.49}_{-0.54}$	0.359	HD50083	B3	S	$11.42^{+3.66}_{-1.76}$	$0.05^{+0.04}_{-0.02}$	0.131
HD305298	O8	F	$17.72^{+2.13}_{-2.04}$	$0.04^{+0.31}_{-0.01}$	0.035	HD50138	AIIb/II	S	$4.17^{+0.46}_{-0.32}$	$0.63^{+0.19}_{-0.18}$	0.377
HD313571	B3Ve	S	$7.26^{+2.94}_{-1.23}$	$0.17^{+0.14}_{-0.11}$	0.103	HD53367	B0IV/Ve	F			1.105
HD31648	A5Vep	S	$1.78^{+0.13}_{-0.09}$	$6.2^{+0.31}_{-1.12}$	0.885	HD56895B	F0V	S	$1.53^{+0.11}_{-0.08}$	$8.31^{+0.42}_{-1.42}$	0.867
HD319896	B4	S	$5.9^{+1.22}_{-0.85}$	$0.3^{+0.18}_{-0.13}$	0.111	HD58647	B9IV	S	$3.87^{+0.33}_{-0.19}$	$0.84^{+0.12}_{-0.18}$	0.45
HD323771	B5Vp	S	$3.84^{+0.36}_{-0.38}$	$1.08^{+0.55}_{-0.26}$	0.134	HD59319	B8V	F	$3.81^{+0.31}_{-0.26}$	$0.96^{+0.24}_{-0.2}$	0.214
HD34282	B9.5V	F	$1.45^{+0.07}_{-0.07}$	$6.54^{+2.41}_{-0.63}$	0.46	HD68695	A3VbekA0mA0_lB	F	$1.83^{+0.09}_{-0.09}$	$7.59^{+1.06}_{-1.21}$	0.362
HD344261	F2	S	$1.34^{+0.07}_{-0.07}$	$12.14^{+3.76}_{-2.04}$	0.477	HD72106	A0IV	F	$2.67^{+1.47}_{-0.67}$	$2.13^{+2.58}_{-1.52}$	1.0
HD34700	G0IV	S	$2.66^{+0.32}_{-0.13}$	$1.4^{+0.23}_{-0.44}$	0.402	HD76534	B2Vn	F	$7.46^{+0.51}_{-0.37}$	$0.17^{+0.02}_{-0.03}$	0.157
HD35187	A2e+A7	S	$2.1^{+0.25}_{-0.25}$	$4.9^{+15.1}_{-1.68}$	0.879	HD85567	B5Vne	F	$6.32^{+0.52}_{-0.39}$	$0.22^{+0.05}_{-0.05}$	0.14
HD35929	F2IV/V	F	$2.92^{+0.15}_{-0.15}$	$1.46^{+0.07}_{-0.17}$	0.37	HD87403	A1III	F	$5.51^{+0.65}_{-0.53}$	$0.28^{+0.11}_{-0.08}$	0.075
HD36112	A8V	S	$1.56^{+0.11}_{-0.08}$	$8.29^{+0.41}_{-1.4}$	0.894	HD87643	B3I[e]	S	$17.72^{+10.87}_{-6.72}$	$0.02^{+0.05}_{-0.01}$	0.071
HD36408	B7III	S	$6.21^{+1.18}_{-1.01}$	$0.22^{+0.17}_{-0.09}$	0.329	HD94509	Bp_sh	F	$5.69^{+1.1}_{-0.77}$	$0.28^{+0.17}_{-0.12}$	0.078
HD36917	B9III/IV	S	$3.71^{+0.94}_{-0.75}$	$0.99^{+0.9}_{-0.5}$	0.302	HD95881	A0	F	$5.5^{+0.5}_{-0.28}$	$0.28^{+0.05}_{-0.07}$	0.123
HD36982	B1.5V	S	$5.2^{+0.42}_{-0.29}$	$0.73^{+0.47}_{-0.17}$	0.351	HD96042	B1(V)ne	F	$20.7^{+3.93}_{-2.86}$	$0.02^{+0.01}_{-0.01}$	0.046
HD37357	A1V	S	$2.95^{+0.96}_{-0.41}$	$1.7^{+0.87}_{-0.93}$	0.221	HD9672	A1V	S	$1.81^{+0.09}_{-0.09}$	$6.89^{+0.34}_{-0.51}$	2.51
HD37371	A0IV	S	$3.85^{+0.63}_{-0.67}$	$0.86^{+0.65}_{-0.34}$	0.348	HD97048	A0Vep	F	$2.25^{+0.11}_{-0.14}$	$4.37^{+1.11}_{-0.32}$	0.775
HD37411	hA3VakA0mA0(eb)_lB	F			1.0	HD98922	B9Ve	F	$6.17^{+0.37}_{-0.31}$	$0.2^{+0.01}_{-0.04}$	0.208

Table C.15: Table C.12 continued

Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg	Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HKOri	A2?e+G0?e	F			1.0	MWC1021	Bpe	S	$32.0^{+1.6}_{-14.44}$	$0.01^{+0.02}_{-0.0}$	0.087
HR5999	A7	F	$2.43^{+0.12}_{-0.12}$	$2.73^{+0.26}_{-0.35}$	0.889	MWC1080	B0eq	S	$16.08^{+6.34}_{-4.22}$	$0.04^{+0.45}_{-0.02}$	0.119
HTCMa	A0e	F	$2.12^{+0.19}_{-0.12}$	$5.21^{+14.79}_{-1.16}$	0.128	MWC137	sgB[e]	S	$23.1^{+10.55}_{-6.53}$	$0.02^{+0.02}_{-0.01}$	0.049
HUCMa	B8e	F	$3.02^{+0.15}_{-0.15}$	$2.04^{+0.34}_{-0.15}$	0.122	MWC297	B1.5Ve	F	$16.9^{+1.87}_{-1.22}$	$0.03^{+0.01}_{-0.01}$	0.382
Hen2-80	B6Ve	S	$3.07^{+0.9}_{-0.39}$	$2.2^{+8.7}_{-1.32}$	0.19	MWC314	B3Ibe	S	$24.58^{+1.23}_{-1.23}$	$0.01^{+0.0}_{-0.0}$	0.048
Hen3-1121	B0.5Ve	S	$10.06^{+18.59}_{-3.66}$	$0.11^{+1.29}_{-0.1}$	0.082	MWC342	B[e]	S	$22.45^{+9.0}_{-6.01}$	$0.02^{+0.02}_{-0.01}$	0.079
Hen3-1121S	B0V	S	$12.1^{+5.21}_{-3.18}$	$0.08^{+0.53}_{-0.04}$	0.125	MWC593	B4Ve	S	$7.99^{+1.8}_{-1.26}$	$0.12^{+0.09}_{-0.06}$	0.107
Hen3-1191	B[e]	S	$8.14^{+2.05}_{-0.41}$	$0.23^{+0.37}_{-0.11}$	0.086	MWC623	K2Ib/II+B4III	S	$18.2^{+4.67}_{-3.12}$	$0.02^{+0.01}_{-0.01}$	0.044
Hen3-1475	B7e	S			1.0	MWC655	B1:IV:nnpe	S	$11.46^{+4.16}_{-2.97}$	$0.07^{+0.54}_{-0.04}$	0.066
Hen3-373	B2?	F	$10.89^{+4.53}_{-3.03}$	$0.06^{+0.1}_{-0.03}$	0.074	MWC657	Be	S	$17.9^{+14.63}_{-9.04}$	$0.02^{+0.5}_{-0.01}$	0.045
Hen3-823	B3Ve	S	$4.83^{+1.57}_{-0.5}$	$0.56^{+1.48}_{-0.32}$	0.123	MWC778	BQ	S			1.0
Hen3-847	B5e	F	$3.0^{+0.6}_{-0.15}$	$2.38^{+8.52}_{-1.09}$	0.183	MWC878	B1Ve	S	$13.51^{+5.69}_{-3.85}$	$0.05^{+0.47}_{-0.03}$	0.081
Hen3-938	O9	S	$25.79^{+6.57}_{-5.98}$	$0.02^{+0.01}_{-0.01}$	0.037	MWC930	B5/9Iaeq	S			0.055
Hen3-949	B4+B7	F	$4.18^{+0.74}_{-0.51}$	$0.82^{+5.56}_{-0.34}$	1.0	MWC953	B2Ve	S	$12.66^{+6.63}_{-3.32}$	$0.04^{+0.06}_{-0.03}$	0.083
ILCep	B3IVe+A3	S	$9.85^{+2.66}_{-1.27}$	$0.07^{+0.04}_{-0.03}$	0.178	NSV2968	B0e	S	$8.97^{+2.04}_{-0.97}$	$0.2^{+0.4}_{-0.1}$	0.14
IPPer	A7III:kA2.5mA3e(r)	S	$1.56^{+0.11}_{-0.12}$	$11.99^{+8.01}_{-3.34}$	0.466	NVOri	F6IIIe	S	$1.84^{+0.25}_{-0.11}$	$4.94^{+0.98}_{-1.46}$	0.371
KKOph	A6Ve+G5Ve	F	$1.51^{+0.08}_{-0.08}$	$18.5^{+1.5}_{-1.4}$	0.648	NXPup	A1e/F2IIIe	S	$5.0^{+1.4}_{-0.75}$	$0.28^{+0.2}_{-0.17}$	0.086
LKHa260	Be	S	$3.03^{+0.53}_{-0.31}$	$2.2^{+8.7}_{-0.8}$	0.116	PDS002	F3Ve	S	$1.38^{+0.09}_{-0.07}$	$10.9^{+0.55}_{-2.25}$	0.346
LKHa338	B9	S	$1.88^{+0.09}_{-0.09}$	$9.33^{+10.67}_{-1.9}$	0.162	PDS004	A0	S	$1.9^{+0.32}_{-0.09}$	$7.43^{+12.57}_{-3.06}$	0.359
LkHa208	F0Ve	S	$1.56^{+0.47}_{-0.14}$	$8.65^{+11.35}_{-4.6}$	0.228	PDS021	B3V[e]	F	$3.94^{+0.51}_{-0.45}$	$1.0^{+3.05}_{-0.34}$	0.162
LkHa215	B7IIne	S	$3.82^{+0.59}_{-0.37}$	$1.03^{+0.26}_{-0.38}$	0.201	PDS022	A0Va-e	S	$2.46^{+0.73}_{-0.24}$	$2.85^{+0.9}_{-1.45}$	0.168
LkHa257	B5/A2e	S	$3.08^{+0.15}_{-0.15}$	$3.62^{+1.09}_{-1.06}$	0.18	PDS025	A3	S	$1.23^{+0.06}_{-0.06}$	$11.34^{+8.36}_{-2.39}$	0.188
LkHa259	A9e	S	$1.7^{+0.1}_{-0.13}$	$6.38^{+1.64}_{-0.9}$	0.19	PDS123		S	$5.03^{+3.66}_{-2.34}$	$0.7^{+10.2}_{-0.63}$	0.096
LkHa324	B9Ve	S	$2.82^{+0.61}_{-0.2}$	$2.12^{+0.44}_{-0.92}$	0.237	PDS124	A0	F	$2.07^{+0.1}_{-0.12}$	$5.92^{+14.08}_{-1.21}$	0.168
LkHa339	A1e	F	$2.59^{+0.13}_{-0.13}$	$2.54^{+0.23}_{-0.16}$	0.167	PDS126	A7V	S	$1.94^{+0.23}_{-0.17}$	$4.71^{+1.21}_{-1.23}$	0.169
MQCas	B	S	$1.8^{+0.3}_{-0.09}$	$11.29^{+8.7}_{-6.22}$	0.18	PDS129	F5V	S	$1.57^{+0.22}_{-0.16}$	$7.24^{+2.09}_{-2.16}$	0.209

Table C.16: Table C.12 continued

Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg	Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PDS133	B6	F	$2.92^{+0.45}_{-0.44}$	$2.77^{+14.33}_{-1.02}$	0.097	PDS520	F3Ve	S	$1.26^{+0.15}_{-0.06}$	$14.13^{+5.87}_{-4.03}$	0.377
PDS134	B6	F	$4.28^{+0.52}_{-0.38}$	$0.73^{+0.22}_{-0.21}$	0.056	PDS530	F0V?	S	$1.89^{+0.52}_{-0.23}$	$5.22^{+2.21}_{-2.45}$	0.103
PDS138	B0	S	$28.32^{+7.39}_{-8.22}$	$0.01^{+0.01}_{-0.0}$	0.031	PDS543	B0.5Ib	S	$30.74^{+4.97}_{-9.1}$	$0.01^{+0.01}_{-0.0}$	0.101
PDS144S	A5V	W			0.958	PDS551	O9	S			0.289
PDS211	B9	S	$2.41^{+0.29}_{-0.28}$	$3.22^{+13.88}_{-1.02}$	0.133	PDS581	B0.5IV[e]	S	$5.35^{+2.85}_{-0.27}$	$0.6^{+1.14}_{-0.4}$	0.208
PDS229		S	$2.51^{+0.47}_{-0.13}$	$4.55^{+12.55}_{-2.51}$	0.163	PVCep	A5	S			0.417
PDS24	B9e	F			1.0	PXVul	F3Ve	S	$2.1^{+0.32}_{-0.22}$	$3.48^{+1.23}_{-1.1}$	0.228
PDS277	F3V	S	$1.54^{+0.08}_{-0.08}$	$7.76^{+0.39}_{-0.87}$	0.414	RCrA	A5IIevar	S			1.502
PDS286	B0V	F	$31.17^{+4.54}_{-5.48}$	$0.01^{+0.01}_{-0.0}$	0.079	RMon	B8IIIe	F			1.0
PDS290	B7	S	$2.36^{+0.29}_{-0.12}$	$4.71^{+15.29}_{-1.49}$	0.165	RRTau	A0:IVe	S	$2.82^{+0.46}_{-0.19}$	$1.98^{+0.4}_{-0.69}$	0.185
PDS297	A7?	S	$2.98^{+0.27}_{-0.31}$	$1.75^{+0.63}_{-0.35}$	0.09	RYOri	F7	S	$1.54^{+0.15}_{-0.08}$	$7.15^{+0.87}_{-1.67}$	0.389
PDS322	B3	S	$5.4^{+2.6}_{-1.5}$	$1.07^{+5.31}_{-0.89}$	0.083	SAO185668	OB	S	$9.38^{+3.61}_{-1.65}$	$0.08^{+0.07}_{-0.05}$	0.097
PDS324	B1	S	$7.74^{+2.26}_{-2.24}$	$0.26^{+1.49}_{-0.15}$	0.05	SAO220669	B4	F	$7.94^{+1.25}_{-1.11}$	$0.13^{+0.08}_{-0.05}$	0.154
PDS33	A0	F	$1.85^{+0.09}_{-0.09}$	$10.75^{+9.26}_{-3.86}$	0.151	SVCep	A2IVe	S	$1.55^{+0.08}_{-0.08}$	$6.22^{+13.48}_{-1.0}$	0.416
PDS34	B2	S	$5.28^{+1.63}_{-1.19}$	$0.65^{+4.83}_{-0.37}$	0.067	TCrA	F0	S			1.0
PDS344	B5?	S	$3.48^{+0.17}_{-0.23}$	$1.75^{+8.35}_{-0.24}$	0.059	TOri	A3IVeb	F	$2.11^{+0.14}_{-0.1}$	$4.15^{+0.56}_{-0.67}$	0.351
PDS364	B8Ve	F	$3.3^{+0.9}_{-0.5}$	$1.46^{+0.75}_{-0.75}$	0.059	TYCrA	B9e	S	$2.06^{+0.22}_{-0.19}$	$6.38^{+13.62}_{-2.01}$	1.05
PDS371	O9	S			1.413	UXOri	A4IVe	F	$1.61^{+0.09}_{-0.08}$	$11.38^{+8.62}_{-2.73}$	0.441
PDS389	A3V	S	$2.91^{+0.36}_{-0.33}$	$1.68^{+0.7}_{-0.48}$	0.178	UYOri	B9	F			0.403
PDS394	F0Ve	S			1.0	V1012Ori	A3e	F	$1.3^{+0.06}_{-0.06}$	$8.47^{+1.06}_{-0.89}$	0.371
PDS406	A5V	S			1.0	V1295Aql	A0IVp+sh	W	$5.89^{+0.8}_{-0.76}$	$0.22^{+0.11}_{-0.07}$	0.164
PDS415N	F0V	S	$1.21^{+0.16}_{-0.09}$	$13.13^{+5.38}_{-4.47}$	0.993	V1478Cyg	Bep	S	$28.3^{+4.23}_{-15.12}$	$0.01^{+0.04}_{-0.0}$	0.11
PDS431	A0	F	$2.52^{+0.27}_{-0.15}$	$2.77^{+0.45}_{-0.73}$	0.079	V1493Cyg	A1/9e	S	$3.35^{+1.11}_{-0.53}$	$1.16^{+0.73}_{-0.67}$	0.13
PDS453	F2V	S			0.815	V1685Cyg	B2Ve	S	$9.08^{+3.93}_{-1.79}$	$0.1^{+0.11}_{-0.07}$	0.157
PDS469	A0	S	$2.15^{+0.55}_{-0.15}$	$4.05^{+1.43}_{-1.85}$	0.133	V1686Cyg	A4:Ve	S	$2.85^{+0.72}_{-0.55}$	$1.24^{+1.14}_{-0.63}$	0.133
PDS477	B1	S	$8.34^{+3.34}_{-2.43}$	$0.17^{+1.58}_{-0.11}$	0.058	V1787Ori	A5e	S	$1.66^{+0.09}_{-0.08}$	$7.43^{+0.59}_{-1.05}$	0.366

Table C.17: Table C.12 continued

Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg	Star	LumClass	Ref LumClass	Mass $M_\odot$	Age Myr	Angle deg
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
V1818Ori	Ae	S	$5.28^{+1.32}_{-1.08}$	$0.37^{+0.39}_{-0.19}$	0.206	V594Cas	B8eq	S	$2.94^{+0.59}_{-0.23}$	$1.89^{+0.49}_{-0.78}$	0.252
V1977Cyg	B8Ve	S	$3.89^{+0.35}_{-0.26}$	$0.84^{+0.19}_{-0.19}$	0.167	V599Ori	A8.9	F	$2.03^{+0.1}_{-0.1}$	$4.29^{+0.42}_{-0.54}$	0.349
V2019Cyg	B8/A0e	S	$3.5^{+0.48}_{-0.64}$	$1.13^{+0.91}_{-0.37}$	0.17	V633Cas	B9e	S			1.0
V346Ori	A8V	F	$1.57^{+0.08}_{-0.08}$	$9.33^{+0.47}_{-0.47}$	0.391	V645Cyg	O6/9eq	S			0.182
V350Ori	A7V	F	$1.71^{+0.09}_{-0.09}$	$12.17^{+7.83}_{-4.74}$	0.367	V669Cep	B[e]	S	$4.0^{+0.49}_{-0.48}$	$0.96^{+0.44}_{-0.3}$	0.147
V351Ori	A7V	S	$1.98^{+0.19}_{-0.1}$	$4.51^{+0.23}_{-1.03}$	0.419	V718Sco	A8IV	W	$1.61^{+0.08}_{-0.08}$	$9.8^{+2.8}_{-0.49}$	0.939
V361Cep	B3IV-Vne	S	$5.31^{+0.69}_{-0.48}$	$0.41^{+0.15}_{-0.13}$	0.16	V883Ori	F:I:	S			1.0
V373Cep	B5Ve	S	$3.18^{+0.51}_{-0.39}$	$1.63^{+0.75}_{-0.6}$	0.155	V892Tau	A0Ve	S			1.22
V374Cep	B5Vep	S	$6.84^{+0.96}_{-0.87}$	$0.19^{+0.09}_{-0.07}$	0.164	V921Sco	B0IVe	S	$19.96^{+6.98}_{-5.0}$	$0.02^{+0.03}_{-0.01}$	0.093
V375Lac	A7Ve	S			1.0	VVSer	A5Ve	S	$2.89^{+0.14}_{-0.14}$	$2.77^{+8.13}_{-0.21}$	0.341
V376Cas	A3/F2e	S			1.0	VXCas	A0Vep	S	$1.88^{+0.18}_{-0.09}$	$9.33+10.67_{-3.85}$	0.267
V380Ori	A1e	F	$2.82^{+0.59}_{-0.38}$	$1.96^{+1.03}_{-0.85}$	0.297	VYMon	A5:Vep	F	$8.8^{+7.46}_{-4.05}$	$0.08^{+0.49}_{-0.07}$	0.098
V388Vel		S	$4.09^{+1.02}_{-0.9}$	$0.68^{+0.72}_{-0.35}$	0.058	WRAY15-1435	B1Ve	S	$8.74^{+2.93}_{-2.55}$	$0.15^{+1.25}_{-0.09}$	0.077
V431Sct	B1Ia+[e]	S	$9.36^{+4.49}_{-2.28}$	$0.1^{+0.5}_{-0.07}$	0.102	WWVul	A2IVe	S	$1.95^{+0.11}_{-0.1}$	$5.08^{+0.84}_{-0.71}$	0.284
V586Ori	A3Ve	F	$1.88^{+0.14}_{-0.11}$	$7.93^{+12.07}_{-2.45}$	1.0	XYPere+w	A2II+B6	S	$2.82^{+0.29}_{-0.2}$	$1.96^{+0.43}_{-0.44}$	1.0
V590Mon	B4Ve	F	$2.3^{+0.13}_{-0.12}$	$5.93^{+14.07}_{-1.22}$	0.175	ZCMa	B5/8eq+F5/7	F	$3.76^{+1.98}_{-0.81}$	$0.8^{+0.83}_{-0.59}$	0.624

Appendix D

Different stages on the selection process made by CEREAL for the sample of Herbig Ae/Be stars

The following figures represent an example of the different stages of the selection process made by CEREAL for the sample of Herbig stars that were found in clusters in section 3.2.4. These stars are part of the catalogue shown in the tables 3.1 and 3.2 in chapter 3.

The HAeBe stars shown below are HD37371 (spectral type B9), HD46060 (spectral type B8), HUCMa (spectral type B8), and V590Mon (spectral type B7), which have been classified as yes by CEREAL due to the distribution of their parallax and their proper motions appearing to be evenly distributed around the known values; the density profiles show a smooth decrease in the density as a function of the distance, and the spatial and proper motion distribution show that the low mass companions around the HAeBe stars share similar parallax values.

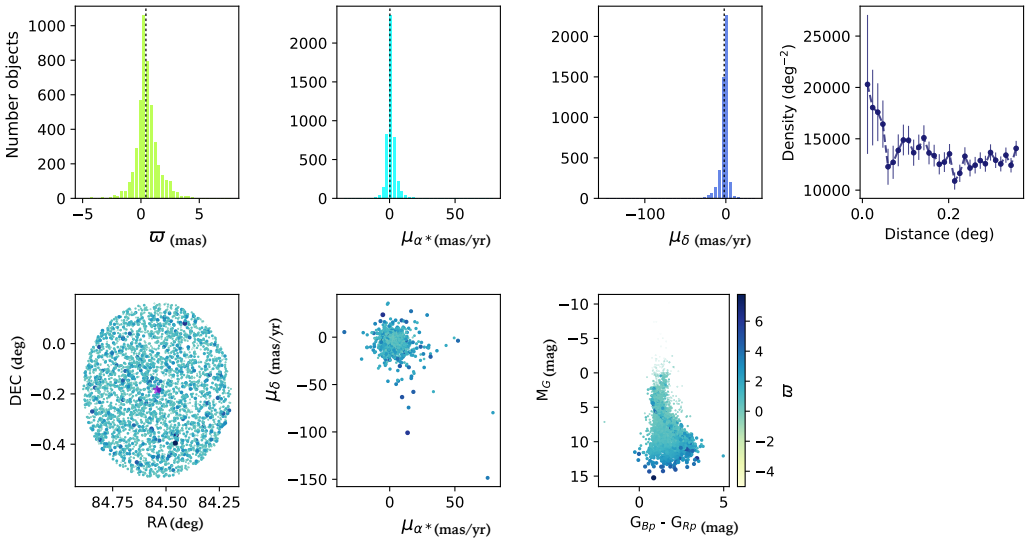


Figure D.1: First iteration of CEREAL for HD37371. The histograms (top panel) represent the astrometric parameters and the black dashed vertical lines show the position of the known values. The density profile is also shown in the top panel. The bottom panel shows the spatial distribution, CMD and the proper motions with a colour code showing how the parallax values are distributed in the three figures. The purple star in the spatial distribution box represents the position of the HAeBe.

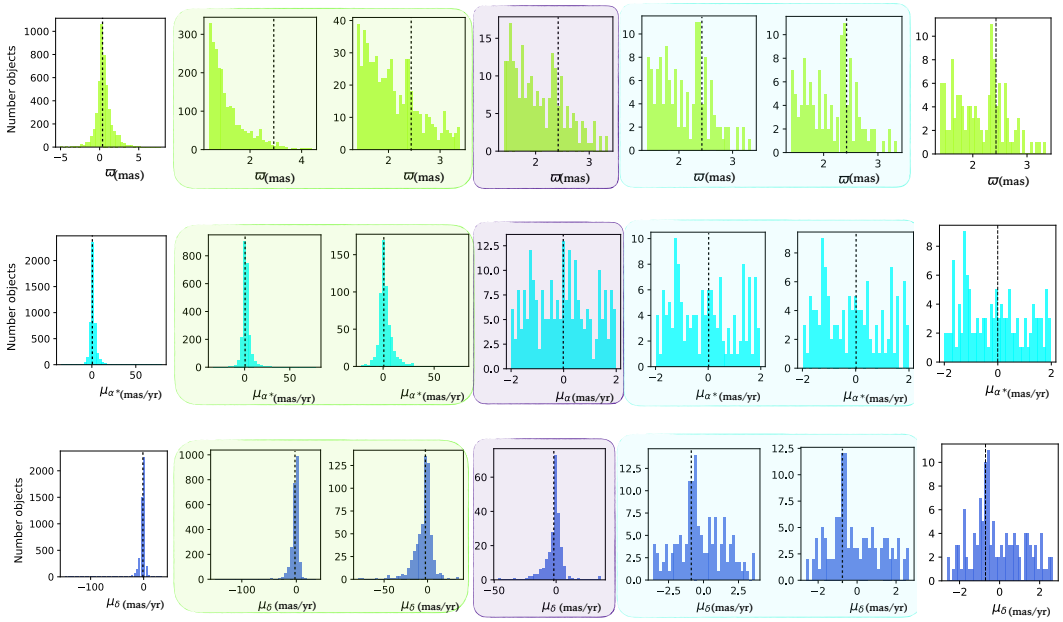


Figure D.2: Step by step of evolution of the astrometric parameters selection made with CEREAL for the HD37371 with the Gaia DR2 data. This shows the expected peak for the astrometric parameter around 2.43 mas in parallax, -0.0034 mas/yr in proper motion in RA and -0.70 mas/yr in proper motion in DEC. The black dashed vertical line shows the known value for each parameter; and, each coloured box represents how the astrometric parameters were affected during the different stages of the selection process made by CEREAL. These coloured boxes represent the selection made in parallax (green), proper motions in RA (light blue) and proper motion in DEC (light purple). From the top to bottom each panel represents an astrometric parameter, on the top the parallax, the middle the proper motion in RA and the bottom the proper motion in DEC.

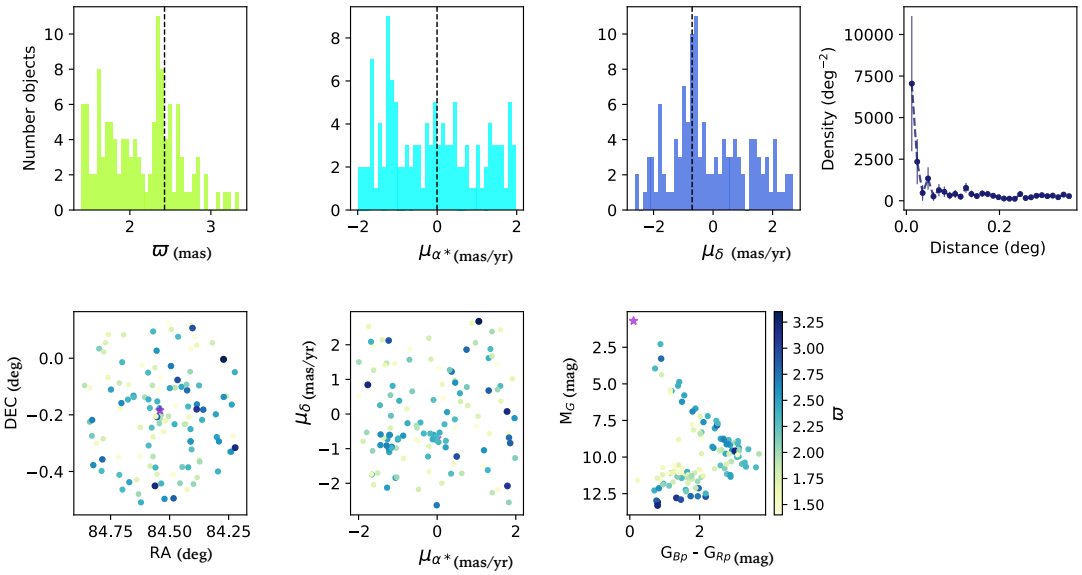


Figure D.3: Images from the final iteration of CEREAL to establish if the stars were a yes, maybe or no. This image represent a *yes* and shows the data for the HAeBe star HD37371. Same label from figure D.1

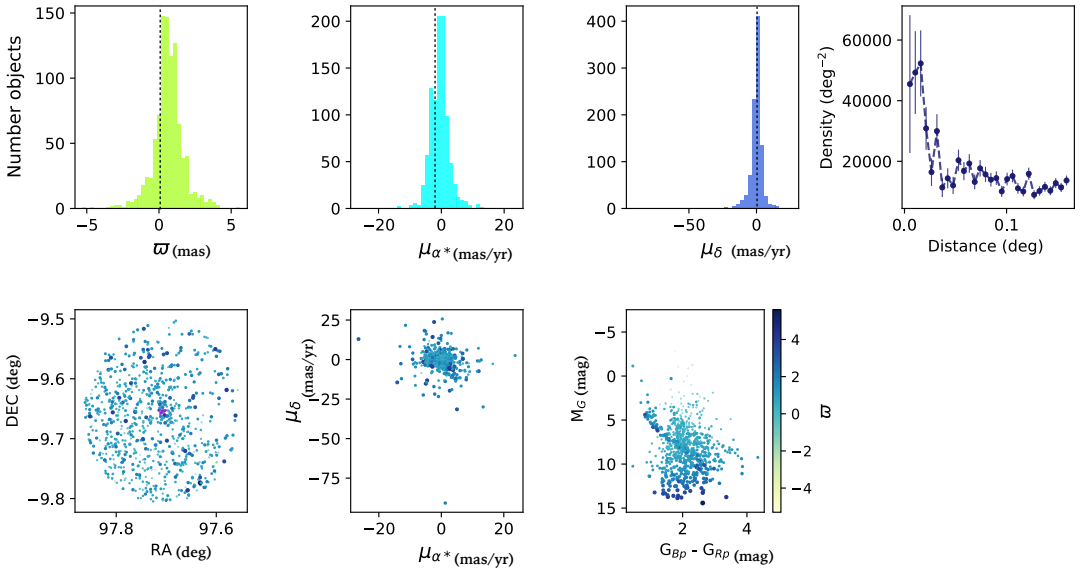


Figure D.4: Images from the first iteration of CEREAL for HD46060. The histograms (top panel) represents the astrometric parameter and the black dashed vertical line shows the position of the known value. The density profile is also given in the top panel. The bottom panel shows the spatial distribution, CMD and a proper motion distribution with a colour code showing how the parallax values were distributed in the three figures. The purple star in the spatial distribution box represents the position of the HAEBE.

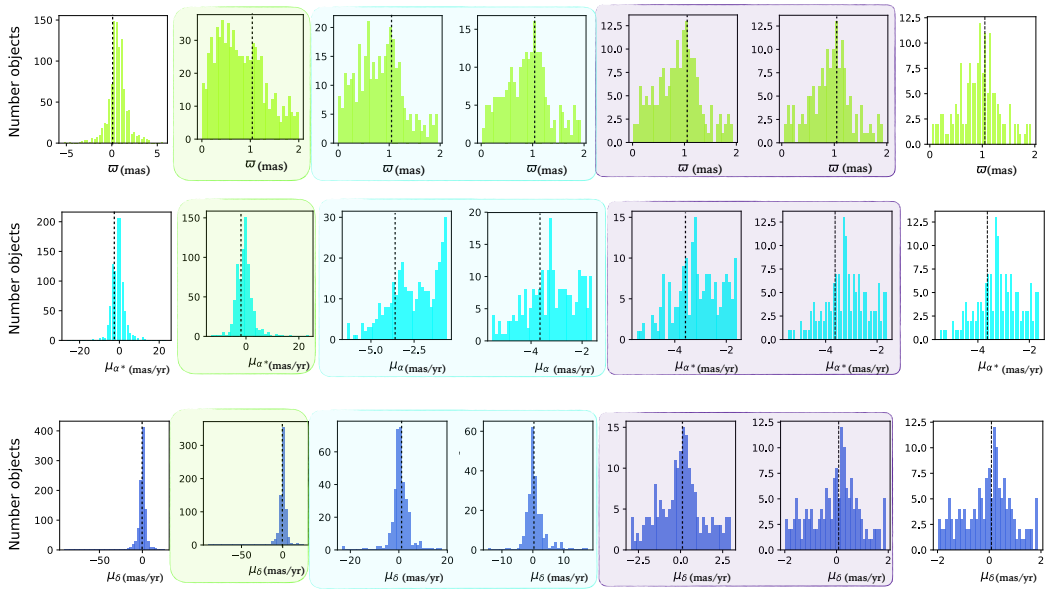


Figure D.5: Step by step progression of the astrometric parameters selections made with CEREAL for the HD46060 with the Gaia DR2 data. This show the expected peak for the astrometric parameter around 1.05 mas in parallax, -3.63 mas/yr in proper motion in RA and 0.08 mas/yr in proper motion in DEC. The black dashed vertical line shows the known value for each parameter; and, each colour box represents how the astrometric parameters are affected during the different stages on the selection process made by CEREAL. The coloured boxes represent the selections made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple). From the top to the bottom each panel represents a different astrometric parameter, on the top the parallax, the middle the proper motion in RA and the bottom the proper motion in DEC.

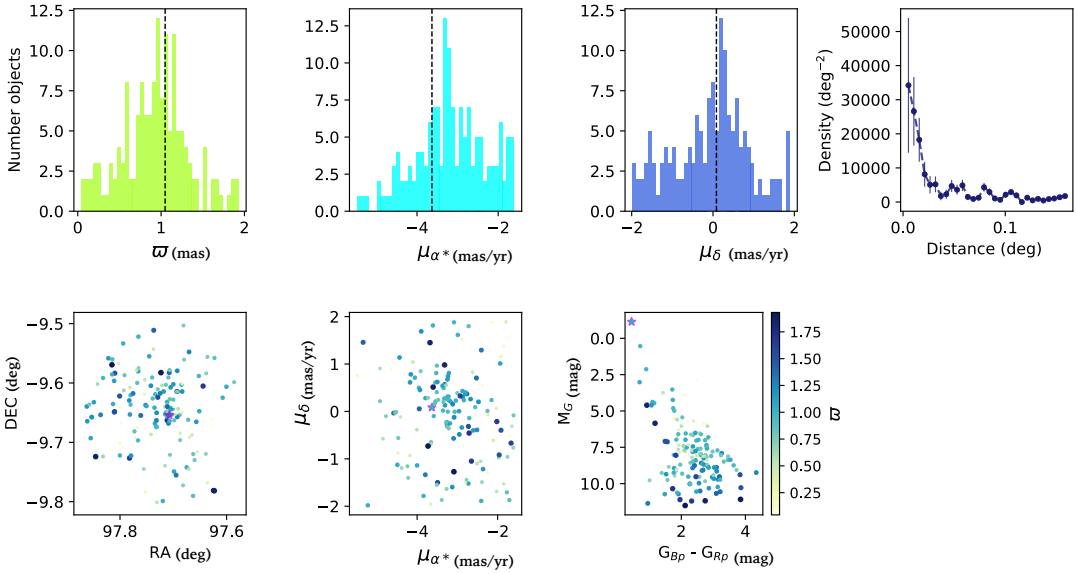


Figure D.6: Images from the final iteration with CEREAL to establish if the stars were a yes, maybe or no. This image represents a *yes* and shows the data for the HAeBe star HD46060. Same labels as used in figure D.4

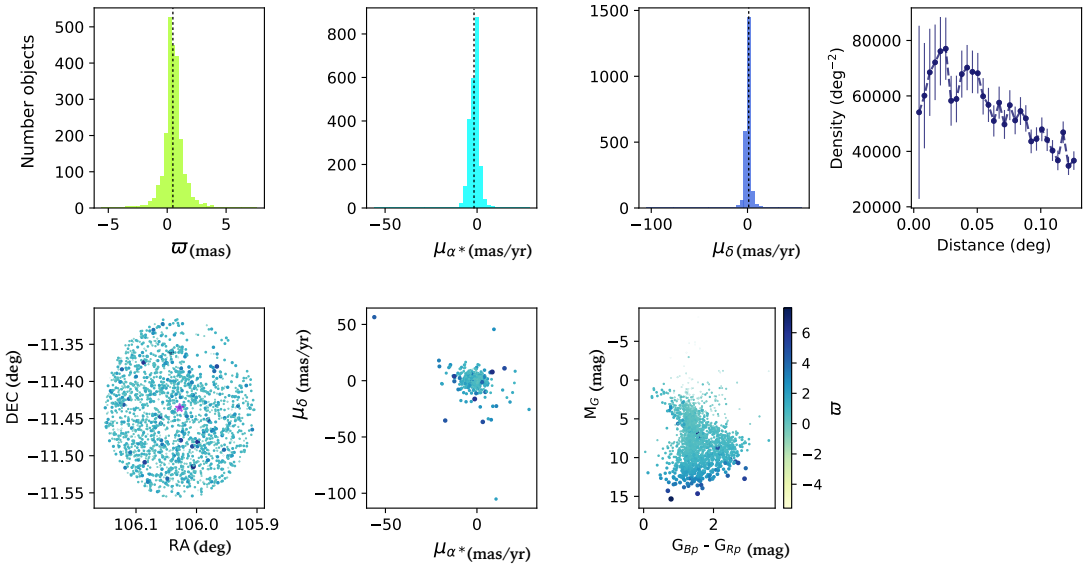


Figure D.7: First iteration of CEREAL for HUCMa. The histograms (top panel) represent the astrometric parameters and the black dashed vertical lines show the position of the known values. The density profile is also given in the top panel. The bottom panel shows the spatial distribution, CMD and a proper motion distribution with a colour code showing how the parallax values are distributed in the three figures. The purple star in the spatial distribution box represents the position of the HAeBe star.

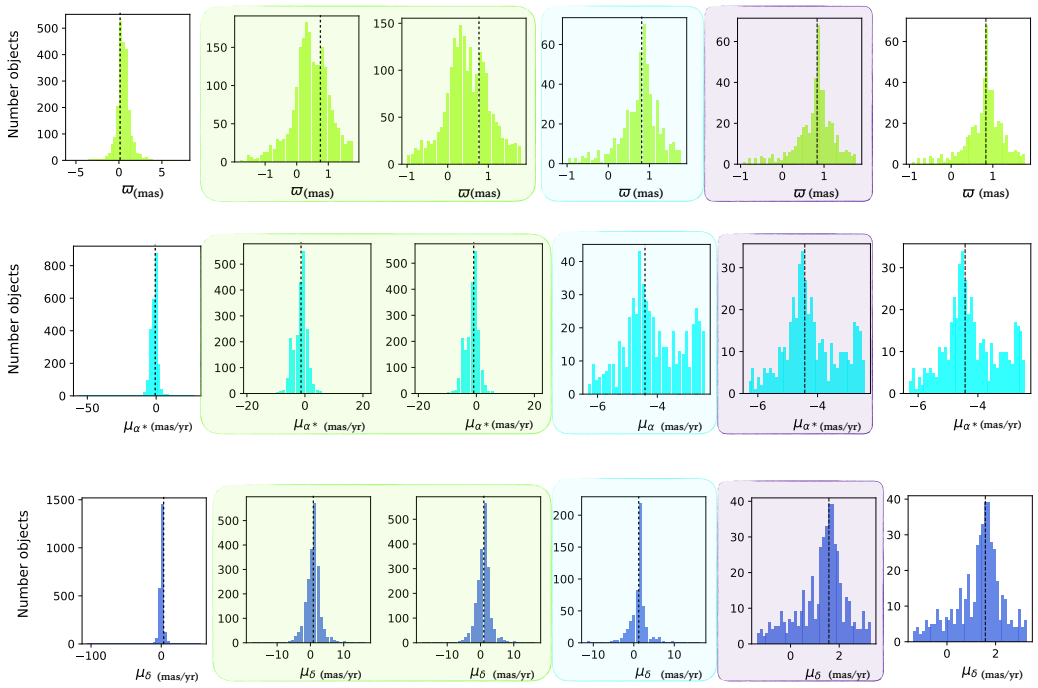


Figure D.8: Step by step progression of the astrometric parameters selections made by CEREAL for HUCMa with the Gaia DR2 data. This shows the expected peak for the astrometric parameter around 0.83 mas in parallax, -4.43 mas/yr in proper motion in RA and 1.58 mas/yr in proper motion in DEC. The black dashed vertical lines show the known value for each parameter; and, each coloured box represents how the astrometric parameters were affected during the different stages of the selection process made with CEREAL. The coloured boxes represent the selections made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple). From the top to bottom each panel represents an astrometric parameter, on the top the parallax, in the middle the proper motion in RA and at the bottom the proper motion in DEC.

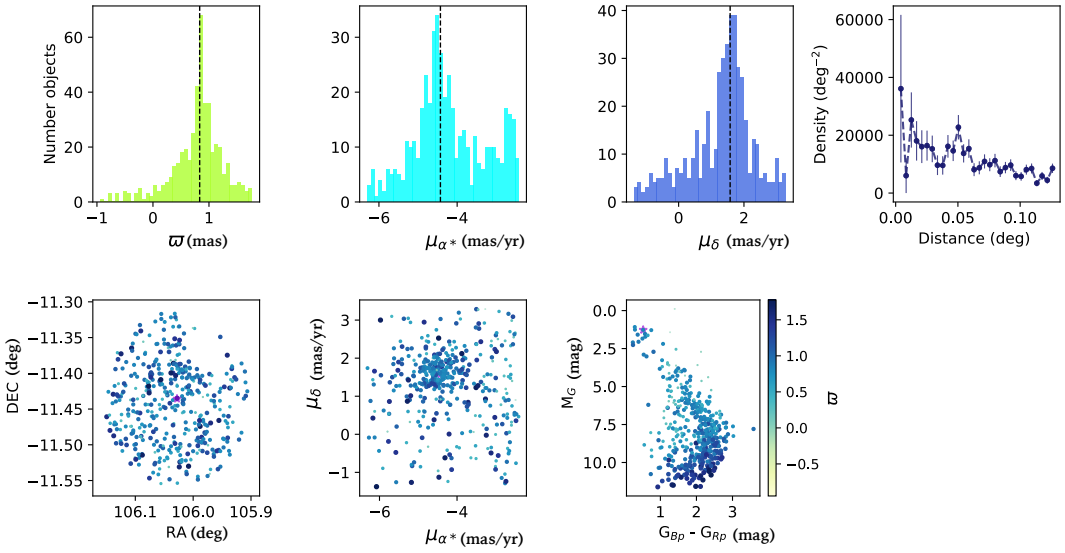


Figure D.9: Images from the final iteration of CEREAL used to establish if the star was a yes, maybe or no. These images represents a *yes* and show the data for the HAeBe star HUCMa. Same labels as used in figure D.7

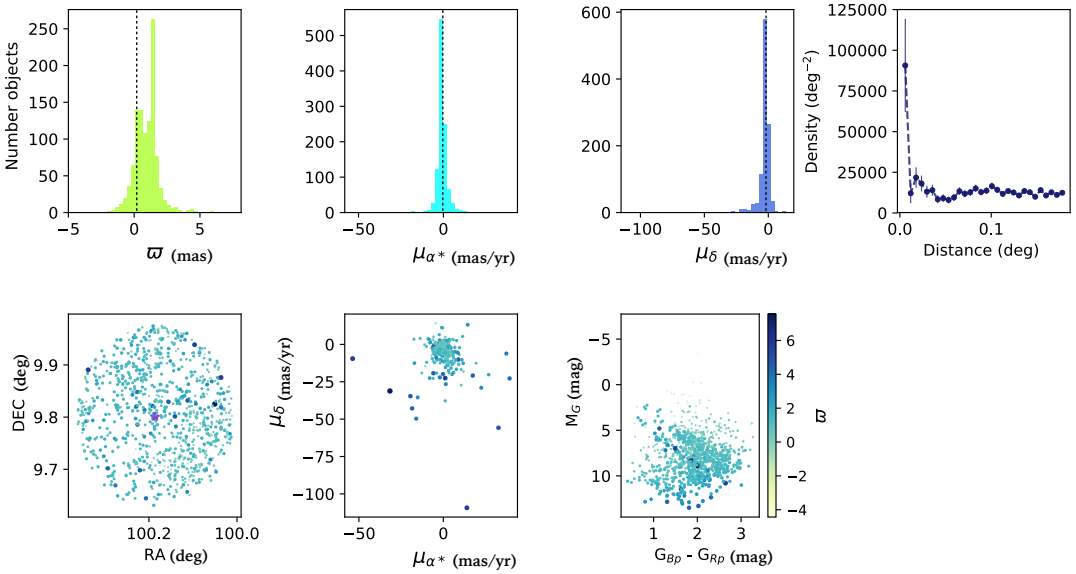


Figure D.10: Images from the first iteration of CEREAL for V590Mon. The histograms (top panel) represent the astrometric parameters and the black dashed vertical lines show the position of the known values. The density profile is also given in the top panel. The bottom panel shows the spatial distribution, CMD and a proper motion distribution with a colour code showing how the parallax values were distributed in the three parts. The purple star in the spatial distribution box represents the position of the HAEBE.

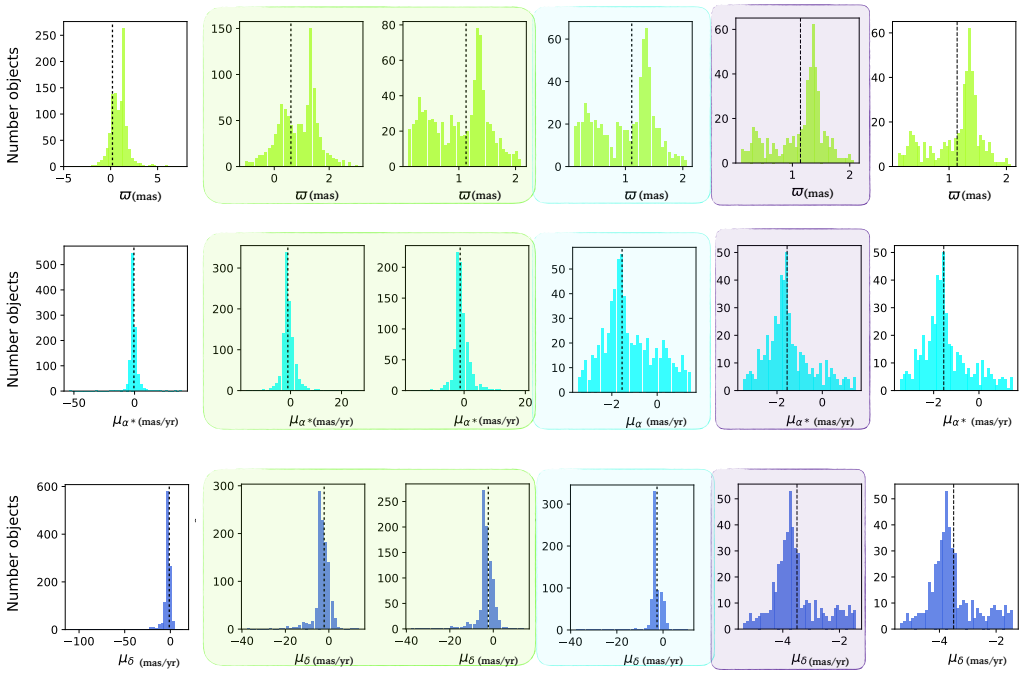


Figure D.11: Step by step progression through the astrometric parameters selection made with CEREAL for V590Mon with the Gaia DR2 data. This shows the expected peak for the astrometric parameter around 1.14 mas in parallax, −1.55 mas/yr in proper motion in RA and −3.50 mas/yr in proper motion in DEC. The black dashed vertical lines show the known values for each parameter; and, each coloured box represents how the astrometric parameters were affected during the different stages of the selection process made with CEREAL. The coloured boxes represent the selections made in parallax (green), proper motion in RA (light blue) and proper motion in DEC (light purple). From the top to bottom each panel represents an astrometric parameter, at the top the parallax, in the middle the proper motion in RA and at the bottom the proper motion in DEC.

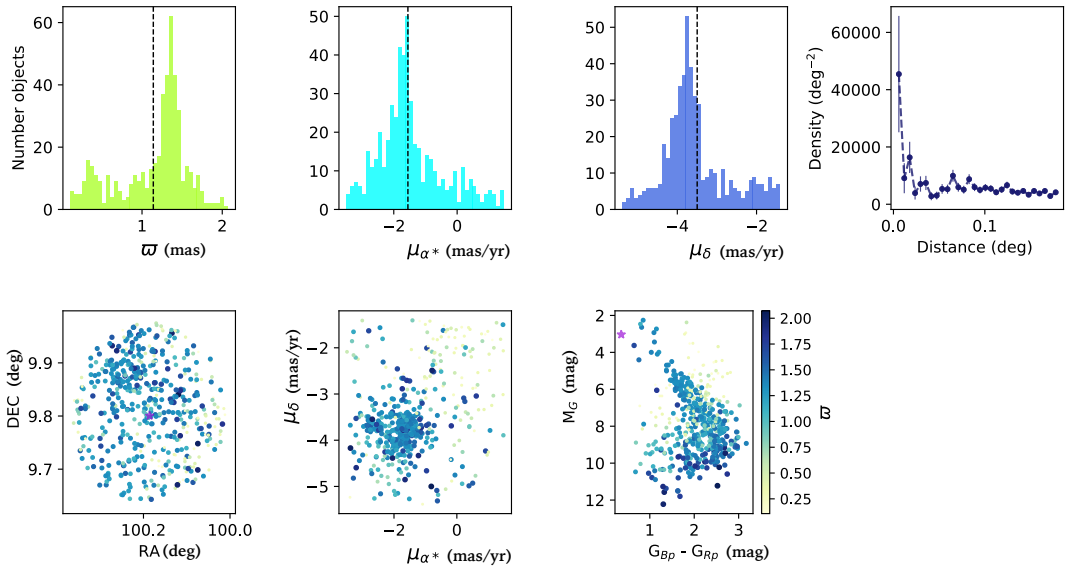


Figure D.12: Images from the final iteration of CEREAL to establish if the stars was a yes, maybe or no. This image represents a *yes* and shows the data for the HAeBe star V590Mon. The same labels are used as in figure D.10

Appendix E

Catalogue of stars defined by CEREAL to not be in a cluster

The following tables represent all the HAeBe stars that were classified by CEREAL to be not found in clusters from section 3.2.4.

Table E.1: Catalogue of Herbig Ae/Be stars not found in clusters by CEREAL

Star	RA h:m:s	DEC deg:m:s	Spectral type	CEREAL	Know Cluster
ABAur	04:55:45.90	+30:33:04.0	A0	NO	Taurus Dark Cloud[1]
AKSco	16:54:44.80	-36:53:19.0	F5	NO	Upper Centaurus Lupus[2]
BHCep	22:01:42.90	+69:44:36.0	F5	NO	
BOCep	22:16:54.10	+70:03:45.0	F4	NO	
BPPsc	23:22:24.70	-02:13:41.0	G9	NO	
CPM25	06:23:56.30	+14:30:28.0	B2	NO	
CQTau	05:35:58.50	+24:44:54.0	F5	NO	
DGCir	15:03:23.80	-63:22:58.0	B5	NO	Circinus cloud[3]
DKCha	12:53:17.20	-77:07:11.0	A0	NO	Cha II in Chan Dark Cloud[4]

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Roccatagliata *et al.* (2020), (2) Luhman & Esplin (2020), (3) Shimoikura & Dobashi (2011), (4) Li *et al.* (2019), (5) Bachiller *et al.* (1998), (6) Soubiran *et al.* (2018), (7) Lee & Lim (2008), (8) Zucker *et al.* (2020), (9) Hara *et al.* (1999), (10) Kauffmann *et al.* (2008), (11) Sen *et al.* (2010),(12) Briceño *et al.* (2019), (13) Barrado *et al.* (2018), (14) Magakian (2003b), (15) Saul (2015)

Table E.2: Table E.1 continued

Star	RA	DEC	Spectral type	CEREAL	Know Cluster
	h:m:s	deg:m:s			
DWCMa	07:19:35.90	-17:39:18.0	B2	NO	
GSC1829-0331	04:30:50.29	+23:00:08.5	F1	NO	Taurus Dark Cloud[1]
GSC1876-0892	06:07:15.40	+29:57:55.0	B2	NO	
GSC4805-1306	06:58:44.36	-03:41:10.0	O9	NO	Iron-clad Nebula[5]
GSC4823-0146	07:08:38.80	-04:19:05.0	B0	NO	
GSC5379-0359	06:49:58.60	-07:38:52.0	B9	NO	
GSC5987-1399	07:31:48.80	-19:27:34.0	B1	NO	
GSC5988-2257	07:41:41.00	-20:00:13.0	B3	NO	
GSC5990-0021	06:48:41.70	-16:48:06.0	B9	NO	
GSC6542-2339	07:24:36.98	-24:34:47.4	O9	NO	
GSC6546-3156	07:24:17.60	-26:16:05.0	A0	NO	
GSC8143-1225	07:59:11.60	-50:22:47.0	F3	NO	
GSC8581-2002	08:44:23.60	-59:56:58.0	A5	NO	
GSC8645-1401	12:17:47.50	-59:43:59.0	F2	NO	
GSC8993-0397	13:03:21.50	-62:13:26.0	B3	NO	
GSC8994-3902	13:19:04.00	-62:34:10.0	B2	NO	HIP 67014 Cluster[6]
GUCMa	07:01:49.51	-11:18:03.3	B2	NO	
HBC1	00:07:02.60	+65:38:38.0	A5	NO	
HBC324	00:07:30.70	+65:39:52.0	A5	NO	
HBC334	02:16:30.71	+55:23:00.1	B3	NO	Per OB1 Association[7]
HBC7	00:43:25.34	+61:38:23.3	B2	NO	
HBC705	20:51:02.70	+43:49:31.0	B0	NO	
HBC717	20:52:06.10	+44:17:16.0	F6	NO	
HD100453	11:33:05.50	-54:19:29.0	A9	NO	NAME ASSOC II SCO[8]
HD100546	11:33:25.30	-70:11:41.0	B9	NO	NAME ASSOC II SCO[8]
HD101412	11:39:44.40	-60:10:28.0	A0	NO	Cha Association[4]
HD104237	12:00:04.90	-78:11:35.0	A4	NO	
HD114981	13:14:40.70	-38:39:06.0	B5	NO	
HD130437	14:50:50.20	-60:17:10.0	B8	NO	
HD132947	15:04:56.10	-63:07:53.0	A0	NO	
HD135344	15:15:48.42	-37:09:16.4	A0	NO	
HD135344B	15:15:48.40	-37:09:16.0	F8	NO	
HD139614	15:40:46.40	-42:29:54.0	A8	NO	Sco OB[9]
HD141569	15:49:57.70	-03:55:17.0	A0	NO	
HD141926	15:54:21.80	-55:19:44.0	B2	NO	
HD142527	15:56:41.90	-42:19:24.0	F6	NO	

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Roccatagliata *et al.* (2020), (2) Luhman & Esplin (2020), (3) Shimoikura & Dobashi (2011), (4) Li *et al.* (2019), (5) Bachiller *et al.* (1998), (6) Soubiran *et al.* (2018), (7) Lee & Lim (2008), (8) Zucker *et al.* (2020), (9) Hara *et al.* (1999), (10) Kauffmann *et al.* (2008), (11) Sen *et al.* (2010),(12) Briceño *et al.* (2019), (13) Barrado *et al.* (2018), (14) Magakian (2003b), (15) Saul (2015)

Table E.3: Table E.1 continued

Star	RA	DEC	Spectral type	CEREAL	Know Cluster
	h:m:s	deg:m:s			
HD142666	15:56:40.00	-22:01:40.0	A8	NO	Sco OB[9]
HD143006	15:58:36.90	-22:57:16.0	G6	NO	Sco OB[9]
HD144432	16:06:57.90	-27:43:10.0	A8	NO	Sco OB[9]
HD149914	16:38:28.60	-18:13:14.0	B9	NO	Sco OB[9]
HD150193	16:40:17.90	-23:53:45.0	A0	NO	Sco OB[9]
HD155448	17:12:58.80	-32:14:34.0	B9	NO	
HD158643	17:31:24.95	-23:57:45.9	A0	NO	
HD163296	17:56:21.30	-21:57:22.0	A0	NO	
HD169142	18:24:29.80	-29:46:50.0	B9	NO	
HD17081	02:44:07.34	-13:51:31.6	B7	NO	
HD174571	18:50:47.20	+08:42:10.0	B2	NO	
HD176386	19:01:38.90	-36:53:27.0	A0	NO	
HD179218	19:11:11.30	+15:47:15.0	A0	NO	
HD199603	20:58:41.78	-14:28:59.5	A9	NO	
HD203024	21:16:03.05	+68:54:52.1	A5	NO	
HD235495	21:21:27.50	+50:59:48.0	A0	NO	LDN 1021[10]
HD244314	05:30:19.00	+11:20:20.0	A1	NO	
HD244604	05:31:57.30	+11:17:41.0	A0	NO	
HD245906	05:39:30.50	+26:19:55.0	A6	NO	
HD249879	05:58:55.80	+16:39:57.0	B8	NO	
HD250550	06:02:00.00	+16:30:57.0	B9	NO	[CB88] 39[11]
HD259431	06:33:05.20	+10:19:20.0	B6	NO	
HD290409	05:27:05.50	+00:25:08.0	B9	NO	
HD290500	05:29:48.10	-00:23:43.0	A2	NO	
HD290764	05:38:05.30	-01:15:22.0	F0	NO	
HD305298	10:33:05.00	-60:19:51.0	O8	NO	
HD313571	18:01:07.20	-22:15:04.0	B3	NO	
HD31648	04:58:46.30	+29:50:37.0	A5	NO	Taurus Dark Cloud[1]
HD319896	17:31:05.80	-35:08:29.0	B4	NO	
HD34282	05:16:00.50	-09:48:35.0	A0	NO	
HD344261	19:21:53.54	+21:31:50.8	F2	NO	
HD34700	05:19:41.40	+05:38:43.0	G0	NO	
HD35187	05:24:01.20	+24:57:37.0	A2	NO	
HD35929	05:27:42.80	-08:19:39.0	A5	NO	
HD36112	05:30:27.50	+25:19:57.0	A8	NO	
HD37357	05:37:47.10	-06:42:30.0	A0	NO	LDN 1641 in Orion A[12]

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Roccatagliata *et al.* (2020), (2) Luhman & Esplin (2020), (3) Shimoikura & Dobashi (2011), (4) Li *et al.* (2019), (5) Bachiller *et al.* (1998), (6) Soubiran *et al.* (2018), (7) Lee & Lim (2008), (8) Zucker *et al.* (2020), (9) Hara *et al.* (1999), (10) Kauffmann *et al.* (2008), (11) Sen *et al.* (2010),(12) Briceño *et al.* (2019), (13) Barrado *et al.* (2018), (14) Magakian (2003b), (15) Saul (2015)

Table E.4: Table E.1 continued

Star	RA	DEC	Spectral type	CEREAL	Know Cluster
	h:m:s	deg:m:s			
HD37411	05:38:14.51	-05:25:13.3	B9	NO	
HD37490	05:39:11.10	+04:07:17.0	B3	NO	
HD38087	05:43:00.60	-02:18:45.0	B5	NO	
HD38120	05:43:11.90	-04:59:50.0	A0	NO	
HD39014	05:44:46.40	-65:44:08.0	A7	NO	
HD41511	06:04:59.10	-16:29:04.0	A2	NO	
HD45677	06:28:17.42	-13:03:11.1	B2	NO	
HD50083	06:51:45.80	+05:05:04.0	B3	NO	
HD50138	06:51:33.40	-06:57:59.0	B8	NO	
HD53367	07:04:25.50	-10:27:16.0	B0	NO	
HD56895B	07:18:31.80	-11:11:34.0	F2	NO	
HD58647	07:25:56.10	-14:10:44.0	B9	NO	
HD59319	07:28:36.80	-21:57:49.0	B8	NO	
HD72106	08:29:34.89	-38:36:21.0	A0	NO	
HD76534	08:55:08.70	-43:28:00.0	B2	NO	
HD85567	09:50:28.50	-60:58:03.0	B2	NO	
HD87643	10:04:30.27	-58:39:52.0	B3	NO	
HD94509	10:53:27.20	-58:25:24.0	B8	NO	
HD95881	11:01:57.60	-71:30:48.0	A0	NO	
HD9672	01:34:37.88	-15:40:34.9	A1	NO	
HD98922	11:22:31.70	-53:22:11.0	B9	NO	
HKOri	05:31:28.05	+12:09:10.1	A2	NO	Barnard 30[13]
Hen3-1121S	15:58:09.66	-53:51:35.3	B0	NO	
Hen3-1191	16:27:15.10	-48:39:26.0	B0	NO	
Hen3-1475	17:45:14.18	-17:56:46.9	B7	NO	
Hen3-373	10:10:00.30	-57:02:07.0	B2	NO	
Hen3-823	12:48:42.40	-59:54:35.0	B3	NO	
Hen3-847	13:01:17.80	-48:53:19.0	B5	NO	
Hen3-938	13:52:42.90	-63:32:49.0	O9	NO	
IPPer	03:40:47.00	+32:31:54.0	A6	NO	
KKOph	17:10:08.10	-27:15:19.0	A5	NO	
LkHa208	06:07:49.50	+18:39:26.0	A7	NO	NGC 2163[14]
LkHa259	23:58:41.60	+66:26:13.0	A9	NO	
MQCas	00:09:37.50	+58:13:10.0	A0	NO	
MWC314	19:21:33.97	+14:52:56.8	B3	NO	
MWC342	20:23:03.60	+39:29:50.0	B0	NO	

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Roccatagliata *et al.* (2020), (2) Luhman & Esplin (2020), (3) Shimoikura & Dobashi (2011), (4) Li *et al.* (2019), (5) Bachiller *et al.* (1998), (6) Soubiran *et al.* (2018), (7) Lee & Lim (2008), (8) Zucker *et al.* (2020), (9) Hara *et al.* (1999), (10) Kauffmann *et al.* (2008), (11) Sen *et al.* (2010),(12) Briceño *et al.* (2019), (13) Barrado *et al.* (2018), (14) Magakian (2003b), (15) Saul (2015)

Table E.5: Table E.1 continued

Star	RA	DEC	Spectral type	CEREAL	Know Cluster
	h:m:s	deg:m:s			
MWC593	17:49:10.20	-24:14:21.0	B4	NO	
MWC623	19:56:31.50	+31:06:20.0	B0	NO	
MWC655	22:38:31.80	+55:50:05.0	B1	NO	
MWC657	22:42:41.80	+60:24:00.0	B0	NO	
MWC778	05:50:13.90	+23:52:17.7	B1	NO	
MWC878	17:24:44.70	-38:43:51.0	B1	NO	
MWC930	18:26:25.20	-07:13:18.0	B0	NO	
MWC953	18:43:28.40	-03:46:17.0	B2	NO	
NSV2968	06:26:53.90	10:15:35.0	B0	NO	
NXPup	07:19:28.30	-44:35:11.0	A0	NO	
PDS002	01:17:43.50	-52:33:31.0	F3	NO	
PDS004	03:39:00.60	+29:41:46.0	A0	NO	
PDS021	06:02:14.90	-10:00:59.0	B5	NO	
PDS022	06:03:37.10	-14:53:03.0	A0	NO	
PDS025	06:54:27.90	-25:02:16.0	A3	NO	
PDS123	05:50:54.77	+20:14:47.7	B0	NO	
PDS124	06:06:58.50	-05:55:07.0	A0	NO	
PDS133	07:25:04.90	-25:45:49.0	B6	NO	
PDS138	11:53:13.20	-62:05:21.0	B0	NO	
PDS144S	15:49:15.31	-26:00:55.1	A5	NO	
PDS211	06:10:17.30	+29:25:17.0	B9	NO	
PDS229	06:55:40.00	-03:09:50.0	A0	NO	
PDS24	06:48:41.68	-16:48:05.5	B9	NO	
PDS286	09:06:00.00	-47:18:58.0	B0	NO	
PDS290	09:26:11.10	-52:42:27.0	B7	NO	
PDS297	09:42:40.30	-56:15:34.0	A7	NO	
PDS322	10:52:08.68	-56:12:06.8	B3	NO	
PDS33	08:48:45.70	-40:48:21.0	A0	NO	
PDS34	08:49:58.50	-45:53:06.0	B2	NO	
PDS344	11:40:32.80	-64:32:06.0	B5	NO	
PDS364	13:20:03.60	-62:23:54.0	B2	NO	
PDS371	13:47:31.40	-36:39:50.0	O9	NO	
PDS389	15:14:47.05	-62:16:59.8	A3	NO	
PDS394	15:35:17.13	-61:59:04.2	F0	NO	
PDS406	16:05:03.91	-39:45:03.8	A5	NO	
PDS431	16:54:59.20	-43:21:50.0	A0	NO	

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Roccatagliata *et al.* (2020), (2) Luhman & Esplin (2020), (3) Shimoikura & Dobashi (2011), (4) Li *et al.* (2019), (5) Bachiller *et al.* (1998), (6) Soubiran *et al.* (2018), (7) Lee & Lim (2008), (8) Zucker *et al.* (2020), (9) Hara *et al.* (1999), (10) Kauffmann *et al.* (2008), (11) Sen *et al.* (2010),(12) Briceño *et al.* (2019), (13) Barrado *et al.* (2018), (14) Magakian (2003b), (15) Saul (2015)

Table E.6: Table E.1 continued

Star	RA	DEC	Spectral type	CEREAL	Know Cluster
	h:m:s	deg:m:s			
PDS453	17:20:56.10	-26:03:31.0	F2	NO	
PDS469	17:50:58.10	-14:16:12.0	A0	NO	
PDS477	18:00:30.30	-16:47:26.0	B1	NO	
PDS530	18:41:34.40	+08:08:21.0	F0	NO	
PDS543	18:48:00.70	+02:54:17.0	B1	NO	
PDS581	19:36:18.90	+29:32:50.0	B0	NO	
PVCep	20:45:53.98	+67:57:38.6	A5	NO	Cep Flare[8]
PXVul	19:26:40.30	+23:53:51.0	F3	NO	
RCrA	19:01:53.69	-36:57:08.6	A5	NO	IRAS 18585-3701[15]
RMon	06:39:09.95	+08:44:09.6	B8	NO	
RYOri	05:32:09.90	-02:49:47.0	F6	NO	
SAO185668	17:43:55.60	-22:05:45.0	B3	NO	
SAO220669	08:55:45.90	-44:25:14.0	B4	NO	
SVCep	22:21:33.20	+73:40:27.0	A0	NO	
TCrA	19:01:58.79	-36:57:50.3	F0	NO	IRAS 18585-3701[15]
TYCrA	19:01:40.80	-36:52:34.0	B9	NO	
UXOri	05:04:30.00	-03:47:14.0	A2	NO	LDN 1616[8]
V1012Ori	05:11:36.55	-02:22:48.5	A3	NO	
V1295Aql	20:03:02.50	+05:44:17.0	A0	NO	
V1493Cyg	20:52:04.60	+44:37:30.0	A2	NO	
V1686Cyg	20:20:29.40	+41:21:28.0	A4	NO	
V1818Ori	05:53:42.60	-10:24:01.0	B7	NO	
V351Ori	05:44:18.80	+00:08:40.0	A7	NO	
V374Cep	23:05:07.50	+62:15:36.0	B4	NO	
V375Lac	22:34:40.99	+40:40:04.3	A7	NO	
V376Cas	00:11:26.56	+58:50:03.8	B5	NO	
V431Sct	18:29:25.70	-06:04:37.0	B1	NO	
V633Cas	00:11:26.50	+58:49:29.0	B9	NO	
V645Cyg	21:39:58.30	+50:14:21.0	O7	NO	
V669Cep	22:26:38.70	+61:13:32.0	B5	NO	
V718Sco	16:13:11.60	-22:29:07.0	A8	NO	Upper Scorpius in Sco OB II[8]
V883Ori	05:38:18.10	-07:02:26.0	B0	NO	LDN 1641 in Orion A[12]
V892Tau	04:18:40.60	+28:19:16.0	A0	NO	Taurus Dark Cloud[1]
V921Sco	16:59:06.80	-42:42:08.0	B0	NO	
VVSer	18:28:47.90	+00:08:40.0	B6	NO	Serpens Cloud[8]
VXCas	00:31:30.70	+61:58:51.0	A0	NO	
VYMon	06:31:06.90	+10:26:05.0	B5	NO	
WRAY15-1435	16:13:06.70	-50:23:20.1	B1	NO	
XYPere+w	03:49:36.35	+38:58:55.4	A2	NO	
ZCMa	07:03:43.20	-11:33:06.0	B5	NO	

Reference of the *Known Cluster* taken from SIMBAD (Wenger *et al.*, 2000): (1) Roccatagliata *et al.* (2020), (2) Luhman & Esplin (2020), (3) Shimoikura & Dobashi (2011), (4) Li *et al.* (2019), (5) Bachiller *et al.* (1998), (6) Soubiran *et al.* (2018), (7) Lee & Lim (2008), (8) Zucker *et al.* (2020), (9) Hara *et al.* (1999), (10) Kauffmann *et al.* (2008), (11) Sen *et al.* (2010),(12) Briceño *et al.* (2019), (13) Barrado *et al.* (2018), (14) Magakian (2003b), (15) Saul (2015)

Appendix F

Machine Learning algorithms

In recent years a wide variety of different methodologies have begun to be implemented for the analysis of the large volume of data available. More specific, the high volume of astronomical data. These techniques include *machine learning* algorithms, which is the field of study that gives computers the ability to learn without being explicitly programmed (Alpaydin, 2014; Géron, 2019).

In the literature it was possible to find a large number of machine learning algorithms depending on the analysis that needed to be performed. *Clustering* is a branch within the entire machine learning which only needs the data to be analyzed to work. Within the clustering methods, it was also possible to identify a variety of algorithms which studied overdensities in large samples. The *Density-based clustering* algorithms (DCAs) are an example of the clustering methods that analyse over densities. In these algorithms, the over densities represent the clusters in the sample (Pedregosa *et al.*, 2011). Additionally, the clusters from the DCAs are defined as an area of higher density with an arbitrary shape with respect to the remainder of the data (Pedregosa *et al.*, 2011).

To process the data, the DCAs required two main parameters as an input. These parameters were:

* ϵ *or eps-distance*, which is the distance between two points in the N-dimensional space; and,

* *mPts* which is the minimum number of points required to form a dense region.

To define a cluster, the DCAs needed to establish a *core* and *border* points, which are represented in figure F.1. The cores were the points which had at least mPts of neighbours within a distance d $= \epsilon$. The border points were separated by a distance d $\leq \epsilon$ from a core point and had a number of neighbours n $<$ mPts (i.e., they are less dense than core points; Ankerst *et al.*, 1999; Ester *et al.*, 1996). The noise points did not belong to the cluster (Ester *et al.*, 1996). These parameters can be modified at any point during the process of finding the clusters, to improve the results.

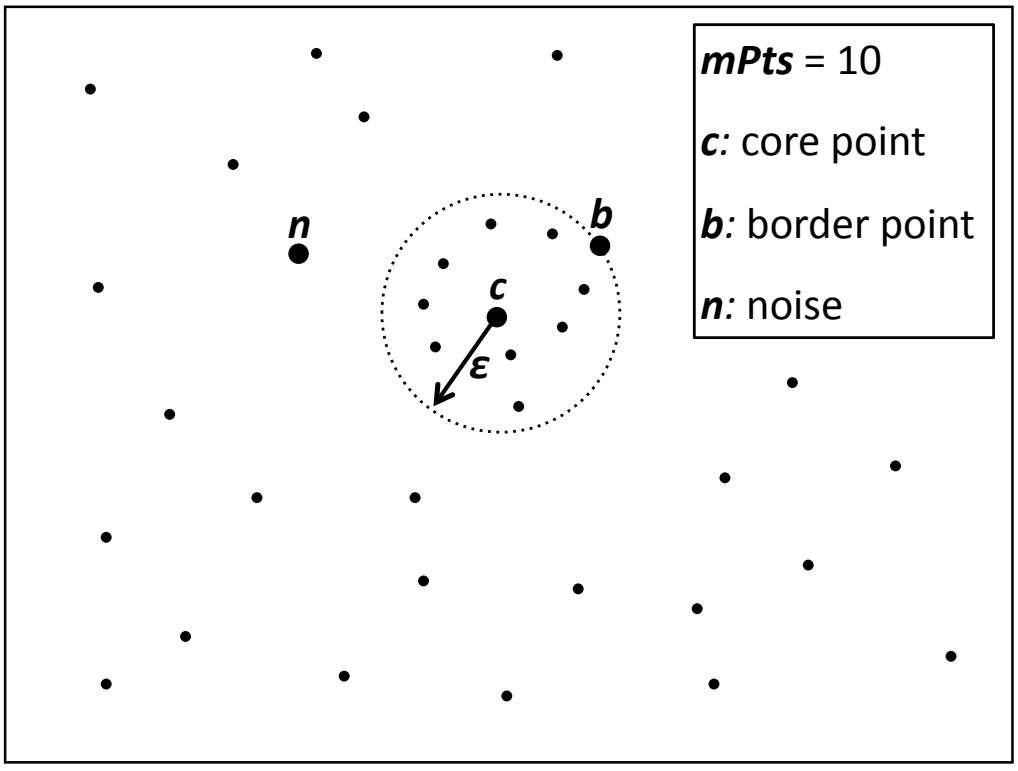


Figure F.1: Fundamental parameters used by DBSCAN and OPTICS to find cluster. ϵ *and mPts* are the main parameters that can be modified during the process of finding clusters. The *core* and *border* points define the limit of the cluster. See the text for more information.

For this study, three density-based clustering algorithms (DCAs) were chosen from the literature to compare their efficiency in finding clusters with `CEREAL` (see §3.3). The DCAs selected were:

* **DBSCAN.** This code views clusters as areas of high density separated by areas of low density which can have any shape (Pedregosa *et al.*, 2011). This gives them the advantage of separating clusters of high density in contrast with clusters of low density within the same data (Ester *et al.*, 1996).

* **OPTICS.** As DBSCAN, finds clusters in high density areas and is capable of distinguishing between the clusters densities. However, if the data contains clusters of different densities it would be difficult to detect all the members of an each cluster (Ankerst *et al.*, 1999).

* **HDBSCAN.** OPTICS, can identify clusters with different densities and is sensitive to the density gradients inside a cluster (Campello *et al.*, 2015; McInnes *et al.*, 2017). This method only uses one parameter called *the minimum clusters size (mCls)* to determinate the presence of clusters, which is similar to *mPts* defined previously.

Appendix G

CEREAL and DCAs: Comparison and quality condition samples

This appendix shows how the raw data from Gaia DR2 is affected when different quality conditions are applied to it. The quality conditions applied have been described in sections § 3.3 and § 3.3.3. These condition were: RUWE $\leq$ 1.40 and parallax signal to noise $>$ 10.

Table G.1 show the number of stars used as an input in this analysis, specifically column 3 and 7 shows the original CEREAL sample and column 4 and 8 the sample used in §3.3.3.

This appendix also shows the number of stars found by each of the DCAs (*DBSCAN*, *OPTICS* and *HDBSCAN*; Ankerst *et al.*, 1999; Campello *et al.*, 2015; Ester *et al.*, 1996; McInnes *et al.*, 2017, respectively) used in section § 3.3 in table G.2. The table shows the combined results found by the DCAs and the stars found in common and not in common by each of the DCAs. The combined results of the DCAS (column 5 and 12) were the ones used to compare with the data found by CEREAL.

Table G.1: Number of stars from Gaia DR2 catalogue by applying different quality conditions

Star	Gaia raw	Gaia RUWE	Gaia RUWE/$\frac{\varpi}{\sigma_\varpi}$	Star	Gaia raw	Gaia RUWE	Gaia RUWE/$\frac{\varpi}{\sigma_\varpi}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A097415	3389	3175	454	HTCMa	1969	1898	146
AS310	1681	1571	71	HUCMa	2689	2487	169
AS477	6628	6318	431	Hen3-1121	2822	2665	157
BD+30549	3919	3773	333	Hen3-949	830305	804248	46103
BD+413731	1091	990	178	ILCep	2561	2393	176
BFOri	1520	1375	324	LKHa260	2332	1605	175
COOri	6357	6112	459	LKHa338	1030	973	126
GSC3975-0579	9743	9310	436	LkHa257	6731	6317	422
GSC5360-1033	2030	1937	233	LkHa324	7930	7525	386
HBC217	1900	1714	307	LkHa339	1132	1074	123
HBC222	1753	1549	321	PDS126	1924	1833	119
HBC442	2485	1773	766	PDS129	4476	4208	178
HBC694	2756	2573	430	PDS324	1607	1515	51
HD200775	7202	6855	770	PDS415N	110480	107251	3441
HD245185	7220	6925	592	PDS520	10830	10279	232
HD287823	7383	6295	633	TOri	2904	1931	845
HD288012	7068	6724	671	V1685Cyg	1191	1117	179
HD290380	6466	6192	699	V1787Ori	1586	1445	241
HD290770	6373	6019	589	V346Ori	7260	6909	782
HD36917	1920	1091	656	V361Cep	2115	1958	190
HD36982	2648	1641	886	V373Cep	1868	1716	172
HD37371	5796	5231	515	V380Ori	846	773	213
HD46060	1147	1087	106	V590Mon	1404	1247	242
HD87403	2116	1952	153	V594Cas	10590	10107	651
HD96042	2398	2212	50	WWVul	99683	92722	1122
HD97048	39227	37810	3237				

- Column 2 and 6: raw data from Gaia DR2
- Column 3 and 7: Gaia DR2 data reduced with the RUWE ≤ 1.40
- Column 4 and 8: Gaia DR2 data reduced with the RUWE ≤ 1.40, $\frac{\varpi}{\sigma_\varpi} > 10$

Table G.2: Number of stars found by each DCas

Star	DBSCAN	OPTICS	HDBSCA		All DCAs		Star	DBSCAN	OPTICS	HDBSCA		All DCAs	
				Combined	Common	Not common					Combined	Common	Not common
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
A097415	17	33	16	33	16	17	HTCMa	9	17	15	21	7	14
AS310	6	6	5	11	3	8	HUCMa	40	24	37	40	24	16
AS477	26	45	17	45	17	28	Hen31121	9	13	8	18	3	15
BD+30549	40	53	43	53	40	13	Hen3949	15	19	10	22	9	13
BD+413731	21	28	8	30	8	22	ILCep	10	24	7	24	7	17
BFOri	40	122	88	122	40	82	LKHa260	21	31	11	32	11	21
COOri		84	17	85	16	69	LKHa338	8	14	5	14	5	9
GSC3975-0579	18	76	41	76	16	60	LkHa257			5	5	5	
GSC5360-1033	6	11	8	11	6	5	LkHa324	27	38	16	38	16	22
HBC217		7	9	11	5	6	LkHa339	15	14	11	16	11	5
HBC222		9	6	9	6	3	PDS126	4	10	3	10	3	7
HBC442			29	29	29		PDS129	8	20	11	20	5	15
HBC694	11	11	35	35	9	26	PDS324	14	16	22	26	10	16
HD200775	18	31	13	31	13	18	PDS415N			64	64	64	
HD245185	60	102	20	102	20	82	PDS520	11	11	19	22	7	15
HD287823	15	35	16	35	14	21	TOri	6	8	5	9	5	4
HD288012			11	11	11		V1685Cyg	13	22	12	24	9	15
HD290380			8	8	8		V1787Ori	66	62	41	70	41	29
HD290770	26	111	70	115	20	95	V346Ori	12	15	36	43	8	35
HD36917	97	164	195	231	77	154	V361Cep	12	16	10	16	10	6
HD36982		17	7	17	7	10	V373Cep	12	16	19	19	12	7
HD37371	16	11	10	17	8	9	V380Ori	21	33	11	33	11	22
HD46060	12	21	22	22	12	10	V590Mon	12	11	12	16	7	9
HD87403	14	31	9	31	9	22	V594Cas	10	8	11	12	8	4
HD96042	7	21	10	22	6	16	WWVul	18	52	13	52	13	39
HD97048	30	88	22	88	22	66							

Appendix H

Mass magnitude relation

For the analysis in section 5.2 a mass ratio (mass-magnitude) relation was created to find the lower mass limit detected for each cluster, using the mass of the full sample of HAeBe stars (with has masses reported by Vioque *et al.*, 2018, see §3.2) and their absolute magnitudes.

The absolute magnitudes of the HAeBe stars used to find the mass ratio relation were corrected by extinction, as it is known that the dust present along the line of sight towards the stars leads to a dimming and reddening of their observed light (Gaia Collaboration *et al.*, 2018a). This decision was taken even though the extinction values were not available for the low mass companions (possible cluster members) as it is probable that the extinction for the HAeBe star is higher than the possible cluster members.

Figure H.1 shows the relation between the mass and absolute magnitude of the HAeBe stars. The stars show a relation close to a straight line that can be represented as $log(M) = m * M_G + c$, with M in solar units. The top panel show a value of M_G not corrected by A_G even if the data show a large scatter, the stars show a relation close to a straight line where m and c as -0.124 ± 0.007 and 0.700 ± 0.018, respectively. The bottom panel, show a tighter distribution around the fit for the sample, where M_G was corrected by A_G using the extinction relation found by Wang & Chen (2019, $A_G = 0.789 A_V$); the A_V for the Herbig stars were taken from Vioque *et al.* (2018). This data also appeared to be very close to linearity and was fitted with a straight line where m and c as -0.131 ± 0.002 and 0.532 ± 0.006, respectively.

The bottom panel of figure H.1 shows the lowest scatter around the fit to the data and the general importance of the correction for extinction, since all the stars look less faint compared to the top panel of figure H.1 but even with the correction the slope (m) in both figures are very similar. The small variation of the slope (compared to the change in ΔM_G) between the corrected and uncorrected data will not significantly change the result for the mass ratio found, $\frac{M}{M_{HAeBe}} = 10^{m\Delta M_G}$, with ΔM_G is the magnitude difference between the object in the cluster and the Herbig Ae/Be star. However, the change in ΔM_G will lead to a much larger minimum mass than those calculated in section 5.2.2.

To asses the influence that correcting only the absolute magnitude of the HAeBe

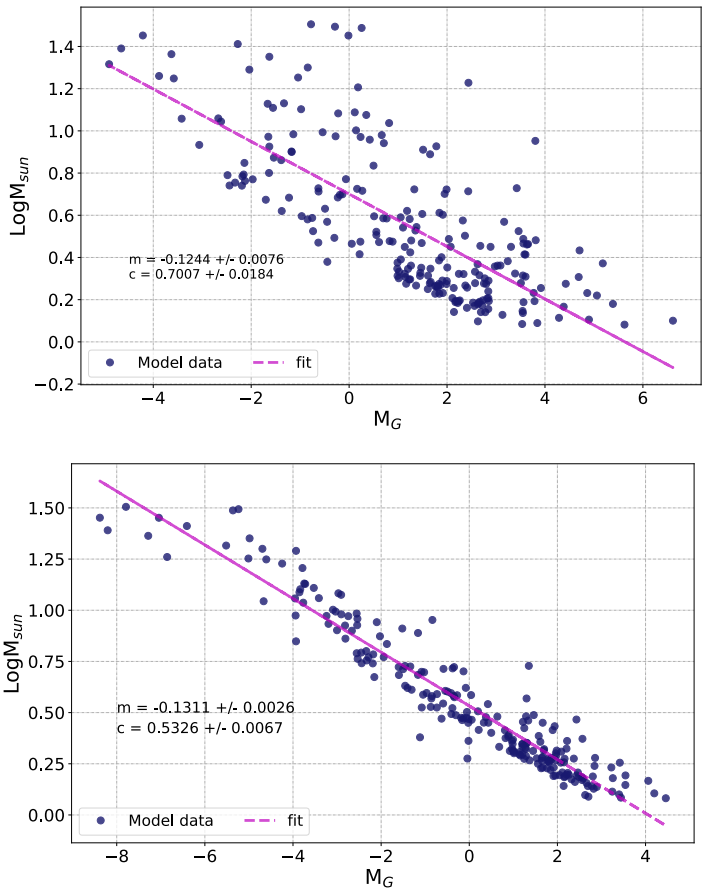


Figure H.1: Mass magnitude relation for a HAeBe star. The full blue represents he full sample of HAeBe stars (see §3.2) and their absolute magnitude. The data show a relation close to a straight line (magenta dashed line) which is represent as $log(M) = m * x + c$. *Top panel:* data not corrected by A_G *Bottom panel:* data corrected by A_G.

star for extinction could have on the results obtained in section 5.2.2 ΔM_G was also calculated using the uncorrected absolute magnitudes of the HAeBe stars. Using the M_G not corrected by A_G allowed a direct comparison for the possible cluster members found around the HAeBe stars since these companions do not have a known value of A_G. When the M_G of the HAeBe star is corrected by A_G and the M_G of the possible cluster members are not, then the value of ΔM_G will be larger than if both were uncorrected. As it is probable that the HAeBe star has higher extinction that its possible cluster members then when M_G of the HAeBe star and not that of its low mass companion will produce a ΔM_G that is more representative of the true difference in magnitudes than using the difference in the uncorrected magnitudes. Although, Andrae *et al.* (2018) state that, due to large uncertainties, extinction corrections are less useful at the individual star level but have good performance for correcting large samples of stars, as can be seen in the improved correlation in figure H.1. Also, it has been shown that inside clusters a large spread of A_G can be found (Arenou *et al.*, 2018, see their figure 47). Future analysis of A_G, similar to the work carried out by Vioque *et al.* (2020), for both the HAeBe stars and their companions would allow for improved estimation of the lower mass limit detected for each cluster and would improve the characterisation of these samples.

The slope from the plot of the masses of the HAeBe stars against their uncorrected absolute magnitudes (top panel on figure H.1) was used in conjunction with the uncorrected ΔM_G to generate minimum masses for all the clusters. These minimum masses were then used to generate IMF correction factors and corrected stellar densities. The results of this are presented in table H.1 and figure H.2 which are the equivalents of table 5.1 and figure 5.10 in sections 5.2.6 and 5.2.7 respectively. The key difference between these plots is in the absolute values for the stellar densities calculated with the IMF correction. With the minimum mass found in section 5.2.2 after correcting the M_G of the HAeBe stars for extinction, the IMF correction accounted for $>84\%$ of the total stellar density observed. When the uncorrected M_G were used to calculate ΔM_G the proportion of the stellar densities that comes from the IMF correction increases to $>95\%$ of the overall stellar densities calculated.

Even with the higher densities found without applying the extinction corrections to the HAeBe stars the results differ from the work of Bonnell *et al.* (2003). As discussed in section 5.2.7 there are elements of the model that are replicated. There is again an early time increase in density followed by a general trend for the stellar densities to decrease

in age. The magnitude of these changes are however still not in good agreement with the results of Bonnell *et al.* (2003). The stellar densities found here are roughly a factor of 3 higher than those found in section 5.2.7 but the multiple order of magnitude changes in stellar densities (from Bonnell *et al.*, 2003) are not seen in these results.

Table H.1: Summary of the cluster parameters

Star	Spectral type	Distance	Mass	Age	Cluster Radius	Num Stars	Low Mass Limit	IMF ratio	Stellar Density	Complete Stellar Density
		pc	$M_\odot$	Myr	pc		$M_\odot$		stars pc^{-3}	stars pc^{-3}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
A0974-15	B3	$397.6^{+9.6}_{-8.9}$	$2.71^{+0.36}_{-0.23}$	$2.0^{+18.0}_{-1.0}$	0.34 ± 0.26	4 ± 2	$0.64^{+0.1}_{-0.08}$	$23.0^{+16.0}_{-18.0}$	20.0 ± 47.0	500^{+1100}_{-1100}
AS310	B1	$2110.0^{+360.0}_{-240.0}$	$11.9^{+4.8}_{-3.4}$	$0.06^{+0.55}_{-0.04}$	0.387 ± 0.043	9 ± 3	$2.1^{+0.86}_{-0.64}$	$108.0^{+95.0}_{-74.0}$	38.0 ± 18.0	4100^{+4100}_{-3400}
BD+30549	B8	$295.0^{+13.0}_{-11.0}$	$2.28^{+0.37}_{-0.19}$	$5.0^{+15.0}_{-2.0}$	0.73 ± 0.21	18 ± 4	$0.166^{+0.04}_{-0.03}$	$3.3^{+1.5}_{-1.2}$	11.0 ± 10.0	36^{+37}_{-35}
HD36982	B1.5	$408.0^{+19.0}_{-16.0}$	$5.2^{+0.42}_{-0.29}$	$0.73^{+0.47}_{-0.17}$	0.64 ± 0.13	24 ± 5	$0.399^{+0.07}_{-0.067}$	$10.8^{+3.9}_{-3.8}$	22.0 ± 14.0	230^{+170}_{-170}
HD37371	B9	$411.0^{+19.0}_{-16.0}$	$3.85^{+0.63}_{-0.67}$	$0.86^{+0.65}_{-0.34}$	0.67 ± 0.36	6 ± 2	$0.241^{+0.06}_{-0.06}$	$5.4^{+2.4}_{-2.5}$	4.5 ± 7.3	24^{+41}_{-41}
HD46060	B8	$933.0^{+96.0}_{-71.0}$	$9.6^{+3.4}_{-2.4}$	$0.09^{+0.12}_{-0.05}$	1.07 ± 0.3	29 ± 5	$0.45^{+0.18}_{-0.14}$	$12.5^{+9.3}_{-7.3}$	5.7 ± 4.9	71^{+81}_{-74}
HUCMa	B8	$1174.2^{+100.0}_{-77.0}$	$3.02^{+0.15}_{-0.15}$	$2.04^{+0.34}_{-0.15}$	0.81 ± 0.24	25 ± 5	$0.235^{+0.039}_{-0.039}$	$5.3^{+1.9}_{-1.9}$	11.0 ± 10.0	58^{+58}_{-58}
Hen3-949	B3	$643.0^{+33.0}_{-28.0}$	$4.18^{+0.74}_{-0.52}$	$0.8^{+5.6}_{-0.3}$	0.9 ± 0.13	20 ± 4	$0.227^{+0.057}_{-0.05}$	$5.0^{+2.3}_{-2.0}$	6.5 ± 3.1	32^{+22}_{-20}
ILCep	B2	$805.0^{+31.0}_{-27.0}$	$9.9^{+2.7}_{-1.3}$	$0.07^{+0.044}_{-0.033}$	0.63 ± 0.13	32 ± 6	$0.47^{+0.15}_{-0.11}$	$13.2^{+8.1}_{-5.6}$	31.0 ± 19.0	410^{+350}_{-310}
LKHa338	B9	$885.0^{+63.0}_{-50.0}$	$1.885^{+0.094}_{-0.094}$	$9.0^{+11.0}_{-2.0}$	1.05 ± 0.13	31 ± 6	$0.409^{+0.043}_{-0.043}$	$12.3^{+9.7}_{-9.7}$	6.4 ± 2.6	78^{+70}_{-70}
LkHa324	B9	$605.0^{+16.0}_{-14.0}$	$2.82^{+0.61}_{-0.2}$	$2.12^{+0.44}_{-0.92}$	0.47 ± 0.13	17 ± 4	$0.42^{+0.1}_{-0.06}$	$12.2^{+6.7}_{-5.7}$	40.0 ± 34.0	480^{+490}_{-470}
MWC137	B0	$2910.0^{+600.0}_{-400.0}$	$23.1^{+10.6}_{-6.5}$	$0.018^{+0.019}_{-0.008}$	0.75 ± 0.36	12 ± 4	$1.05^{+0.52}_{-0.36}$	$39.0^{+36.0}_{-25.0}$	7.0 ± 10.0	280^{+470}_{-440}
V361Cep	B2	$893.0^{+35.0}_{-31.0}$	$5.31^{+0.69}_{-0.48}$	$0.41^{+0.15}_{-0.13}$	0.288 ± 0.041	18 ± 4	$0.403^{+0.082}_{-0.074}$	$10.9^{+4.4}_{-4.0}$	180.0 ± 88.0	2000^{+1200}_{-1200}
V373Cep	B5	$922.0^{+33.0}_{-29.0}$	$3.18^{+0.51}_{-0.39}$	$1.63^{+0.75}_{-0.6}$	0.291 ± 0.073	5 ± 2	$0.57^{+0.11}_{-0.09}$	$18.7^{+10.1}_{-9.9}$	44.0 ± 39.0	820^{+850}_{-850}

Columns (3), (4) and (5) taking from Vioque *et al.* (2018)

Columns (7): Number of stars found using column (6) (§5.2.1) and Gaia DR2 completeness (§5.2.3).

Columns (8): Low mass limit detected for each cluster (§5.2.2).

Columns (9): Initial mass function ratio of each cluster (§5.2.5).

Columns (10): Stellar density for each cluster (§5.2.4).

Columns (11): Stellar density for each cluster corrected by their IMF (§5.2.6).

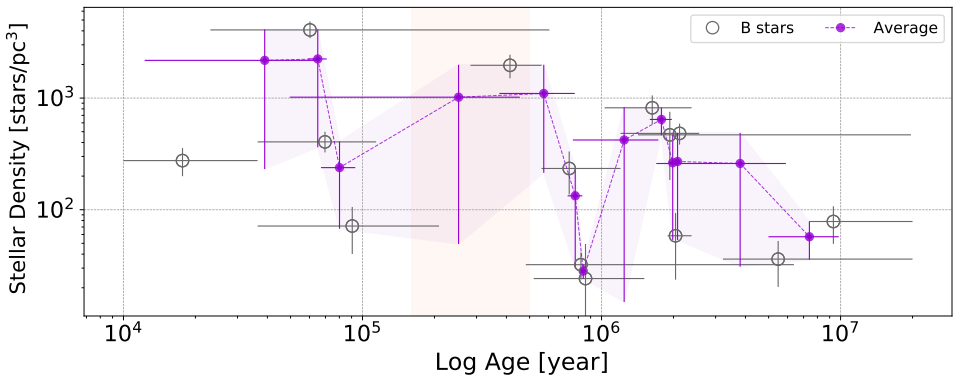


Figure H.2: Real Stellar densities vs Age. The open grey circles represent the B stars found to be in a cluster by CEREAL and the purple filled circles represent the median rolling average (for every 2 points) of those stars. The purple shadows represent the median error of the rolling average. The light orange area represents the time scale depicted in figure 5.9.

References

Alpaydin, E. (2014). *Introduction to Machine Learning*. The MIT Press. 58, 191

Andrae, R., Fouesneau, M., Creevey, O., Ordenovic, C., Mary, N., Burlacu, A., Chaoul, L., Jean-Antoine-Piccolo, A., Kordopatis, G., Korn, A., Lebreton, Y., Panem, C., Pichon, B., Thévenin, F., Walmsley, G. & Bailer-Jones, C.A.L. (2018). Gaia Data Release 2. First stellar parameters from Apsis. *A&A*, **616**, A8. 199

Andre, P., Ward-Thompson, D. & Barsony, M. (1993). Submillimeter Continuum Observations of rho Ophiuchi A: The Candidate Protostar VLA 1623 and Prestellar Clumps. *ApJ*, **406**, 122. 7

André, P., Men'shchikov, A., Bontemps, S., Könyves, V., Motte, F., Schneider, N., Didelon, P., Minier, V., Saraceno, P., Ward-Thompson, D., di Francesco, J., White, G., Molinari, S., Testi, L., Abergel, A., Griffin, M., Henning, T., Royer, P., Merín, B., Vavrek, R., Attard, M., Arzoumanian, D., Wilson, C.D., Ade, P., Aussel, H., Baluteau, J.P., Benedettini, M., Bernard, J.P., Blommaert, J.A.D.L., Cambrésy, L., Cox, P., di Giorgio, A., Hargrave, P., Hennemann, M., Huang, M., Kirk, J., Krause, O., Launhardt, R., Leeks, S., Le Pennec, J., Li, J.Z., Martin, P.G., Maury, A., Olofsson, G., Omont, A., Peretto, N., Pezzuto, S., Prusti, T., Roussel, H., Russeil, D., Sauvage, M., Sibthorpe, B., Sicilia-Aguilar, A., Spinoglio, L., Waelkens, C., Woodcraft, A. & Zavagno, A. (2010). From filamentary clouds to prestellar cores to the stellar IMF: Initial highlights from the Herschel Gould Belt Survey. *A&A*, **518**, L102. 19

Ankerst, M., Breunig, M.M., peter Kriegel, H. & Sander, J. (1999). Optics: Ordering points to identify the clustering structure. 49–60, ACM Press. 58, 66, 127, 192, 193, 194

Arenou, F., Luri, X., Babusiaux, C., Fabricius, C., Helmi, A., Muraveva, T., Robin, A.C., Spoto, F., Vallenari, A., Antoja, T., Cantat-Gaudin, T., Jordi, C., Leclerc, N., Reylé, C., Romero-Gómez, M., Shih, I.C., Soria, S., Barache, C., Bossini, D., Bragaglia, A., Breddels, M.A., Fabrizio, M., Lambert, S., Marrese, P.M., Massari, D., Moitinho, A., Robichon, N., Ruiz-Dern, L., Sordo, R., Veljanoski, J., Eyer, L., Jasniewicz, G., Pancino, E., Soubiran, C., Spagna, A., Tanga, P., Turon, C. & Zurbach, C. (2018). Gaia Data Release 2. Catalogue validation. *A&A*, **616**, A17. 59, 106, 107, 108, 199

Arnold, B., Wright, N.J. & Parker, R.J. (2022). Quantifying kinematic substructure in star-forming regions with statistical tests of spatial autocorrelation. *MNRAS*, **515**, 2266–2279. 124, 136

Ascenso, J. (2018). *Embedded Clusters*, vol. 424, 1. 98

Bachiller, R., Perez Gutierrez, M. & Garcia-Lario, P. (1998). IRAS06562-0337, the Iron-clad Nebula: a young star embedded in a molecular cloud. *A&A*, **331**, L45–L48. 185, 186, 187, 188, 189, 190

Bailer-Jones, C.A.L. (2015). Estimating Distances from Parallaxes. *PASP*, **127**, 994. 153

Banerjee, S. & Kroupa, P. (2017a). How can young massive clusters reach their present-day sizes? *A&A*, **597**, A28. 17

Banerjee, S. & Kroupa, P. (2017b). How can young massive clusters reach their present-day sizes? *A&A*, **597**, A28. 103

Barrado, D., de Gregorio Monsalvo, I., Huélamo, N., Morales-Calderón, M., Bayo, A., Palau, A., Ruiz, M.T., Rivière-Marichalar, P., Bouy, H., Morata, Ó., Stauffer, J.R., Eiroa, C. & Noriega-Crespo, A. (2018). Early phases in the stellar and substellar formation and evolution. Infrared and submillimeter data in the Barnard 30 dark cloud. *A&A*, **612**, A79. 185, 186, 187, 188, 189, 190

Bastian, N. (2016). Young Massive Clusters: Their Population Properties, Formation and Evolution, and Their Relation to the Ancient Globular Clusters. In *EAS Publications Series*, vol. 80-81 of *EAS Publications Series*, 5–37. 15

Belloche, A., Maury, A.J., Maret, S., Anderl, S., Bacmann, A., André, P., Bontemps, S., Cabrit, S., Codella, C., Gaudel, M., Gueth, F., Lefèvre, C., Lefloch, B., Podio, L. & Testi, L. (2020). Questioning the spatial origin of complex organic molecules in young protostars with the CALYPSO survey. *A&A*, **635**, A198. 56, 57, 58

Beltrán, M.T. (2018). Perspectives on star formation: the formation of high-mass stars. In A. Tarchi, M.J. Reid & P. Castangia, eds., *Astrophysical Masers: Unlocking the Mysteries of the Universe*, vol. 336, 193–200. 102, 103, 116

Bennett, J., Donahue, M., Schneider, N. & Voit, M. (2011). *The Cosmo Perspective*. PEARSON. 2

Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M. & Bartlett, J.G. (2000). The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources. *A&AS*, **143**, 33–40. 63

Bonnell, I.A. & Bate, M.R. (2006). Star formation through gravitational collapse and competitive accretion. *MNRAS*, **370**, 488–494. 20, 21

Bonnell, I.A., Bate, M.R. & Zinnecker, H. (1998). On the formation of massive stars. *MNRAS*, **298**, 93–102. 20

Bonnell, I.A., Bate, M.R. & Vine, S.G. (2003). The hierarchical formation of a stellar cluster. *MNRAS*, **343**, 413–418. 17, 18, 19, 101, 102, 112, 115, 116, 117, 118, 120, 121, 122, 123, 132, 199, 200

Boubert, D. & Everall, A. (2020). Completeness of the Gaia verse II: what are the odds that a star is missing from Gaia DR2? *MNRAS*, **497**, 4246–4261. 107, 108

Bressan, A., Marigo, P., Girardi, L., Salasnich, B., Dal Cero, C., Rubele, S. & Nanni, A. (2012). PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code. *MNRAS*, **427**, 127–145. 135, 138, 154

Briceño, C., Calvet, N., Hernández, J., Vivas, A.K., Mateu, C., Downes, J.J., Loerincs, J., Pérez-Blanco, A., Berlind, P., Espaillat, C., Allen, L., Hartmann, L., Mateo, M. & Bailey, I., John I. (2019). The CIDA Variability Survey of Orion OB1. II. Demographics of the Young, Low-mass Stellar Populations. *AJ*, **157**, 85. 56, 57, 58, 63, 185, 186, 187, 188, 189, 190

Buckner, A.S.M., Khorrami, Z., González, M., Lumsden, S.L., Moraux, E., Oudmaijer, R.D., Clark, P., Joncour, I., Blanco, J.M., de la Calle, I., Hacar, Á., Herrera-Fernandez, J.M., Motte, F., Salgado, J. & Valero-Martín, L. (2020). The spatial evolution of young massive clusters. II. Looking for imprints of star formation in NGC 2264 with Gaia DR2. *A&A*, **636**, A80. 119

Calvet, N., Muzerolle, J., Briceño, C., Hernández, J., Hartmann, L., Saucedo, J.L. & Gordon, K.D. (2004). The Mass Accretion Rates of Intermediate-Mass T Tauri Stars. *AJ*, **128**, 1294–1318. 49

Campello, R.J.G.B., Moulavi, D., Zimek, A. & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Trans. Knowl. Discov. Data*, **10**, 5:1–5:51. 58, 66, 127, 193, 194

Cánovas, H., Cantero, C., Cieza, L., Bombrun, A., Lammers, U., Merín, B., Mora, A., Ribas, Á. & Ruíz-Rodríguez, D. (2019). Census of ρ Ophiuchi candidate members from Gaia Data Release 2. *A&A*, **626**, A80. 59, 65, 66, 69, 71, 127, 133

Cantat-Gaudin, T., Jordi, C., Vallenari, A., Bragaglia, A., Balaguer-Núñez, L., Soubiran, C., Bossini, D., Moitinho, A., Castro-Ginard, A., Krone-Martins, A., Casamiquela, L., Sordo, R. & Carrera, R. (2018a). A Gaia DR2 view of the open cluster population in the Milky Way. *A&A*, **618**, A93. 56, 57, 58

Cantat-Gaudin, T., Jordi, C., Vallenari, A., Bragaglia, A., Balaguer-Núñez, L., Soubiran, C., Bossini, D., Moitinho, A., Castro-Ginard, A., Krone-Martins, A., Casamiquela, L., Sordo, R. & Carrera, R. (2018b). A Gaia DR2 view of the open cluster population in the Milky Way. *A&A*, **618**, A93. 56, 57, 58

Carroll, B.W. & Ostlie, D.A. (2007). *An Introduction to Modern Astrophysics*. PEARSON. 5

Carruthers, G.R. (1970). Atomic and Molecular Hydrogen in Interstellar Space. *Space Sci. Rev.*, **10**, 459–482. 5

Chen, P.S., Shan, H.G. & Zhang, P. (2016). A new photometric study of Herbig Ae/Be stars in the infrared. *New∼Astron.*, **44**, 1–11. 48, 49, 125, 127, 153

de Grijs, R. (2010). Young massive star clusters: achievements and challenges. In R. de Grijs & J.R.D. Lépine, eds., *Star Clusters: Basic Galactic Building Blocks Throughout Time and Space*, vol. 266, 49–57. 16

de Wit, W.J., Oudmaijer, R.D., van den Ancker, M.E. & Calvet, N. (2014). Report on the Workshop Herbig Ae/Be Stars: The Missing Link in Star Formation. *The Messenger*, **157**, 50–53. 11

Dotter, A. (2016). MESA Isochrones and Stellar Tracks (MIST) 0: Methods for the Construction of Stellar Isochrones. *ApJS*, **222**, 8. 135, 138

ESA, ed. (1997). *The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission*, vol. 1200 of *ESA Special Publication*. 22, 25, 26

Ester, M., Kriegel, H.P., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. 226–231, AAAI Press. 58, 66, 127, 192, 193, 194

Evans, M.G. (2018). Gravitational instabilities in a protosolar-like disc. 8

Fairlamb, J.R., Oudmaijer, R.D., Mendigutía, I., Ilee, J.D. & van den Ancker, M.E. (2015). A spectroscopic survey of Herbig Ae/Be stars with X-shooter - I. Stellar parameters and accretion rates. *MNRAS*, **453**, 976–1001. 49, 153

Field, A. (2016). *An adventure in statistics. The Reality enigma*. SAGE. 78

Frost, A.J. (2019). The circumstellar environment of massive young stellar objects - a multi-scale analysis. 10

Frost, A.J., Oudmaijer, R.D., de Wit, W.J. & Lumsden, S.L. (2021). Unveiling the traits of massive young stellar objects through a multi-scale survey. *arXiv e-prints*, arXiv:2102.05087. 4, 7

Gaia Collaboration, Brown, A.G.A., Vallenari, A., Prusti, T., de Bruijne, J.H.J., Mignard, F., Drimmel, R., Babusiaux, C., Bailer-Jones, C.A.L., Bastian, U. & et al. (2016a). Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties. *A&A*, **595**, A2. 22, 25, 26, 30

Gaia Collaboration, Prusti, T., de Bruijne, J.H.J., Brown, A.G.A., Vallenari, A., Babusiaux, C., Bailer-Jones, C.A.L., Bastian, U., Biermann, M., Evans, D.W. & et al. (2016b). The Gaia mission. *A&A*, **595**, A1. 22, 125

Gaia Collaboration, van Leeuwen, F., Vallenari, A., Jordi, C., Lindegren, L., Bastian, U., Prusti, T., de Bruijne, J.H.J., Brown, A.G.A., Babusiaux, C. & et al. (2017). Gaia Data Release 1. Open cluster astrometry: performance, limitations, and future prospects. *A&A*, **601**, A19. 14, 30, 31, 32, 33, 34, 35, 45, 126, 133

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., Barstow, M.A., Jordi, C., Vallenari, A., Bossini, D., Bressan, A., Cantat-Gaudin, T., van Leeuwen & et al. (2018a). Gaia Data Release 2. Observational Hertzsprung-Russell diagrams. *A&A*, **616**, A10. 36, 197

Gaia Collaboration, Brown, A.G.A. & *et al.* (2018b). Gaia Data Release 2. Summary of the contents and survey properties. *A&A*, **616**, A1. 22, 23, 26, 30, 106, 107, 125

Garcia, P.J.V. (2011). *Physical processes in circumstellar disks around young stars*. The University of Chicago Press. 3

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Lear, Karas TensorFlow. Concepts,Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media. 58, 191

Getman, K.V., Feigelson, E.D., Kuhn, M.A., Bate, M.R., Broos, P.S. & Garmire, G.P. (2018). Intracluster age gradients in numerous young stellar clusters. *MNRAS*, **476**, 1213–1223. 56, 57, 58

Gomez, M., Hartmann, L., Kenyon, S.J. & Hewett, R. (1993). On the spatial distribution of pre-main-sequence stars in Taurus. *AJ*, **105**, 1927–1937. 4, 125, 151

Grosssschedl, J.E., Alves, J., Teixeira, P.S., Bouy, H., Forbrich, J., Lada, C.J., Meingast, S., Hacar, Á., Ascenso, J., Ackerl, C., Hasenberger, B., Köhler, R., Kubiak, K., Larreina, I., Linhardt, L., Lombardi, M. & Möller, T. (2019). VISION - Vienna survey in Orion. III. Young stellar objects in Orion A. *A&A*, **622**, A149. 63

Hara, A., Tachihara, K., Mizuno, A., Onishi, T., Kawamura, A., Obayashi, A. & Fukui, Y. (1999). A Study of Dense Cloud Cores and Star Formation in Lupus: $C^{18}O$ J = 1-0 Observations with NANTEN. *PASJ*, **51**, 895–910. 185, 186, 187, 188, 189, 190

Hartmann, J. (1904). Investigations on the spectrum and orbit of delta Orionis. *ApJ*, **19**. 1

Hartmann, L. (2009). *Accretion Processes in Star Formation*. Cambridge Astrophysics Series. 2, 5, 11

Harwit, M. (2006). *Astrophysical concepts*. Springer. 5

Herbig, G.H. (1960). The Spectra of Be- and Ae-Type Stars Associated with Nebulosity. *ApJS*, **4**, 337. 11

Herbig, G.H. (1962). The Properties and Problems of T Tauri Stars and Related Objects. *Advances in Astronomy and Astrophysics*, **1**, 47–103. 14

Hernández, J., Calvet, N., Briceño, C., Hartmann, L. & Berlind, P. (2004). Spectral Analysis and Classification of Herbig Ae/Be Stars. *AJ*, **127**, 1682–1701. 11

Hernández, J., Calvet, N., Hartmann, L., Briceño, C., Sicilia-Aguilar, A. & Berlind, P. (2005). Herbig Ae/Be Stars in nearby OB Associations. *AJ*, **129**, 856–871. 11

Hillenbrand, L.A., Meyer, M.R., Strom, S.E. & Skrutskie, M.F. (1995). Isolated star-forming regions containing Herbig Ae/Be stars. 1: The young stellar aggregate associated with BD +40deg 4124. *AJ*, **109**, 280–297. 12, 104, 123, 127, 132, 150, 151, 152

Hodges, J.L. (1958). The significance probability of the smirnov two-sample test. *Ark. Mat.*, **3**, 469–486. 78

Høg, E., Fabricius, C., Makarov, V.V., Urban, S., Corbin, T., Wycoff, G., Bastian, U., Schwekendiek, P. & Wicenec, A. (2000). The Tycho-2 catalogue of the 2.5 million brightest stars. *A&A*, **355**, L27–L30. 22, 25, 26, 48

Houk, N. & Swift, C. (1999). Michigan catalogue of two-dimensional spectral types for the HD Stars, Vol. 5. *Michigan Spectral Survey*, **5**, 0. 63

Janes, K. (2001). Star clusters: Encyclopedia of astronomy and astrophysics. http://www.astro.caltech.edu/~george/ay20/eaa-starclus.pdf. 14, 27

Jeans, J.H. (1902). The Stability of a Spherical Nebula. *Philosophical Transactions of the Royal Society of London Series A*, **199**, 1–53. 5

Jonkheid, B.J. (2006). *Chemistry in evolving protoplanetary disks*. Ph.D. thesis, Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands. 8

Juárez, C., Liu, H.B., Girart, J.M., Palau, A., Busquet, G., Galván-Madrid, R., Hirano, N. & Lin, Y. (2019). Extreme fragmentation and complex kinematics at the center of the L1287 cloud. *A&A*, **621**, A140. 71

Kahn, F.D. (1974). Cocoons around early-type stars. *A&A*, **37**, 149–162. 18

Kauffmann, J., Bertoldi, F., Bourke, T.L., Evans, I., N. J. & Lee, C.W. (2008). MAMBO mapping of Spitzer c2d small clouds and cores. *A&A*, **487**, 993–1017. 185, 186, 187, 188, 189, 190

Kos, J., Bland-Hawthorn, J., Asplund, M., Buder, S., Lewis, G.F., Lin, J., Martell, S.L., Ness, M.K., Sharma, S., De Silva, G.M., Simpson, J.D., Zucker, D.B., Zwitter, T., Čotar, K. & Spina, L. (2019). Discovery of a 21 Myr old stellar population in the Orion complex⋆. *A&A*, **631**, A166. 56, 57, 58

Krause, M.G.H., Offner, S.S.R., Charbonnel, C., Gieles, M., Klessen, R.S., Vázquez-Semadeni, E., Ballesteros-Paredes, J., Girichidis, P., Kruijssen, J.M.D., Ward, J.L. & Zinnecker, H. (2020). The Physics of Star Cluster Formation and Evolution. *Space Sci. Rev.*, **216**, 64. 16

Krumholz, M.R. & Bonnell, I.A. (2007). Models for the Formation of Massive Stars. *arXiv e-prints*, arXiv:0712.0828. 17, 18, 102

Krumholz, M.R., McKee, C.F. & Klein, R.I. (2005). The formation of stars by gravitational collapse rather than competitive accretion. *Nature*, **438**, 332–334. 21

Krumholz, M.R., McKee, C.F. & Bland-Hawthorn, J. (2019). Star Clusters Across Cosmic Time. *ARA&A*, **57**, 227–303. 13, 14, 15, 16, 22

Kutner, M.L. (2003). *Astronomy: A Physical Perspective*. Cambridge University Press. 3

Lada, C.J. & Lada, E.A. (2003). Embedded Clusters in Molecular Clouds. *ARA&A*, **41**, 57–115. 2, 3

Lada, E.A., Strom, K.M. & Myers, P.C. (1993). Environments of star formation - Relationship between molecular clouds, dense cores and young stars. In E.H. Levy & J.I. Lunine, eds., *Protostars and Planets III*, 245–277. 4, 7, 13, 14, 15, 101

Lee, H.T. & Lim, J. (2008). On the Formation of Perseus OB1 at High Galactic Latitudes. *ApJ*, **679**, 1352–1363. 185, 186, 187, 188, 189, 190

Leitherer, C. (1994). Massive Stars in Starburst Galaxies and the Origin of Galactic Superwinds. *Reviews in Modern Astronomy*, **7**, 73–102. 7

Li, J., Myers, P.C., Kirk, H., Gutermuth, R.A., Dunham, M.M. & Pokhrel, R. (2019). Catalog of High Protostellar Surface Density Regions in Nearby Embedded Clusters. *ApJ*, **871**, 163. 185, 186, 187, 188, 189, 190

Lindegren, L., Hernández, J., Bombrun, A., Klioner & *et al.* (2018a). Gaia Data Release 2. The astrometric solution. *A&A*, **616**, A2. 27

Lindegren, L., Hernández, J., Bombrun, A., Klioner, S., Bastian, U. & *et al.* (2018b). Gaia Data Release 2. The astrometric solution. *A&A*, **616**, A2. 26

Lindegren, L., Bastian, U., Biermann, M., Bombrun, A., de Torres, A., Gerlach, E., Geyer, R., Hernández, J., Hilger, T., Hobbs, D., Klioner, S.A., Lammers, U., McMillan, P.J., Ramos-Lerate, M., Steidelmüller, H., Stephenson, C.A. & van Leeuwen, F. (2020). Gaia Early Data Release 3: Parallax bias versus magnitude, colour, and position. *arXiv e-prints*, arXiv:2012.01742. 22, 26

Lindegren, L., Klioner, S.A., Hernández, J., Bombrun, A., Ramos-Lerate, M., Steidelmüller, H., Bastian, U., Biermann, M. & et al. (2021). Gaia Early Data Release 3. The astrometric solution. *A&A*, **649**, A2. 22, 26

Louvet, F. (2018). Review on high-mass star formation. In P. Di Matteo, F. Billebaud, F. Herpin, N. Lagarde, J.B. Marquette, A. Robin & O. Venot, eds., *SF2A-2018: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, Di. 4, 8, 102, 103

Luhman, K.L. & Esplin, T.L. (2020). Refining the Census of the Upper Scorpius Association with Gaia. *AJ*, **160**, 44. 185, 186, 187, 188, 189, 190

Luri, X., Brown, A.G.A., Sarro, L.M., Arenou, F., Bailer-Jones, C.A.L., Castro-Ginard, A., de Bruijne, J., Prusti, T., Babusiaux, C. & Delgado, H.E. (2018). Gaia Data Release 2. Using Gaia parallaxes. *A&A*, **616**, A9. 27, 153

Magakian, T.Y. (2003a). Merged catalogue of reflection nebulae. *A&A*, **399**, 141–145. 56, 57, 58

Magakian, T.Y. (2003b). Merged catalogue of reflection nebulae. *A&A*, **399**, 141–145. 185, 186, 187, 188, 189, 190

Marigo, P., Girardi, L., Bressan, A., Rosenfield, P., Aringer, B., Chen, Y., Dussin, M., Nanni, A., Pastorelli, G., Rodrigues, T.S., Trabucchi, M., Bladh, S., Dalcanton, J., Groenewegen, M.A.T., Montalbán, J. & Wood, P.R. (2017). A New Generation of PARSEC-COLIBRI Stellar Isochrones Including the TP-AGB Phase. *ApJ*, **835**, 77. 135, 138, 154

McInnes, L., Healy, J. & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, **2**, 9–22. 58, 66, 127, 193, 194

McKee, C.F. & Ostriker, J.P. (1977). A theory of the interstellar medium - Three components regulated by supernova explosions in an inhomogeneous substrate. *ApJ*, **218**, 148–169. 1

McKee, C.F. & Tan, J.C. (2003). The Formation of Massive Stars from Turbulent Cores. *ApJ*, **585**, 850–871. 19

Mehner, A., de Wit, W.J., Groh, J.H., Oudmaijer, R.D., Baade, D., Rivinius, T., Selman, F., Boffin, H.M.J. & Martayan, C. (2016). VLT/MUSE discovers a jet from the evolved B[e] star MWC 137. *A&A*, **585**, A81. 49, 51, 56, 57, 58, 74

Mendigutía, I., Mora, A., Montesinos, B., Eiroa, C., Meeus, G., Merín, B. & Oudmaijer, R.D. (2012). Accretion-related properties of Herbig Ae/Be stars. Comparison with T Tauris. *A&A*, **543**, A59. 4, 125

Michalik, D., Lindegren, L., Hobbs, D. & Butkevich, A.G. (2015). Gaia astrometry for stars with too few observations. A Bayesian approach. *A&A*, **583**, A68. 22

Motte, F., Bontemps, S. & Louvet, F. (2018). High-Mass Star and Massive Cluster Formation in the Milky Way. *ARA&A*, **56**, 41–82. 11, 17

Oudmaijer, R.D. & Parr, A.M. (2010). The binary fraction and mass ratio of Be and B stars: a comparative Very Large Telescope/NACO study. *MNRAS*, **405**, 2439–2446. 104

Palla, F. & Stahler, S.W. (1993). The Pre-Main-Sequence Evolution of Intermediate-Mass Stars. *ApJ*, **418**, 414. 71

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830. 58, 191, 193

Peretto, N., André, P. & Belloche, A. (2006). Probing the formation of intermediate- to high-mass stars in protoclusters. A detailed millimeter study of the NGC 2264 clumps. *A&A*, **445**, 979–998. 119, 124, 136

Pérez-Blanco, A., Maucó, K., Hernández, J., Calvet, N., Espaillat, C., McClure, M., Briceño, C., Robinson, C., Feldman, D., Villarreal, L. & D'Alessio, P. (2018). A Transitional Disk around an Intermediate-mass Star in the Sparse Population of the Orion OB1 Association. *ApJ*, **867**, 116. 49

Perryman, M.A.C., Lindegren, L., Kovalevsky, J., Hog, E., Bastian, U., Bernacca, P.L., Creze, M., Donati, F., Grenon, M., Grewing, M., van Leeuwen, F., van der Marel, H., Mignard, F., Murray, C.A., Le Poole, R.S., Schrijver, H., Turon, C., Arenou, F., Froeschle, M. & Petersen, C.S. (1997). The Hipparcos Catalogue. *A&A*, **500**, 501–504. 22, 26

Portegies Zwart, S.F., McMillan, S.L.W. & Gieles, M. (2010). Young Massive Star Clusters. *ARA&A*, **48**, 431–493. 13

Prialnik, D. (2000). *The Theory of Stellar Structure and Evolution*. Cambridge University Press. 2, 5, 111

Rivilla, V.M., Martín-Pintado, J., Jiménez-Serra, I. & Rodríguez-Franco, A. (2013). The role of low-mass star clusters in massive star formation. The Orion case. *A&A*, **554**, A48. 18, 20, 102, 103, 118, 124, 133, 135, 136

Roccatagliata, V., Franciosini, E., Sacco, G.G., Rand ich, S. & Sicilia-Aguilar, A. (2020). A 3D view of the Taurus star-forming region by Gaia and Herschel. Multiple populations related to the filamentary molecular cloud. *A&A*, **638**, A85. 185, 186, 187, 188, 189, 190

Salaris, M. & Cassisi, S. (2005). *Evolution of Stars and Stellar Populations*. John Wiley Sons. 112

Salpeter, E.E. (1955). The Luminosity Function and Stellar Evolution. *ApJ*, **121**, 161. 111, 112, 115

Samus', N.N., Kazarovets, E.V., Durlevich, O.V., Kireeva, N.N. & Pastukhova, E.N. (2017). General catalogue of variable stars: Version GCVS 5.1. *Astronomy Reports*, **61**, 80–88. 71

Saul, M. (2015). Millimeter Source 13 S in R CrA:Observations of a Proto-Herbig Ae System Candidate. *ApJ*, **809**, 86. 185, 186, 187, 188, 189, 190

Schulz, N.S. (2005). *From Dust to Stars*. Springer-Praxis. 1, 2

Segovia, J.C. (2016). Gaia tap+ (astroquery.gaia). astroquery gaia package. https://astroquery.readthedocs.io/en/latest/gaia/gaia.html. 27

Sen, A.K., Polcaro, V.F., Dey, I. & Gupta, R. (2010). Photopolarimetric study of the star-forming clouds CB3, CB25, and CB39. *A&A*, **522**, A45. 185, 186, 187, 188, 189, 190

Shimoikura, T. & Dobashi, K. (2011). Star Formation and Distributions of Gas and Dust in the Circinus Cloud. *ApJ*, **731**, 23. 185, 186, 187, 188, 189, 190

Shu, F.H., Adams, F.C. & Lizano, S. (1987). Star formation in molecular clouds: observation and theory. *ARA&A*, **25**, 23–81. 8

Smilgys, R. & Bonnell, I.A. (2017). Formation of stellar clusters. *MNRAS*, **472**, 4982–4991. 17, 124, 133

Smith, N. & Brooks, K.J. (2008). *The Carina Nebula: A Laboratory for Feedback and Triggered Star Formation*, vol. 5, 138. 5

Solomon, P.M., Edmunds, M.G. & Turner, B.E. (1980). Book-Review - Giant Molecular Clouds in the Galaxy. *Nature*, **287**, 176. 5

Soubiran, C., Cantat-Gaudin, T., Romero-Gómez, M., Casamiquela, L., Jordi, C., Vallenari, A., Antoja, T., Balaguer-Núñez, L., Bossini, D., Bragaglia, A., Carrera, R., Castro-Ginard, A., Figueras, F., Heiter, U., Katz, D., Krone-Martins, A., Le Campion, J.F., Moitinho, A. & Sordo, R. (2018). Open cluster kinematics with Gaia DR2. *A&A*, **619**, A155. 185, 186, 187, 188, 189, 190

Tan, J.C., Beltrán, M.T., Caselli, P., Fontani, F., Fuente, A., Krumholz, M.R., McKee, C.F. & Stolte, A. (2014). Massive Star Formation. In H. Beuther, R.S. Klessen, C.P. Dullemond & T. Henning, eds., *Protostars and Planets VI*, 149. 4

Testi, L., Palla, F., Prusti, T., Natta, A. & Maltagliati, S. (1997). A search for clustering around Herbig Ae/Be stars. *A&A*, **320**, 159–166. 4, 12, 13, 23, 28, 29, 47, 54, 71, 73, 74, 89, 90, 94, 95, 96, 97, 99, 103, 104, 123, 125, 127, 131, 136, 137, 150, 151, 152

Testi, L., Palla, F. & Natta, A. (1998). A search for clustering around Herbig Ae/Be stars. II. Atlas of the observed sources. *A&AS*, **133**, 81–121. 12, 23, 47, 71, 89, 103, 104, 123, 127, 136, 137, 150, 151

Testi, L., Palla, F. & Natta, A. (1999). The onset of cluster formation around Herbig Ae/Be stars. *A&A*, **342**, 515–523. 11, 14, 18, 12, 23, 47, 71, 72, 73, 74, 75, 89, 97, 103, 104, 123, 127, 128, 129, 136, 137, 150, 151

Testi, L., Sargent, A.I., Olmi, L. & Onello, J.S. (2000). Star Formation in Clusters: Early Subclustering in the Serpens Core. *ApJ*, **540**, L53–L56. 12, 103, 119

The, P.S., de Winter, D. & Perez, M.R. (1994). A new catalogue of members and candidate members of the Herbig Ae/Be (HAEBE) stellar group. *A&AS*, **104**, 315–339. 11, 48, 153, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165

Trumpler, R.J. (1930). Preliminary results on the distances, dimensions and space distribution of open star clusters. *Lick Observatory Bulletin*, **14**, 154–188. 1, 14

van Leeuwen, F. (2009). Parallaxes and proper motions for 20 open clusters as based on the new Hipparcos catalogue. *A&A*, **497**, 209–242. 14, 31, 33, 34, 35, 45, 126, 133

Vázquez-Semadeni, E., Palau, A., Ballesteros-Paredes, J., Gómez, G.C. & Zamora-Avilés, M. (2019). Global hierarchical collapse in molecular clouds. Towards a comprehensive scenario. *MNRAS*, **490**, 3061–3097. 102

Vieira, S.L.A., Corradi, W.J.B., Alencar, S.H.P., Mendes, L.T.S., Torres, C.A.O., Quast, G.R., Guimarães, M.M. & da Silva, L. (2003). Investigation of 131 Herbig Ae/Be Candidate Stars. *AJ*, **126**, 2971–2987. 11

Vioque, M., Oudmaijer, R.D., Baines, D., Mendigutía, I. & Pérez-Martínez, R. (2018). Gaia DR2 study of Herbig Ae/Be stars. *ArXiv e-prints*. 12, 22, 27, 48, 49, 103, 104, 109, 112, 115, 123, 125, 127, 153, 154, 197, 200

Vioque, M., Oudmaijer, R.D., Schreiner, M., Mendigutía, I., Baines, D., Mowlavi, N. & Pérez-Martínez, R. (2020). Catalogue of new Herbig Ae/Be and classical Be stars. A machine learning approach to Gaia DR2. *A&A*, **638**, A21. 12, 135, 199

Wall, J.V. & Jenkins, C.R. (2003). *Practical Statistics for Astronomers*. Cambridge University Press. 29, 30, 78

Wang, S. & Chen, X. (2019). The Optical to Mid-infrared Extinction Law Based on the APOGEE, Gaia DR2, Pan-STARRS1, SDSS, APASS, 2MASS, and WISE Surveys. *ApJ*, **877**, 116. 104, 197

Wang, S. & Looney, L.W. (2007). Young Stellar Groups around Herbig Ae/Be Stars: A Low-Mass YSO Census. *ApJ*, **659**, 1360–1372. 12

Ward-Thompson, D. & Whitworth, A. (2011). *An Introduction to Star Formation*. Cambridge University Press. 2, 112

Waters, L.B.F.M. & Waelkens, C. (1998). Herbig Ae/Be Stars. *ARA&A*, **36**, 233–266. 3, 11, 12

Wenger, M., Ochsenbein, F., Egret, D., Dubois, P., Bonnarel, F., Borde, S., Genova, F., Jasniewicz, G., Laloë, S., Lesteven, S. & Monier, R. (2000). The SIMBAD astronomical database. The CDS reference database for astronomical objects. *A&AS*, **143**, 9–22. 21, 48, 49, 56, 57, 58, 185, 186, 187, 188, 189, 190

Wichittanakom, C., Oudmaijer, R.D., Fairlamb, J.R., Mendigutía, I., Vioque, M. & Ababakr, K.M. (2020). The accretion rates and mechanisms of Herbig Ae/Be stars. *MNRAS*, **493**, 234–249. 12, 49, 153

Williams, J.P., Blitz, L. & McKee, C.F. (2000). The Structure and Evolution of Molecular Clouds: from Clumps to Cores to the IMF. In V. Mannings, A.P. Boss & S.S. Russell, eds., *Protostars and Planets IV*, 97. 6

Wright, N.J. (2015). The Kinematics of Star Formation: Theory and Observation in the Gaia Era. *arXiv e-prints*, arXiv:1512.06854. 101, 103

Wright, N.J. (2020). OB Associations and their origins. *New∼Astron.∼Rev.*, **90**, 101549. 14, 22

Wright, N.J. & Parker, R.J. (2019). Kinematic signatures of cluster formation from cool collapse in the Lagoon Nebula cluster NGC 6530. *MNRAS*, **489**, 2694–2701. 119

Zacharias, N., Urban, S.E., Zacharias, M.I., Wycoff, G.L., Hall, D.M., Germain, M.E., Holdenried, E.R. & Winter, L. (2003). UCAC2 Catalogue (Zacharias+ 2004). *VizieR Online Data Catalog*, I/289. 48

Zacharias, N., Finch, C.T., Girard, T.M., Henden, A., Bartlett, J.L., Monet, D.G. & Zacharias, M.I. (2012). VizieR Online Data Catalog: UCAC4 Catalogue (Zacharias+, 2012). *VizieR Online Data Catalog*, I/322A. 48

Zinnecker, H. & Yorke, H.W. (2007). Toward Understanding Massive Star Formation. *ARA&A*, **45**, 481–563. 9, 10

Zinnecker, H., McCaughrean, M.J. & Wilking, B.A. (1993). The initial stellar population. In E.H. Levy & J.I. Lunine, eds., *Protostars and Planets III*, 429–495. 4

Zucker, C., Speagle, J.S., Schlafly, E.F., Green, G.M., Finkbeiner, D.P., Goodman, A. & Alves, J. (2020). A compendium of distances to molecular clouds in the Star Formation Handbook. *A&A*, **633**, A51. 56, 57, 58, 116, 185, 186, 187, 188, 189, 190