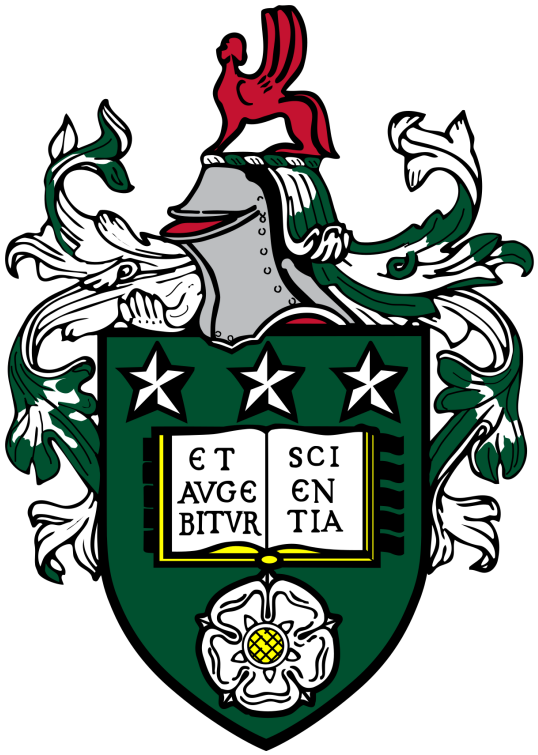
A census of Herbig Ae/Be stars

New candidates and analysis from a Gaia perspective



Miguel Vioque de Lózar

School of Physics and Astronomy

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

August 2020

Under English trees I contemplated that lost labyrinth, imagining it pristine and inviolate in a mountain fastness. I imagined it obliterated by paddies or under water; I pictured it endless, no longer consisting of octagonal pavilions and of paths that turn back on themselves but of rivers and provinces and kingdoms. I thought of a labyrinth of labyrinths, of a meandering, ever-growing labyrinth that would encompass the past and future and would somehow take in the heavenly bodies.

The Garden of Branching Paths, 1941

Jorge Luis Borges

Declaration

I hereby declare that the contents of this dissertation are my own, except where work which has formed part of jointly authored publications has been included. My contribution and the other authors contribution to this work has been explicitly indicated in the text. 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.

Miguel Vioque de Lózar

August 2020

Preface

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

Refereed publications

1. **Vioque, M.**, Oudmaijer, R. D., Schreiner, M., Mendigutía, I., Baines, D., Mowlavi, N., and 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.

2. Wichittanakom, C., Oudmaijer, R. D., Fairlamb, J. R., Mendigutía, I., **Vioque, M.**, and Ababakr, K. M. (2020). *The accretion rates and mechanisms of Herbig Ae/Be stars*. MNRAS, 493(1):234-249.

3. **Vioque, M.**, Oudmaijer, R. D., Baines, D., Mendigutía, I., and Pérez-Martínez, R. (2018). *Gaia DR2 study of Herbig Ae/Be stars*. A&A, 620:A128.

Non-refereed publications

1. **Vioque, M.**, Wichittanakom, C., Oudmaijer, R. D., Schreiner, M., Mendigutía, I., Baines, D., Mowlavi, N., and Pérez-Martínez, R. (2020). *Cataloguing new high-mass Pre-Main Sequence and Classical Be stars using Machine Learning and Gaia*. Contributions to the XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society, id. 192.

2. **Vioque, M.**, Oudmaijer, R. D., Baines, D., Pérez-Martínez, R. (2019). *New catalogue of intermediate mass Pre-Main Sequence objects in Gaia DR2 using Machine Learning*. 53rd ESLAB Symposium: The Gaia Universe, id. 52.

3. **Vioque, M.**, Oudmaijer, R. D., Baines, D., Mendigutía, I., Pérez-Martínez, R. (2019). *Gaia study on the formation of intermediate mass stars*. Highlights on Spanish Astrophysics X, Proceedings of the XIII Scientific Meeting of the Spanish Astronomical Society, p. 437.

4. **Vioque, M.**, Oudmaijer, R. D., Baines D. (2018). *Herbig Ae/Be stars with TGAS parallaxes in the HR diagram*. Astrometry and Astrophysics in the Gaia sky, Proceedings of the International Astronomical Union, 330:277-278.

5. **Vioque, M.**, Oudmaijer, R. D., Baines D. (2017). *HR diagram of Herbig Ae/Be stars and their infrared excesses*. Memorie della Societa Astronomica Italiana, 88:824-825.

The publications I sign as first author are included in their totality in this thesis. In particular, Vioque et al. (2018) and Vioque et al. (2020) almost entirely constitute Chapters 2 and 3, respectively. Everything in the publications I sign as first author is the result of my own work, upon revisions and suggestions of my supervisory team. Other co-authors made minor corrections at late stages in the preparation of the documents. Nothing from Wichittanakom et al. (2020) was included in this dissertation, except when direct reference is made to that publication. I note that my contribution to that work was very minor.

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

Acknowledgements

Without my parents' endless love and dedication this work, or any of my other achievements in life, would have not been possible. I owe them everything. The man I am today also would not have existed without the life education I received at my *alma mater*, the Colegio Mayor Isabel de España. The feelings of gratitude and love I profess to all the good friends I made there cannot be put into words.

The quality of this research and the results obtained must be credited to my project supervisors. I want to stress the importance of Professor René D. Oudmaijer as a science educator and mentor, both professionally and personally. I feel honoured that he signs as the main supervisor of this PhD thesis. The results of this dissertation are merely footnotes of his ideas. Dr. Deborah Baines and Dr. Ricardo Pérez-Martínez, in their respective fields of expertise, guided me through the always somewhat obscure paths of scientific politics and industry and were always ready to help and advise me. Although they were not in the supervisory team, I also owe gratitude to Dr. Ignacio Mendigutía, Dr. Benjamín Montesinos and Dr. Olja Panić, whose eagerness for astronomy and passion for work greatly influenced this project and brought me joy and resolution in moments of doubt. In addition, I wish to extend these acknowledgements to the Centro de Astrobiología personnel, who treated me with immense affection and kindness during my secondment there.

As in any period of life, my time in Leeds did not pass without some moments of sorrow. The days that was not the case can be correlated with my presence at the Fenton pub or in the company of the extraordinary people I met throughout it. In particular, Lena Kilian gave me the strength to finish this project with the same resolution with which I started it, and Dr. James Miley helped me find the next life adventure with unforgettable honour and elegance.

Lastly, I want to extend my feelings of gratitude to Antonio Rodríguez Cerrillo, my secondary school physics teacher, whose enthusiasm and patience with a younger version of me led, in time, to the accomplishment of this doctoral thesis.

Abstract

Herbig Ae/Be stars are pre-main sequence sources, canonically defined as having masses 2 $M_\odot \lesssim M \lesssim$ 10 $M_\odot$, which are key to understanding the differences in formation mechanisms between low- and high-mass stars. The study of the general properties of these objects is hampered by the lack of a well-defined, homogeneous sample, and because few and mostly serendipitously discovered sources are known. As a consequence, many open problems involving high-mass star formation suffer from biases and lack of completeness.

I study the general properties of the 252 known and proposed Herbig Ae/Be stars with parallaxes from Gaia DR2 at the start of this doctoral thesis. High-mass stars were found to have a much smaller infrared excess and optical variability compared to lower-mass stars, with the break at around 7 $M_\odot$. Different or differently acting (dust-) disc-dispersal mechanisms are proposed for this break. The variability indicator developed in this dissertation shows that $\sim 25\%$ of all Herbig Ae/Be stars are strongly variable. Evidence is provided to support the hypothesis that this variability is in most cases due to asymmetric dusty disc structures seen edge-on.

Using that well-characterised sample of Herbig Ae/Be stars as a training sample for a bespoke machine learning algorithm, a homogeneous and well-defined catalogue of 8470 new pre-main sequence candidates was obtained. In parallel, a catalogue of 693 new classical Be candidates was produced. At least 1361 sources are potentially new Herbig Ae/Be stars according to their position in the Hertzsprung-Russell diagram. This increases the number of known objects of the class by an order of magnitude.

In addition, I discuss the results of independent spectroscopic observations conducted for a selection of 145 new Herbig Ae/Be and 14 new classical Be stars. These independent observations further confirm the quality and the accuracy of the classification. I conclude with an analysis of the general properties of the new catalogues that validates the results and conclusions obtained for the set of previously known Herbig Ae/Be stars.

Table of contents

Nomenclature

Acronyms / Abbreviations

2MASS	Two Micron All-Sky Survey
AGB	Asymptotic giant branch
AI	Artificial intelligence
ALMA	Atacama Large Millimeter/submillimeter Array
ANN	Artificial neural network
APASS	The American Association of Variable Star Observers Photometric All-Sky Survey
BeSS Database	Be Star Spectra Database
BSC	Bright Star Catalogue
CAFOS	Calar Alto Faint Object Spectrograph
CAHA2.2m	Calar Alto 2.2 metres telescope
CBe	Classical Be
CCD	Charge-coupled device
CK-model	Castelli and Kurucz (2003) model
CV	Cross-validation
D	Deuterium
DR2	Data Release 2
EFOSC2	European Southern Observatory Faint Object Spectrograph and Camera (v.2)
EW	Equivalent width

FN	Number of false negatives
FP	Number of false positives
GMC	Giant molecular cloud
HAe	Herbig Ae
HAeBe	Herbig Ae/Be
HB catalogue	Herbig-Bell catalogue
HBe	Herbig Be
HI	Neutral atomic hydrogen
HII	Ionised atomic hydrogen
HR	Hertzsprung-Russell
IDS	Intermediate Dispersion Spectrograph
IMF	Initial mass function
IMTT	Intermediate Mass T Tauri
INT	Isaac Newton Telescope
IPHAS	INT Photometric Hα Survey of the Northern Galactic Plane
IR	Infrared
ISM	Interstellar medium
KS	Kolmogorov-Smirnov
MIR	Mid-infrared
ML	Machine learning
MS	Main sequence
MYSO	Massive Young Stellar Object
NIR	Near-infrared
NTT	New Technology Telescope
P	Precision
PARSEC	Padova and Trieste Stellar Evolution Code
PCA	Principal component analysis
PMS	Pre-main sequence

R	Recall
RUWE	Re-normalised unit weight error
SDSS	Sloan Digital Sky Survey
SED	Spectral energy distribution
SIMBAD	Set of Identifications, Measurements and Bibliography for Astronomical Data
SoSt	Sample of Study
SPIRE	Spectral and Photometric Imaging Receiver
TP	Number of true positives
UMAP	Uniform Manifold Approximation and Projection
UV	Ultraviolet
UXOR	UX Ori type star
VPHAS+	VST Photometric Hα Survey of the Southern Galactic Plane and Bulge
WEAVE	William Herschel Telescope Enhanced Area Velocity Explorer
WISE	Wide-Field Infrared Survey Explorer
YSO	Young stellar object
ZAMS	Zero age main sequence

List of figures

List of tables

Chapter 1

Introduction

Every sentence I utter must be understood not as an affirmation, but as a question.
New York Times Book Review, 1957 – Niels Bohr

In this introduction, which does not intend to be exhaustive, I summarise for the reader the current state of the star formation field, focusing on the aspects that this dissertation covers in a major depth. I describe star formation chronologically, from large to small scales, and present the motivation for this thesis. In addition, I introduce some of the more common telescopes, tools and instrumentation used to describe and study the discussed topics.

Following the tacit agreement in modern science, I barely mention what is still unknown or poorly understood – which are most things – and restrict myself to describe the few things that seem to have endured the test of the scientific method. As a disclaimer, what is presented here is a gruesome simplification. Many physical interpretations, models, and thoughts have been omitted either for simplicity or because of being beyond the scope of this dissertation.

I chose it necessary and educative to start with a brief historical introduction as the results of this thesis, though original, stand on top of decades of previous efforts without which this work would not have been possible at all. Finally, I conclude with the thesis outline.

1.1 Historical overview

Questions about the internal structure and formation mechanisms of the stars can be traced back to the dawn of natural philosophy. However, limited by technology, stellar structure and evolution theory remained in relative darkness until the recent period of scientific and technological revolution that followed the Second World War. Astronomy, an interdisciplinary field, benefited from the development of the different branches of fundamental physics. At the same time, newer and better generations of ground and space telescopes successively set new eyes to observe deeper into the cosmos, bringing the distant stars slightly closer to our understanding.

One of the first attempts of modern astronomy, what we now call *astrophysics*, to understand the internal mechanisms of the stars was that of Julius Robert von Mayer (1814-1878) in 1848. At that time one of the most fundamental laws of physics, the conservation of energy, started to be understood (first stated by Mayer himself; Mayer, 1842). Understanding how the Sun could generate a huge amount of energy for what seemed to be an unlimited or immeasurable amount of time was therefore the next logical step. In short, he proposed that the Sun's source of heat might be due to the capture of meteors from the surrounding space (Mayer, 1848). This original explanation received, in time, independent support by other astronomers but was practically abandoned by the 1870s, as the number of meteoric bodies required to match the observations was found to be unreasonably high.

On February 7th 1854, a young Hermann von Helmholtz (1821-1894) gave a lecture in Königsberg in which he proposed another gravitational theory to explain the Sun's luminosity. It was due to a small secular contraction. It might not be a coincidence that he chose Königsberg to give such talk, as this contraction idea goes back to an essay from Immanuel Kant (1724-1804) from 1785 (Watkins, 2012; originally published in the *Berlinische Monatsschrift*), in which he wrote: *If it be assumed . . . that the original matter of all celestial bodies, in the whole vast space*

in which they now move, was initially distributed in gaseous form, and was formed initially in accordance with the laws of chemical attraction. In the second half of the 18th century it was broadly accepted that the stars were not much different than our Sun, but they were located at much greater distances. Kant suggested that stars were formed from cold clouds of gaseous material. Since gas is heated up if compressed, this formation mechanism would also be responsible for the surface heat of the stars and perhaps for the planets around them. This impressive piece of bright instinct, as can be seen later in this introduction, was known but ignored for decades, and only marginally and anecdotally considered until Helmholtz's lecture. As Kant, Helmholtz thought that the Sun's luminosity was linked to the hypothesis of a nebular origin of the stars. An even more primitive intuition of star formation being produced by the condensation of diffuse matter in the Universe was written down by the father of gravity itself, Sir Isaac Newton (1642-1727): ... *if the matter were evenly disposed throughout an infinite space, ..., some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered great distances from one to another throughout all that infinite space. And thus might the sun and fixed stars be formed, supposing the matter were of a lucid nature* (as quoted by Jeans, 1928).

Sir William Thomson, 1st Baron Kelvin (1824-1907), who had a reputed career addressing the Sun's source of energy problem, adopted the contraction theory of Helmholtz and made significant contributions to it. This theory robustly stood for over forty years until it was proved that it could not explain the source of stellar energy. Geological evidence showed that the Sun was older than what the theory could predict. Today, the Kelvin–Helmholtz mechanism applies to many other astronomical scenarios, including the star formation processes (see Sect. 1.4).

The situation did not improve much in the following decades. It was agreed that, by process of elimination, the source of stellar energy needed to have a subatomic origin of unclear nature: *by a process of exhaustion we are driven to conclude that the*

only possible source of a star's energy is subatomic (Eddington, 1926). The answer as we understand it today came with the revolution of nuclear physics in the 1930s (see Chandrasekhar, 1939 and references therein), and it could be said that the story ends with the Nobel Prize in Physics of 1967 to Hans Albrecht Bethe (1906-2005) for his work on the creation of chemical elements within stars by nuclear fusion reactions, which generate the amount of energy required to match the observations. However, the understanding of how the primordial masses necessary to start nuclear reactions were assembled (i.e. the formation of the stars) evolved little from Kant's intuition until the mid-20th century. George Herbig (1920-2013) expressed this way which was the state of the star formation field before 1930: *One senses the attitude that the origin of stars was something unknown and unknowable: it happened in the remote past under circumstances that we were unlikely to fathom* (Herbig, 2002).

Earlier in the 20th century, new 'large scale surveys' caused the need for systematisation and the aspiration to find big correlations. One of the most successful endeavours for systematisation was the Harvard classification of stellar spectra by Annie Jump Cannon (1863-1941) and Edward Charles Pickering (1846-1919), which at the present time is still broadly used. Cannon found a phenomenological correlation between the line structure and the colour of the stars, and they ordered the different spectral types in a sequence (O, B, A, F, G, K, M) in the hope that it would represent some sort of stellar evolution (Cannon and Pickering, 1901). At that time, previous definitions of spectral type had been already correlated with colour, and it was an open debate in the community how the different 'spectral types' were related to the evolutionary status of the stars. Panel *b* of Fig. 1.1 shows a sketch Henry Norris Russell (1877-1957) used as lecture notes in 1907, in which he wonders about the correct sequence of 'spectral types' to describe the evolution of a single star. These Russell's sequences are among the first evolutionary tracks recorded and evidence the necessity in the early 20th century to find a proper, well-defined parameter space to construct the star formation and evolution theory from observations.

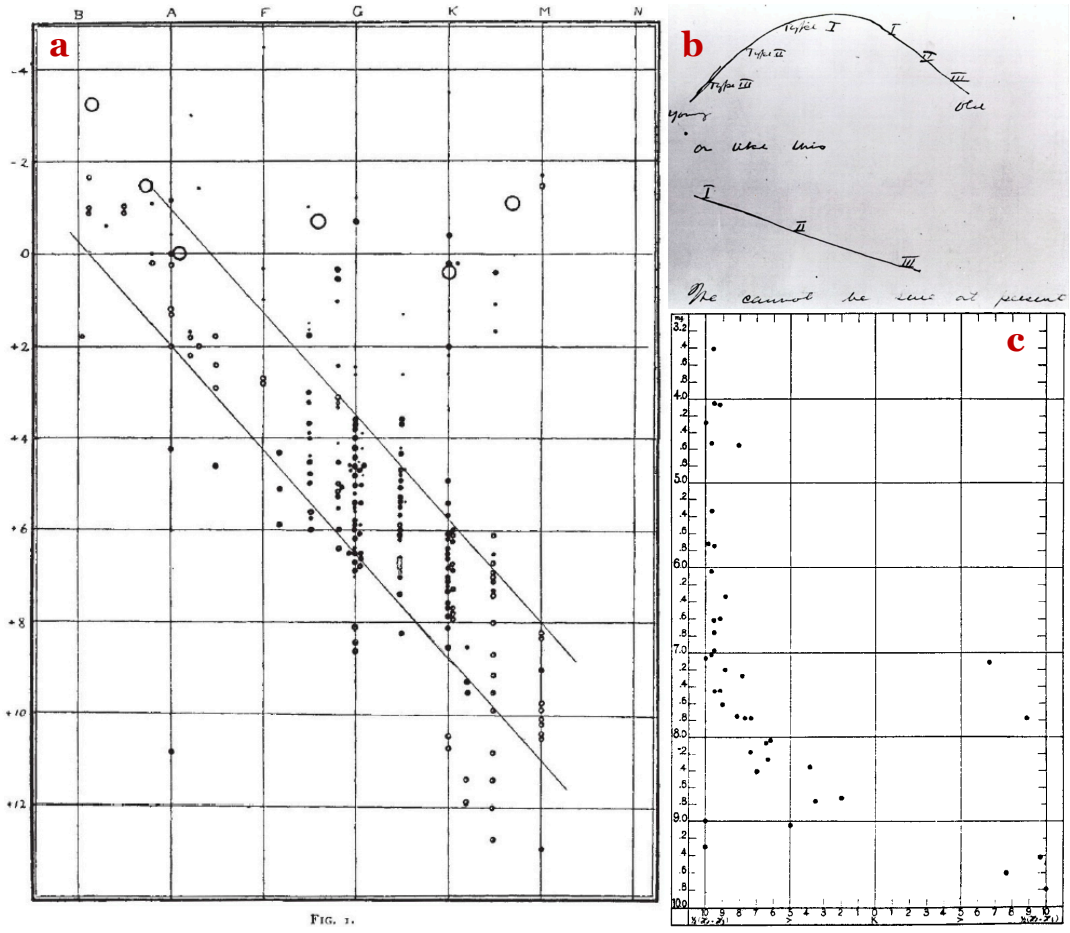


Fig. 1.1 *Panel a:* Russell's original spectral type vs. absolute magnitude diagram as it appeared in Russell (1914). The original version of the plot appeared in Russell (1913). *Panel b:* Russell's 1907 lecture notes showing two different sequences of stellar spectral classes that suggest possible stellar evolutions, from *young* to *old*. At the bottom it can be read: *We cannot be sure at present*. *Panel c:* Original colour-magnitude diagram of the Pleiades by Rosenberg (1910). Late B through late A stars can be seen vs. apparent magnitude.

Another one of the sought correlations was to link colour and spectra with absolute magnitudes. Although at the time it was very hard to directly measure distances, it was soon established that this issue could be avoided by the use of clusters. The first, as we now call it 'colour-magnitude diagram', was published by Hans Rosenberg (1879-1940) in 1910 using the Pleiades and the Cannon classification (Rosenberg, 1910; see panel *c* of Fig. 1.1). It is noteworthy that he followed the suggestion of Karl Schwarzschild (1873-1916) and might have been inspired by previous unpublished attempts of Schwarzschild's student Ejnar Hertzsprung (1873-1967).

Hertzsprung published his first version of the diagram in 1911 (Hertzsprung, 1911), and continued with a series of papers with diagrams of the Pleiades, the Hyades, and other clusters. As it often happens in science, there is no evidence that the original diagram of Rosenberg was ever cited by his contemporary peers.

Russell pioneered the tedious process of systematically deriving parallaxes from photographic plates and obtained the remarkable number of 55 parallax fields by 1906. Russell, in a totally independent fashion (at least according to the papers he cited), published in 1913 the first spectral class vs. absolute magnitude diagram using nearby field stars (Russell, 1913; see panel *a* of Fig. 1.1). He noticed the correlation we now call 'main sequence' and even proposed an evolutionary model in which the stars begin their lives from nebulae as red giants, get to the tip of the main sequence (MS), and then move to cooler temperatures along the MS as they age. Later development, and the outstanding PhD thesis of Cecilia Helena Payne (1900-1979), which broke the degeneracy between abundance and temperature effects in spectral lines (Payne, 1925), confirmed the long considered but never entirely understood hypothesis that the Cannon classification of spectral types was indeed a temperature classification. Finally, both axes of Russell's, Rosenberg's, and Hertzsprung's diagrams acquired full physical meaning.

Although Russell's evolutionary theory did not stand for long, in the 1930s this diagram that relates temperature with luminosity was finally coined the Hertzsprung-Russell (HR) diagram. Since then, it has had a priceless impact on the understanding of star formation, the stellar evolution theories and the stellar classification, and it plays a central role in this PhD thesis.

For a field that evolved so slowly since antiquity, many things happened in the star formation field during the decades of 1940s, 1950s, and 1960s. Astronomers knew that the only sources of material for star formation were massive clouds or nebulae in the interstellar medium. They also knew that stars were not formed that long ago, as the recently discovered nucleosynthesis process allowed them to

calculate stellar lifespans, nor that far away, as the stars did not have much time to travel from their birthplace. Therefore, star formation should be a still ongoing process observable in the night sky, if it was visible at all.

Hoyle and Lyttleton (1939) were the first to propose that stars might acquire most of their mass from the interstellar medium using purely scientific and astronomical arguments. In Joy (1942), Alfred Harrison Joy (1882-1973) discussed a Mount Wilson spectrographic survey of variable stars. He wrote: *The T Tauri and RW Aurigae stars form a small but important group which deserves more attention than it has yet received. Some of the stars are apparently connected with or obscured by shells or wisps of nebulosity. The light-changes are rapid and entirely irregular...* Although in the following years he followed-up on this 'small but important group' of objects (see e.g. Joy, 1945, 1949), it took time until these T Tauri like objects were associated with the efforts of studying nebulae and the star formation within them (led by Fred L. Whipple, Bartholomeus Bok and Lyman Spitzer, among others).

According to Herbig (2002), credit should be given to the Armenian astronomer Viktor Amazaspovich Ambartsumian (1908-1996), who in 1947 proposed for the first time that T Tauri stars were undergoing a formation process through contraction (Ambartsumian, 1947). However, his work and the subsequent research of the Soviet astronomical community did not get much attention in the West, and such theory was not seriously considered in Europe and America until 1954. In 1953 Edwin Ernest Salpeter (1924-2008) spotted that forming stars should have a particular location in the HR diagram from where young objects could be identified (Salpeter, 1954). After a few years of many papers wandering around the evidence, the idea that the T Tauri class is composed of stars undergoing formation processes finally rooted in the astronomical community. In Herbig (1962), to summarise the scientific debate and results of those years, Herbig stated: ... *it is concluded that no acceptable alternative has yet been developed to the interpretation of the T Tauri stars as young objects undergoing their initial gravitational contraction.*

Fig. 1.2 *Left:* George Herbig observing at 20, 1940. *Right:* George Herbig at 90 observing at Keck-I telescope, 2010. Figure from Reipurth (2016).

In 1960, George Herbig (in Fig. 1.2), using arguments similar to those Joy used to find the T Tauri group, looked for the massive counterparts of the T Tauri stars (Herbig, 1960). He imposed three criteria to identify his massive young candidates:

- The spectral type is A or earlier, with emission lines.
- The star lies in an obscured region.
- The star illuminates fairly bright nebulosity in its immediate vicinity.

These famous three conditions led to the discovery of the first 26 sources of the class now called 'Herbig Ae/Be', which is the central topic of this PhD thesis. Since then, many contributions have been made to the understanding of these objects and their formation mechanisms. Currently, none of Herbig's original criteria stands as a necessary requirement of the Herbig Ae/Be group, although these simple conditions shall withstand as historical proof of scientific elegance. As the description of more recent works may exit the realm of history and enter the domains of journalism, I refer the interested reader to the bibliography of this PhD thesis and the references therein contained to judge the contributions made from here.

1.2 Molecular clouds, filaments and cores

The Universe is populated by many galaxies. Different galaxies have different star formation rates. This rate mainly depends on the amount of gas within them; which in turn varies with age, dynamical circumstances, and galactic environment (see Fig. 1.3). The current star formation rate in the Milky Way is 1.65 ± 0.19 $M_{\odot}$/year (Licquia and Newman, 2015). Although the discussion about stellar evolution in distinct galaxies throughout the history of the Universe is of the major interest, for simplicity I restrict myself to describe the star formation observed at the local scale, in the vicinity of our Sun in the Milky Way. Star formation in any place and time should be expected to be similar, at least in the fundamental principles.

1.2.1 Molecular clouds

The Milky Way is full of gas that permeates the space between the stars; the interstellar medium (ISM). Around half of the Galactic volume is occupied by very hot gas, mainly ionised atomic hydrogen (HII), that was ionised and heated up by supernovae (the big regulators of the ISM). This gas expands in bubbles and vertically through the disc, dissipating in the Galactic halo. The high temperature of $10^6 - 10^7$ K makes the density of this HII gas to be of only $10^{-4} - 10^{-2}$ particles/cm^3. In pressure balance with this super hot gas we find the ionised and neutral warm ISM, which is composed primarily of neutral and ionised atomic hydrogen (HI and HII respectively). At a much lower temperature of around $6000 - 10000$ K, the typical density of this medium is about 1 particle/cm^3. These two states of the ISM sum up to $\sim 95\%$ of the volume of the Galaxy. The remaining 5% corresponds to the cold ISM: mostly cold HI concentrated in the Galactic plane that have a temperature of $50 - 100$ K and a density of a few tens of particles/cm^3.

Through the collision of flows of HI, huge complexes of cold molecular gas are formed (see e.g. Valdivia et al., 2016). These complexes span hundreds or even

thousands of parsecs and are distributed in patchy regions mostly located in the spiral arms. Again, this molecular gas is primarily composed of molecular hydrogen (H_2). Cold molecular hydrogen does not have energy transition levels in the optical, infrared or radio and hence it is technically invisible. Therefore CO, one of the most abundant molecules, is normally used to trace molecular gas. The distribution of molecular gas as traced by CO in 126 close-by galaxies is presented in Fig. 1.3.

Astronomy is a tale of scales, structure, and hierarchical substructure within the scales. The huge complexes of cold molecular gas are subdivided into the so-called giant molecular clouds (GMCs), which typical sizes are in the order of $20-100$ pc and have a mass of $\sim 10^5-10^6$ $M_\odot$. These GMCs are opaque to ultraviolet radiation as it readily dissociates the molecules. This opacity is due to a density of $\sim 10^2-10^6$ particles/cm^3. Therefore, GMCs are super-dense compared with the rest of the ISM. The opacity causes these clouds to be little heated by external radiation. In addition, they are very efficient radiating energy away (CO is also the most important coolant; see e.g. Whitworth and Jaffa, 2018). As a consequence, GMCs are at extremely low temperatures of only $\sim 10-20$ K. An example of GMC, the Polaris Flare cloud, is presented in Fig. 1.4 in the top panels.

There is one extremely important component of the ISM yet to be mentioned; dust. Dust particles of a typical size of 1 μm are present throughout the Galaxy, although they are promptly sublimated at temperatures higher than $1000-2000$ K (depending on composition, Kobayashi et al., 2011). They are of the utmost importance at every scale. They regulate many thermodynamic processes, influence GMCs dynamics and are key for planet formation. Furthermore, most ISM chemistry happens on the surface of dust grains (e.g. H_2 formation). The high density of dust particles causes molecular clouds to also be opaque to visible light but transparent to longer wavelengths. Overall, the ISM is a dynamical and complex environment. Although a detailed description of it is beyond the scope of this introduction, a good book of reference is Tielens (2005).

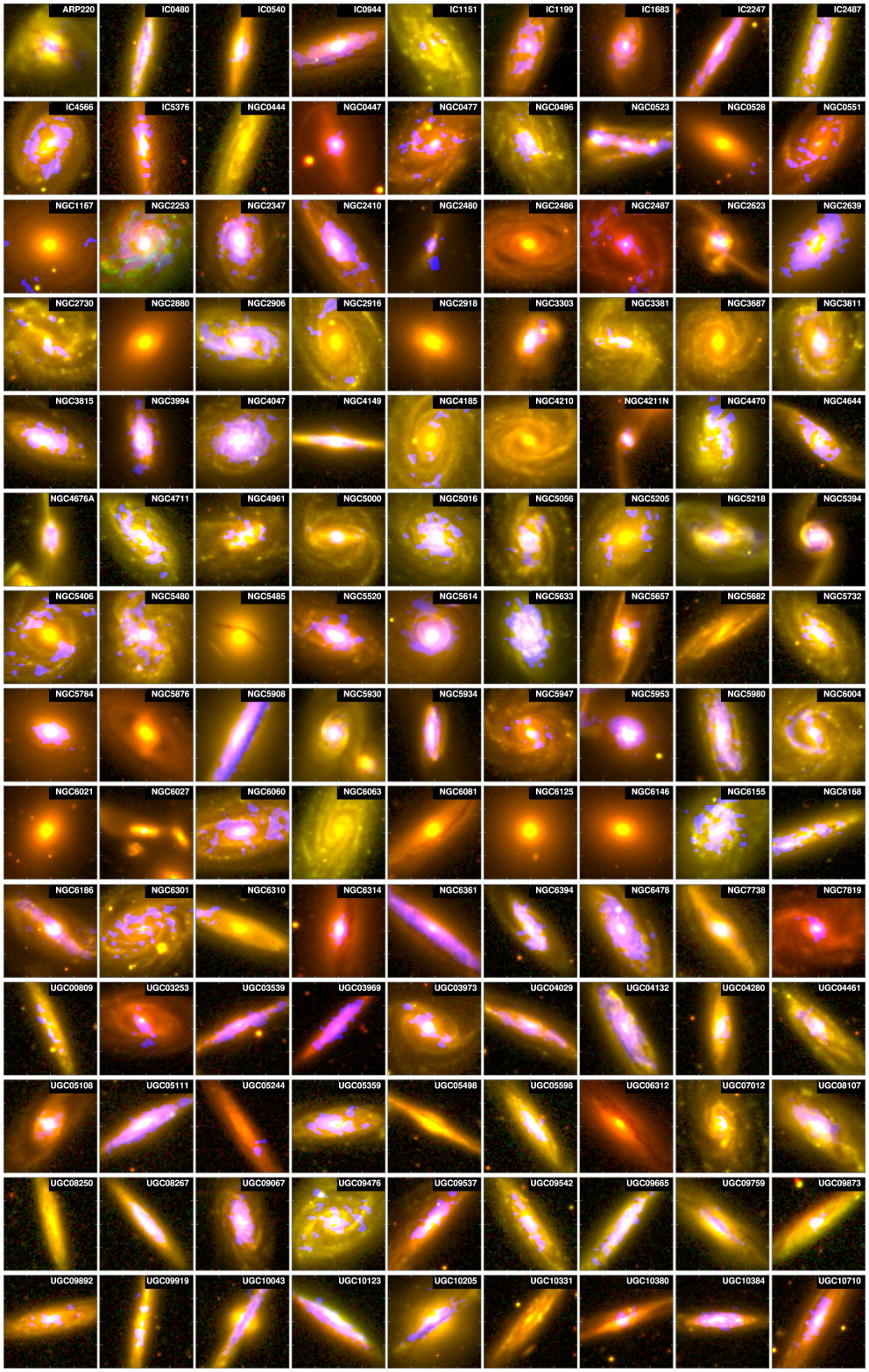


Fig. 1.3 Distribution of stars and molecular gas in 126 local galaxies. The CO intensity map is shown in blue. The stellar light is composed of SDSS *g* (green channel) and *i* (red channel). Figure from Bolatto et al. (2017).

Molecular clouds have complex internal motions. This can be summarised in Larson's empirical law (Larson, 1981), which positively correlates the size of a cloud with its internal velocity dispersion. At the present time, the physics behind Larson's relation is not well understood. It seems to imply that within molecular clouds gas velocities are dominated by a turbulent motion rather than a thermal motion. 'Turbulence' means some kind of size hierarchy in the random motions of the particles, that can generate density fluctuations and shocks (see e.g. Larson, 2003 and Hennebelle and Falgarone, 2012).

One of the key consequences of their turbulence-dominated internal motions is that GMCs do not quickly collapse under the effect of gravity to form stars. On the contrary, they seem to be rather inefficient in forming stars. On large scales, the star formation rate is sensitive to the Galactic environment (e.g. Galactic shear, spiral arm potential or vertical gravity from gas, stars, and dark matter), whereas at smaller scales star formation is subjected to local environmental conditions. These can be summarised into dynamical effects (e.g. rotation, turbulence), gas compression (mainly gravity), chemical heating (e.g. cosmic rays, UV radiation), mechanical heating (e.g. supernovae), abundance of gas and magnetic fields. Significant discrepancies have been found between these various factors in different GMCs, although their precise impact in the star formation efficiency is as yet undetermined (see e.g. Geen et al., 2017). Nonetheless, most observed GMCs show active star formation.

There is the general agreement that regardless of circumstances GMCs cannot last much longer than ~ 10 Myrs, as they cannot survive for long after star formation begins within them. Soon they are dissipated by photoionisation from HII regions, stellar winds, supernova blasts and jets from forming stars, among other things (see e.g. Pellegrini et al., 2011, Federrath et al., 2014, Rosen et al., 2014 or Walch et al., 2015). Thus, it is evident that star formation needs to be a rapid process that begins promptly after the formation of the cloud (see e.g. Povich et al., 2019). All these processes in which new stars give energy back to the cloud are known as 'feedback'.

1.2.2 Filaments

Internally, GMCs organise themselves into a filamentary sub-structure. Filaments within the Polaris Flare cloud are presented as an example in Fig. 1.4 in the top panel. This cloud is not forming stars yet, implying that filaments appear prior to stars and their formation is independent of internal stellar feedback. Filaments can be defined (from André et al., 2014) as: *elongated ISM structures with an aspect ratio larger than* $\sim 5-10$ *that is significantly over-dense with respect to their surroundings*. Therefore, they are not necessarily of a cylindrical geometry, and the possibility of some of them being 'sheets' is considered. The typical width of the filaments is observed to be of around 0.1 pc.

The magnetic field, which is always hard to measure, is found to be orthogonal to the main filaments with a typical strength of $\sim 1-100$ μG (Matthews et al., 2014; Pillai et al., 2015). On the other hand, the magnetic field is aligned with the so-called 'striations', which are also perpendicularly disposed to the filaments. Striations by all means satisfy the definition of filament of André et al., 2014, but they differ in that they are much less dense (Tritsis and Tassis, 2016). It is not yet known whether they are long-lived or transient, but they do not constitute a site for star formation. In Fig. 1.4 in the bottom left panel a filament is presented together with the measured magnetic field in the area and the surrounding striations.

Many studies have been dedicated to the internal kinematics of filaments (e.g. Smith et al., 2016, Hacar et al., 2017 or Dhabal et al., 2018), which found signs of rotation and velocity gradients both along and across the filaments. However, their formation mechanism is still under debate. There are two main competing scenarios; one dominated by turbulence and gravity and one dominated by magnetic fields (or 'magnetohydrodynamic mechanisms'). The problem lies in that filaments can easily be recreated in simulations with and without strong magnetic fields. Cloud to cloud collisions or even thermal instabilities have also been proposed for the

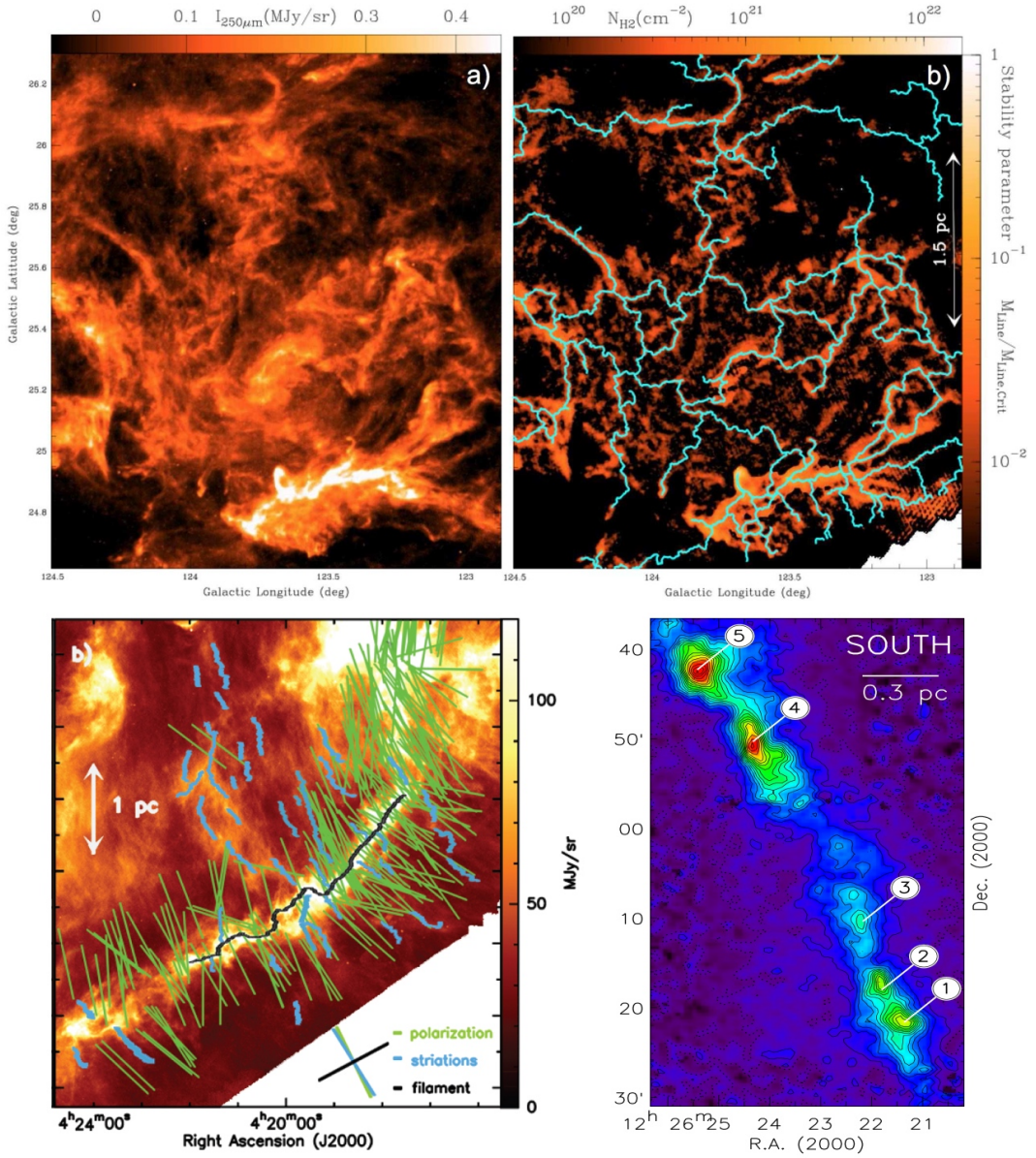


Fig. 1.4 *Top left:* Herschel/SPIRE 250 µm dust continuum emission of the Polaris Flare molecular cloud. *Top right:* Column density map of the same region with the detected filaments in light blue. The typical width of the filaments is 0.1 pc. Figures from André et al. (2014). *Bottom left:* Herschel/SPIRE 250 µm dust continuum emission of a filament in Taurus molecular cloud. The filament is highlighted with a black line and some striations around it are highlighted with blue lines. Polarisation vectors that trace the magnetic field are presented in green. In the bottom right an average of the relative orientation of the lines is shown. Note that the magnetic field and the striations are perpendicular to the filament. Figure from Palmeirim et al. (2013). *Bottom right:* Column density map of hydrogen (atomic and molecular) from near-infrared observations of the Musca cloud. Five cores are numbered in a filament under fragmentation. Figure from Kainulainen et al. (2016).

formation of filaments. Of course, models that combine the previous ingredients in different fashions also populate the literature. Interesting references are: Molinari et al. (2010), André et al. (2014), Banda-Barragán et al. (2016), Federrath (2016), Smith et al. (2016), Clarke et al. (2017), Wu et al. (2017) and Inoue et al. (2018).

1.2.3 Cores

If we move down one level in the structure ladder these filaments fragment into dense, roughly spherical cores (observed in e.g. Beuther et al., 2015; Contreras et al., 2016; Kainulainen et al., 2016). If these cores in turn lead to a spherical collapse they can be considered the first real stage of star formation (see Fig. 1.5 and Sect. 1.3). An example of a filament fragmented into different cores is shown in Fig. 1.4 in the bottom right panel. This fragmentation can be reproduced by simulations (e.g. Gritschneder et al., 2017; Lee et al., 2017) and it is caused by a gravity dominated process. In addition, material appears to be funneled along filaments towards 'hubs', points where various filaments intersect, at a typical timescale of $1-4$ Myrs (Peretto et al., 2014; Tackenberg et al., 2014).

These hubs seem to be the locus of cluster formation (e.g. Kirk et al., 2013; Baug et al., 2018) and massive star formation (e.g. Peretto et al., 2013; Yuan et al., 2018). These are not independent concepts. Clusters are formed through a further fragmentation of the core into smaller sub-cores; fragmentation caused by a combination of magnetic fields, turbulence and the delicate equilibrium between pressure and gravity (see e.g. Hennebelle et al., 2011 and Fontani et al., 2016). Most stars form in clustered environments, but this is particularly acute for high-mass stars as $\sim 95\%$ of them are observed associated with a cluster origin (see Portegies Zwart et al., 2010; Kruijssen, 2012; Gvaramadze et al., 2012 and references therein).

The amount of cores formed as a function of their masses is described by the so-called 'core mass function' (CMF). Equivalently, there is an 'initial mass function'

(IMF) that describes the proportion of stars formed at a given mass. These functions are described both locally in individual clouds and universally, assuming they are a fundamental feature of star formation. The connection between the two is not trivial, but there seems to be indeed a direct correlation, this meaning that the final mass of the stars is heavily influenced by the mass of the cores (Offner et al., 2014; Guszejnov and Hopkins, 2015). Nonetheless, Motte et al. (2018) recently provided evidence that implies that this might not be entirely the case. Observationally, there is a mass offset of a factor of three between the two which is consistent among the different star forming regions (André et al., 2010, Sokol et al., 2019). This implies that at this scale star formation is still considerably inefficient, as most of the mass does not end as stellar material. This is mainly due to radiation feedback (Krumholz et al., 2016), outflows (Offner and Arce, 2014), core fragmentation that can lead to multiple systems (Pineda et al., 2015) and other effects like the existence of starless, perhaps sterile, cores.

1.3 Collapsing cores and protostars

In this section, I start an overview of the different stages of star formation, which are recreated in an artistic impression in Fig. 1.5. After the previous section dedicated to the interstellar medium and to the dynamics and structure of molecular clouds, we are now at the level of single cores with a typical size of ~ 0.1 pc (or ~ 20000 au, see panels *1* and *2* of Fig. 1.5) and a typical lifetime of $\sim 0.1 - 0.3$ Myr (see Offner and Chaban, 2017 and references therein). The balance of forces and interactions in these cores have as main actors the magnetic field, thermal pressure and gravity. Depending on the interplay, collapse starts or not. Many derivations from the fundamental laws of physics have been used to decide this dichotomy. Each one with different subtle assumptions. Commonly used parameters to quantify the typical mass and length that can undergo contraction are variations of the Jeans mass and length (or of the

more robust Bonner-Ebert mass and length) either in isothermal or non-isothermal conditions. I refer the interested reader to Larson, 2003 and Ward-Thompson and Whitworth, 2011 for a proper definition and discussion of these parameters and their implications. For the sake of our story, let's continue with the cores in which collapse begins and follow the evolution of a single core assuming no fragmentation or halt. Collapsing cores also receive the name of Class 0 objects, as they constitute the first stage of star formation (panel *2* of Fig. 1.5).

Collapsing cores are subjected to the virial theorem, and hence half the gravitational energy acquired is radiated away and half goes to increase the thermal energy. Nonetheless, as mentioned before the molecular gas and the dust are very efficient at cooling the system. Hence, the collapse is roughly isothermal up to densities of around 10^{10} particles/cm^3. The characteristic time of the collapse can be approximated with the free-fall time, the time that would take a spherical body of uniform density to collapse under its own gravity:

$$t_{ff} = \left(\frac{3\pi}{32G\rho}\right)^{\frac{1}{2}}. \tag{1.1}$$

Note that the free-fall time is independent of the original size of the core and only depends on the density. However, the collapse is not uniform due to the outward forces, and it is characterised by the runaway growth of density at the centre, while the outer regions collapse slower. This implies that the global collapse is slightly slower than t_{ff}. More importantly, it implies that the centre of the collapsing cloud get to densities high enough to form a star while most of the cloud mass is still around as an infalling envelope.

At the typical free-fall time-scale, the central overdensity becomes optically thick. No longer isothermal, its temperature starts rising substantially from few tens of degrees as hydrostatic equilibrium is achieved. This hydrostatic core has a mass of ~ 0.01 M$_\odot$, a radius of several astronomical units and it is very short-lived

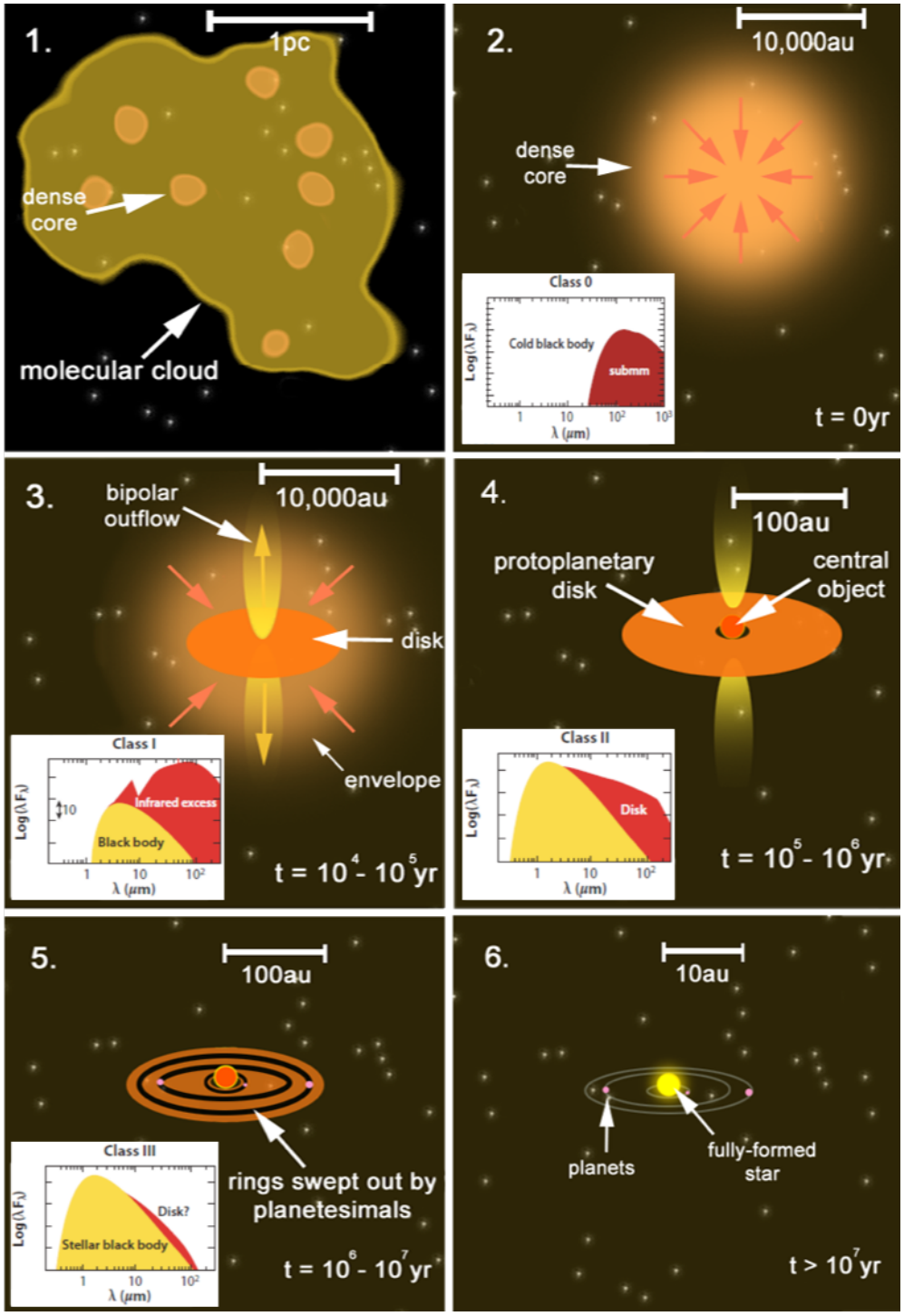


Fig. 1.5 Stages of star formation from molecular clouds (panel 1) to main sequence stars (panel 6). Core phase or Class 0 in panel 2. Protostellar phase or Class I in panel 3. Pre-main sequence phase or Class II in panel 4. Late pre-main sequence phase or Class III in panel 5. Credit of the image goes to A. J. Frost (Frost, 2020); spectral energy distributions from André (2002).

(Larson, 2003). As soon as its temperature rises above 2000 K H_2 dissociates and the hydrostatic core becomes unstable, provoking a second collapse (very high-mass stars, $\gtrsim 30$ $M_\odot$, might go directly to this second phase, Bhandare et al., 2018). In a similar fashion to the original contraction, this second contraction phase generates another central overdensity. Once this second central overdensity is mostly ionised collapse halts and a second hydrostatic core is formed. Rapidly, the first hydrostatic core collapses onto the second one (in the order of ~ 10 years) while most of the material from the original core is still in an envelope slowly falling onto the central overdensity. A protostar (or Class I object) has been born. See panel *3* of Fig. 1.5.

The central object and the envelope evolve decoupled, and the details on how accretion feeds the hydrostatic core at this stage determine what will be the final mass and radius of the star. Although I do not enter into much detail in this introduction, references for the protostellar phase are: Palla and Zinnecker (1999); Larson (2003); Wuchterl and Tscharnuter (2003); Stahler and Palla (2004); Baraffe et al. (2009).

1.3.1 Protostellar discs and outflows

There are two important ingredients that still need to be added to this picture, rotation and magnetic fields. A direct consequence of the increasing angular velocity upon contraction is that a disc is formed around the central overdensity (see Fig. 1.5). This disc, which is not stable and is likely to fragment (which can lead to binary systems), transports material inwards while it is able to remove angular momentum through viscosity and magnetic braking.

Another way of losing angular momentum is through the ejection of material. This happens through jets and wide-angle winds (see Fig. 1.5). Jets are a common feature observed in protostars (e.g. Kraus et al., 2010, Purser et al., 2016, Tafalla et al., 2017). When these jets collide with the parental cloud and are optically visible on the parsec scale they receive the historical name of 'Herbig-Haro objects'.

Wide-angle winds are also detected (e.g. Santiago-García et al., 2009) and arise from both the central protostar and the disc around it. The main difference is that jets are narrow and collimated whereas winds are broad, shell-like and angle-dependent. These outflows are bipolar, mostly ionised and roughly perpendicular to the disc rotation plane, this being particularly true for the jets, although some precession is also observed in them.

Outflows originate very close to the central protostar ($\lesssim$ 100 au, see e.g. Zanni et al., 2007). Material is ionised either by photoionisation or collisional ionisation and then accelerated centrifugally along magnetic field lines away from the disc. The rotation of the disc twists the magnetic field lines, which results in the collimation of the jet. The details and subtleties of the ejection mechanisms are complex, different models exist and the nature of these outflows is yet poorly understood (see Pudritz et al., 2007; Zanni et al., 2007; Seifried et al., 2012; Carrasco-González et al., 2015; Lee, 2020 and references therein).

1.4 Pre-main sequence stars

Now we are at the stage of a central stellar hydrostatic embryo of ~ 0.01 $M_\odot$ accreting material from the surrounding envelope and disc, which shocks the surface. At first, the accretion shock is optically thick and adiabatic, thus heating the outer layers of the protostar and expanding them. However, at some point around 4 $R_\odot$ (this value corresponds to a main sequence star of 1 $M_\odot$, but a similar picture should be expected for other masses up to at least ~ 10 $M_\odot$, see discussion in Sect. 1.4.2) the material surrounding the shock becomes optically thin, the protostar radiates efficiently and stops growing in size (Masunaga and Inutsuka, 2000, Larson, 2003). Now we have an optically thick protostar accreting from an optically thin medium, increasing its mass and radiating energy away due to conversion of gravitational potential. The characteristic lifetime of a structure at luminosity L in hydrostatic con-

ditions sustained only by gravitational contraction is given by the Kelvin–Helmholtz timescale (see Sect. 1.1):

$$t_{KH} = \frac{GM^2}{RL}, \tag{1.2}$$

being R the radius of the spherically symmetric object and M its mass. Note that massive stars ($\gtrsim 8$ M$_\odot$) have a $t_{KH} < t_{ff}$, implying that they evolve so fast that by the time they get to the main sequence the surrounding envelope is still in dynamical collapse around them. When the central protostar reaches ~ 0.2 M$_\odot$ deuterium (D) burning onsets due to high central temperatures of $\sim 10^6$ K, which halts the fast hydrostatic Kelvin–Helmholtz contraction and in turn makes the central protostar fully convective. Approximately 10^5 years after the core contraction began, the star has accreted most of its final mass and it has arrived to the pre-main sequence (PMS) phase (Class II objects, see panel *4* of Fig. 1.5). For the lower mass case, the envelope has now been mostly dissipated through outflows or has settled in the disc (now called 'protoplanetary disc'), from which accretion continues. The scale-size is about two orders of magnitude smaller than in the protostellar case (the typical disc radius is of order ~ 100 au). Outflows, if present, are much weaker and become quiescent with time.

It is worth mentioning that forming stars experience luminosity outbursts through all the stages of their formation, though these tend to get smaller in intensity with time. They are abrupt and varied, and are most likely caused by sudden bursts of accretion from the disc. When they are large and long-term they receive the name of 'FU Ori outburst' (they are often associated to the protostellar phase; see Vorobyov and Basu, 2005, Zhu et al., 2007, Hales et al., 2015, Hartmann et al., 2016) and when they are small, short-term, and repetitive they are known as 'EXor outburst' (see e.g. Stahler, 2017, Hales et al., 2018; the model star is EX Lup but history is capricious). These outbursts are neither a universal phase nor necessarily occur at early stages

of star formation (Herbig, 2008). However, Contreras Peña et al. (2019) found that outbursts in the Class I stage are ~ 10 times more frequent than during the Class II stage.

1.4.1 Hertzsprung-Russell diagram perspective

The PMS evolution is better understood on the Hertzsprung-Russell diagram. In Fig 1.6 a HR diagram of the star formation evolution is presented. Theoretical evolutionary tracks and isochrones of different masses and ages are indicated. Note that isochrones need a choice of an age 'zero' and the assumption that the first photospheres of all stars form at the same time, which is not true. In addition, the location of evolutionary tracks is currently under discussion and permanent revision, and depend upon many issues like, for example, the metallicity of each star.

Before entering into discussion the HR diagram needs further introduction, now from a purely physical perspective (see Sect. 1.1 for the historical introduction). On the right of Fig 1.6 we see the 'birthline', which can be defined as the line where stars become optically visible. Although during the protostellar phase the hydrostatic core radiates energy away thanks to the Kelvin–Helmholtz mechanism and the outflows (Sect. 1.3), its light is blocked by the surrounding cloud until the forming star emerges. This point mainly depends on the mass but also on each source environmental circumstances and, as mentioned before, might even not ever happen for very massive stars. On the other end, low-mass stars ($M \lesssim 1$ M$_\odot$) become optically visible even during the late protostellar phase. This can be seen in the almost vertical lines of Fig 1.6, the so-called 'Hayashi tracks', (Hayashi, 1966), where the core contracts hydrostatically from very high radii and luminosities at almost constant temperature. These objects are sometimes given a more technical definition of birthline as the point where the D-burning halts the fast core contraction (Wuchterl and Tscharnuter, 2003). Nevertheless, there are differing definitions of

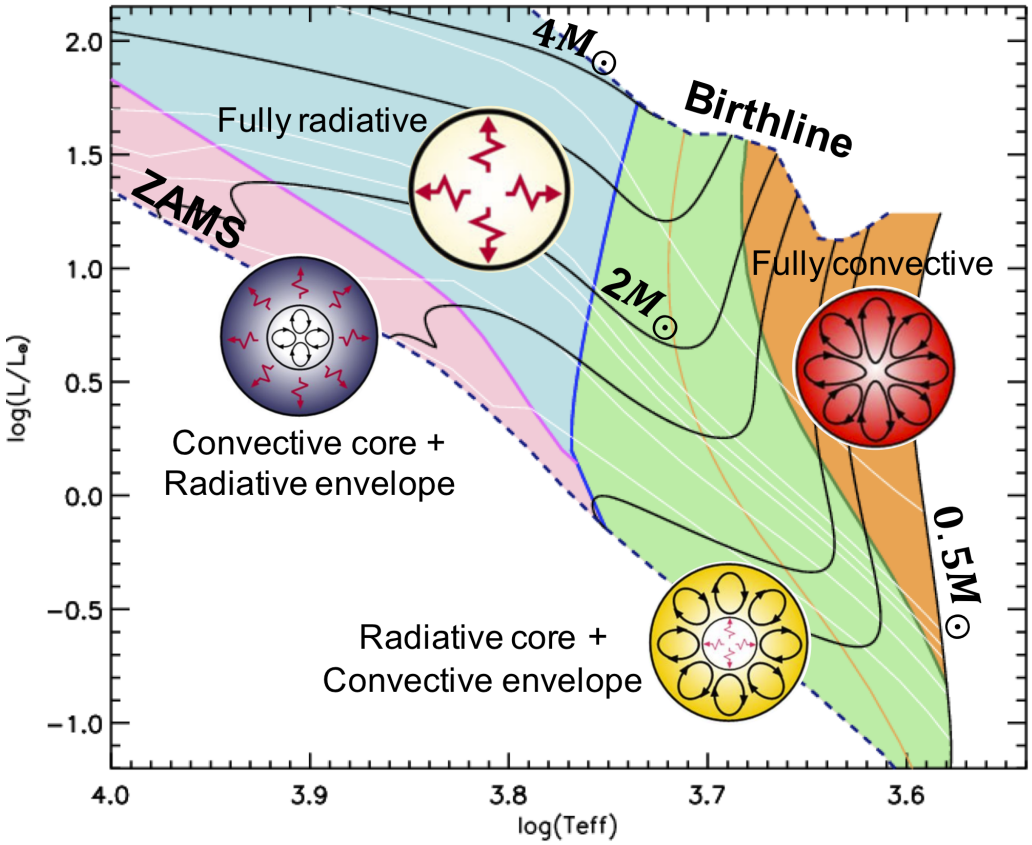


Fig. 1.6 Hertzsprung-Russell diagram of the optical phase of the star formation evolutionary sequence from the birthline to the zero age main sequence (ZAMS). Different colours trace different stellar interiors as indicated by the depictions. Evolutionary tracks appear in black (0.5, 0.8, 1, 1.5, 2, 3, and 4 $M_{\odot}$) and isochrones appear in white (1, 4, 5, 6, 10, and 20 Myr). The orange line indicates the location where 40% of the radius of the star is a convective envelope. Figure adapted from Alecian et al. (2013) and Villebrun et al. (2019).

the birthline in the theoretical literature, and for simplicity I stick to the purely observational definition regarding optical visibility.

The Hayashi tracks are caused by the development of a fully convective interior. In short, the large luminosity and radius cause a temperature gradient which is unstable to convection. Convection is very efficient and thus the temperature is kept almost constant upon contraction ($T \propto M^{1/4}R^{-1/8}$), this causing the almost vertical drop in luminosity. When the luminosity is low enough convective stability arises, diffusion settles throughout the interior and the temperature begins to rise. This part

of the evolutionary tracks at roughly constant luminosity (only dependent on mass, $L \propto M^3$) receive the name of 'Henyey tracks' (Henyey et al., 1955) and is the one that better represent the PMS phase. Note that there are different ways to set the limit between Class 0, protostars and PMS objects, which historically corresponded to where the spectral energy distribution (SED) peak occurs (see Fig. 1.5 and Sect. 1.5), but now those limits are often also defined from a theoretical perspective.

The zero age main sequence (ZAMS), on the very left of Fig. 1.6 is the point where contraction stops under the effects of nuclear reactions after D-burning, which is considered to be the moment when the star has completed its formation and joins the main sequence, where it remains stable for most of its life. The main sequence can also be defined as when significant H-burning begins, which can happen before the star reaches the ZAMS.

With all the aforementioned caveats and assumptions, it is easy to track forming stars on the HR diagram by the evolution of their temperature and luminosity (and thus their radius). In addition, in the HR diagram of Fig. 1.6 the different stellar interiors that develop during evolution towards the MS are also indicated. After the initial fully convective phase, stars develop a radiative core. Low mass stars ($M \lesssim 1$ $M_\odot$) get to the MS still with a large convective envelope, these receive the name of T Tauri stars. More massive stars ($M \gtrsim 1-2$ $M_\odot$) become fully radiative and only develop a convective core when they are about to reach the MS. These receive the name of Herbig Ae/Be stars (HAeBes, although they can be up to F type). Very massive stars are not optically visible at any point of their formation, and they receive the name of Massive Young Stellar Objects (MYSOs). The limit between HAeBes and MYSOs is uncertain and depends upon the particular environmental conditions of each source, but it is canonically set at $M \sim 10$ $M_\odot$ (see e.g. Lumsden et al., 2013). Finally, the sources that still have convective envelopes but that will become fully radiative (i.e. the precursors of the HAeBes) receive the name of Intermediate Mass T Tauri stars (IMTT). As explained in Sect. 1.4.2, the general picture presented

in Figs. 1.5 and 1.6 should be applied with caution to very massive stars (HBe and MYSOs). References for the theoretical evolution of stellar interiors in PMS objects are: Bressan et al. (2012), Choi et al. (2016), Feiden, 2016, Marigo et al. (2017), Munari et al. (2019), Villebrun et al. (2019) and references therein.

1.4.2 T Tauri stars, Herbig Ae/Be stars and MYSOs

As outlined in the previous section, there are three main categories of forming stars, T Tauri stars (and IMTTs), Herbig Ae/Be stars and Massive Young Stellar Objects. Before describing these in detail let us consider the magnetic fields first. Magnetic fields are hard to measure in PMS objects. Normally they are observed through direct Zeeman broadening or spectropolarimetry. The magnetic fields rapidly decay from the early PMS stages (Folsom et al., 2018). Only $\lesssim 10\%$ of Herbig Ae/Be stars have detectable magnetic fields, which rarely exceed 200 G. However, fields up to ~ 3000 G have been reported (e.g. Hubrig et al., 2010). The observed magnetic field in HAeBes is dipolar and stable, but it is not continuously sustained by a dynamo (Alecian et al., 2013; Villebrun et al., 2019). T Tauri stars have a magnetic field of $1000 - 5000$ G, which is variable and supported by a dynamo (Donati et al., 2013, 2011; Folsom et al., 2016). By 'dynamo' I mean a magnetic field generated by internal convection and rotation of charged particles. Therefore, convective stars tend to have strong magnetic fields, whereas radiative stars have small or negligible magnetic fields (Duez and Mathis, 2010). This implies that more massive objects are likely to be less magnetic (see Fig. 1.6). As a consequence, there is a significant difference between the magnetic fields of T Tauri and HAeBe stars, and this has important consequences on their accretion mechanisms.

It is commonly accepted that T Tauri stars accrete through magnetically-funnelled flows arising from the protoplanetary disc, which is truncated by the magnetic field lines at a distance of a few stellar-radii (see Bouvier et al., 2007; Hartmann et al.,

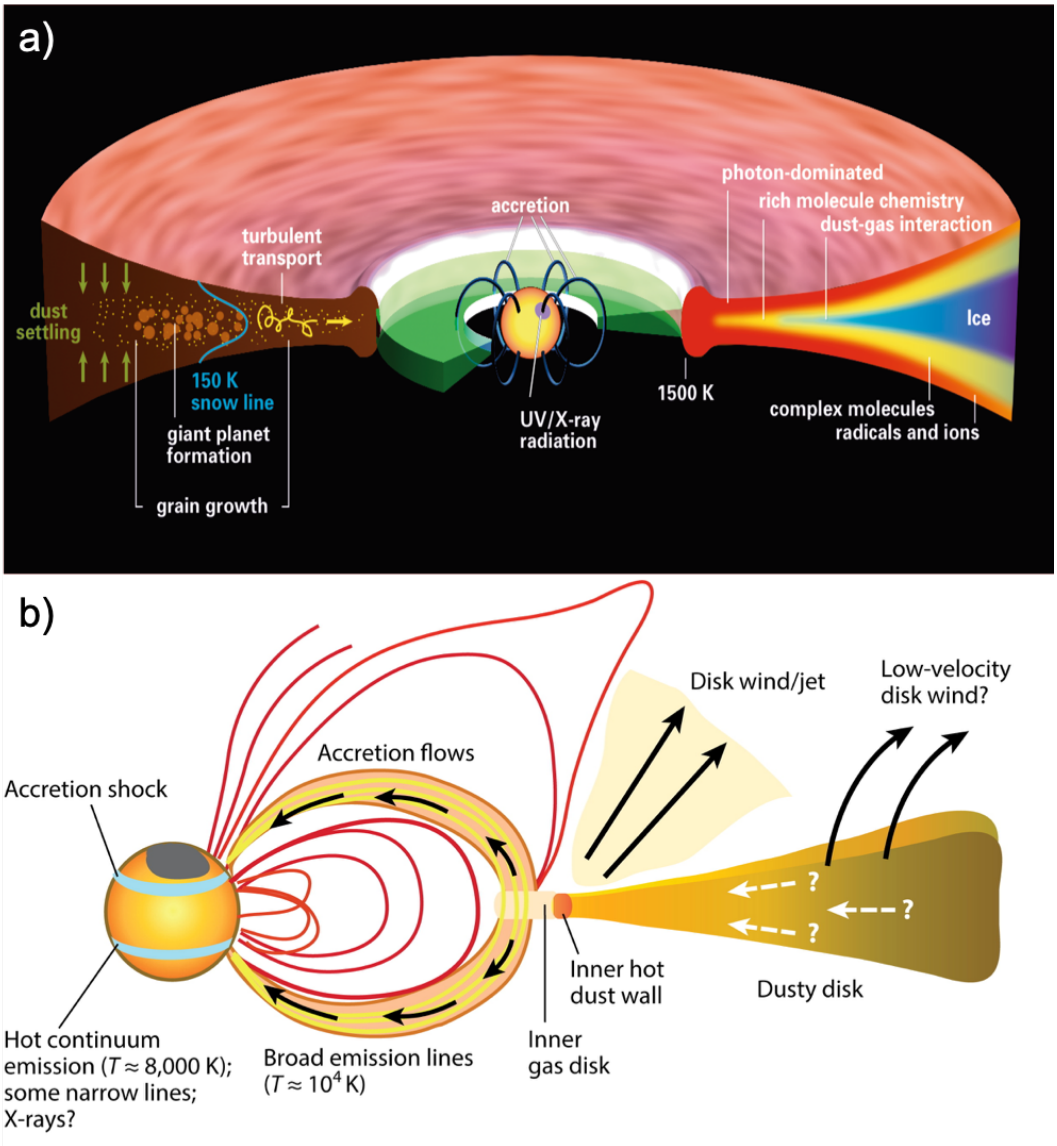


Fig. 1.7 *Panel a:* Protoplanetary disc around a PMS star accreting magnetospherically. Figure from Henning and Semenov (2013). *Panel b:* Depiction detailing magnetospheric accretion, the truncation radius is of only a few stellar radii. Figure from Hartmann et al. (2016).

2016). This is the 'magnetospheric accretion paradigm', and it is illustrated in Fig. 1.7. Each flow generates an accretion shock that hits the photosphere free-falling at $\sim$ 300 km/s with a temperature of $\sim 10^6$ K (Brickhouse et al., 2010). This shock cools down via X-ray emission which in turn heats the surrounding optically thin preshock material to $\sim 10^5$ K. This can be seen in the far UV over the Balmer continuum of the star (see Hartmann et al., 2016; Schneider et al., 2018). The

accretion funnels themselves are thought to be isothermal at lower temperatures of $\sim$ 7000 K (see e.g. Alencar et al., 2012). There is still a vague understanding on how outflows relate to the accretion processes. It has been proposed that accretion energy opens magnetic field lines which drive gas away through winds (Matt et al., 2012). Whether accretion is stable or if there are unstable accretion regimes that might lead to accretion bursts is still under study. Theoretical references for magnetospheric accretion can be found in: Muzerolle et al. (2001); Lima et al. (2010); Romanova et al. (2013).

In general, the accretion rate increases with stellar mass even from very low-mass stars (Manara et al., 2017), with younger objects showing larger accretion rates. These correlations are complex though (e.g. Manara et al., 2020). In addition, mass and age are very degenerate in any data set, as massive objects evolve much faster and hence are typically younger (see isochrones of Fig. 1.6).

Accretion on Herbig Ae/Be stars has been studied by a number of authors, for example: Muzerolle et al. (2004), Donehew and Brittain (2011), Mendigutía et al. (2011a), Cauley and Johns-Krull (2015), Fairlamb et al. (2015, 2017), Wichittanakom et al. (2020). As mentioned before, higher-mass PMS objects have radiative envelopes and hence normally exhibit negligible magnetic fields. Therefore, the magnetospheric accretion model probably cannot apply to them. The transition from magnetospheric accretion to the still unknown accretion mechanism for higher-mass PMS objects takes place within the mass range of the Herbig Ae/Be stars. Indeed, near-IR interferometric (e.g. Monnier et al., 2005), optical- and near-UV spectro-polarimetric (e.g.Vink et al., 2002; Ababakr et al., 2017) and spectro-photometric observations (e.g. Mendigutía et al., 2011a; Cauley and Johns-Krull, 2015; Fairlamb et al., 2015; Patel et al., 2017; Wichittanakom et al., 2020) have shown that the lower mass Herbig Ae stars show accretion signatures consistent with T Tauri stars, whereas Herbig Be stars appear to be inconsistent with magnetospheric accretion (see also Grady et al., 2010; Schöller et al., 2016). This is supported by the evidence of

Herbig Be stars being surrounded by denser and larger inner gaseous discs (Monnier et al., 2005 and Ilee et al., 2014, respectively).

The latest study of Wichittanakom et al. (2020) placed the change in accretion properties at $3.98^{+1.37}_{-0.94}$ $M_{\odot}$ (around B3 type), suggesting the 'boundary layer' accretion model for more massive objects. Boundary layer accretion was introduced in the case of young stars by Bertout et al. (1988) for T Tauri stars, and it has been proposed repeatedly (Blondel and Djie, 2006; Mendigutía et al., 2011a; Cauley and Johns-Krull, 2014, 2015; Kraus, 2015; Wichittanakom et al., 2020) as the main alternative to magnetospheric accretion in high-mass non-magnetic objects. It basically implies that the disc reaches directly to the star. However, it has never been properly observed, tested or modelled in Herbig Be stars or more massive objects (see Mendigutía, 2020 review). In principle, it should be a less energetic process requiring larger accretion rates to match the observations (see Sect. 1.5). In addition to accretion, there is observational evidence that points towards more differences between low- and high-mass PMS sources. For example, Herbig Be stars are more likely to be found in binaries than Herbig Ae stars (Baines et al., 2006) and more massive stars tend to appear in clusters (see Sect. 1.2.3) and rotate faster (which has an impact on the magnetic field, Emeriau and Mathis, 2015).

Regarding the aforementioned differences between low- and high-mass PMS sources, an important issue is that the scarcity of known high-mass PMS sources makes the statistics non-robust. Another important caveat in these studies is that they do not include the less evolved sources in high-mass PMS tracks (most Herbig Be stars considered are very close to the main sequence, see Chapter 2), which are obviously of paramount importance for understanding high-mass star formation. This is because more massive objects become optically visible later and evolve faster, thus implying that they are more likely to be found close to the main sequence. In addition, less evolved massive objects can have a later spectral type than A or B, and thus they could have been missed. This is why the main objective of this thesis is to

study the general properties of Herbig Ae/Be stars as a function of mass and to find new objects of the class in a homogeneous and unbiased fashion, trying to identify new young and massive PMS sources.

Very massive stars present other several particularities in their formation. During the PMS phase MYSOs may even ignite H-burning while accreting, thus blurring the concept of ZAMS for these objects. In addition, they generate enormous amounts of radiation pressure, which has an impact on the infalling material. Theoretically, this radiation pressure can halt the collapse eventually and prevent the formation of super-massive stars, although stars up to ~ 100 $M_{\odot}$ have been observed (e.g. Kashi and Soker, 2010, Crowther et al., 2016). As MYSOs are optically-invisible, heavily extincted, low in number and typically farther away (Mottram et al., 2011); it is challenging to study their super fast formation. In general, significant differences are expected in the MYSO formation process with respect to the lower-mass case, from the core phase to the ZAMS. These differences are probably caused by the high mass gravitational instabilities generated, episodic accretion events, dynamical interactions within clusters, and by the differences in the magnetic field. Hence, the picture presented in this general introduction may only partially apply to them, while the sketch of discrete phases of Fig. 1.5 is not entirely true, as these objects basically skip through the PMS phase. It is important to note that although very few in number (according to the IMF), MYSOs dominate the feedback of the ISM via their outflows, HII regions and eventual supernovae. See Lumsden et al. (2013) for a summary of the properties of these objects.

The pre-main sequence phase is the last stage of star formation. As the objects evolve through the PMS phase; the interactions within the disc, the accretion, and the feedback from the star (radiation and winds) end up clearing the disc (which perhaps has formed a few planets, see Sect. 1.4.3). This is represented in the panel *5* of Fig. 1.5 (also called Class III phase). The timescale on which this occurs depends on the mass of the star, as can be seen from the isochrones of Fig. 1.6, but it happens

typically between $\sim 10^6 - 10^7$ yr. Historically, T Tauri stars have been subdivided into classical T Tauri stars and weak-line T Tauri stars, depending whether they are in the Class II or the Class III respectively. At some point the star reaches the ZAMS, the disc gets mostly dispersed and a formed star starts its life with an orbiting planetary system around, which is composed of planets and other minor bodies, a picture not much different to our own Solar System.

1.4.3 Discs and planets

To complete the picture of star formation it is worth detailing the protoplanetary discs around PMS objects. These discs consist of approximately 99% gas and 1% dust (although this ratio is debated). Discs are rich, complex environments and they evolve from gas-rich protostellar discs to gas-depleted, dissipated discs in a few million years. These discs can be directly detected in the sub-millimiter and radio wavelengths with telescopes like the Atacama Large Millimeter/submillimeter Array (ALMA) or the forthcoming Square Kilometre Array (SKA). Examples of discs around PMS objects detected with ALMA are presented in Fig. 1.8. In addition, near-infrared scattered light traces the micron-sized grains that populate the optically thick surface of the discs. This scattered light can also evidence planets or sub-structure within protoplanetary discs (e.g. Canovas et al., 2017, Mendigutía et al., 2017, Avenhaus et al., 2018).

A glance of the internal composition of these discs can be seen in the *a* panel of Fig. 1.6. Mostly Keplerian in nature (except for some MYSOs), they have radial mass displacements driven mostly through viscosity and turbulence. The first thing to note is that because dust sublimates at a temperature of ~ 1500 K only gas reaches to the inner radii. However, dust is the most important component of the disc, as many important chemical processes happen on the surface of dust grains and they are the first step in the formation of planetesimals and planets through grain

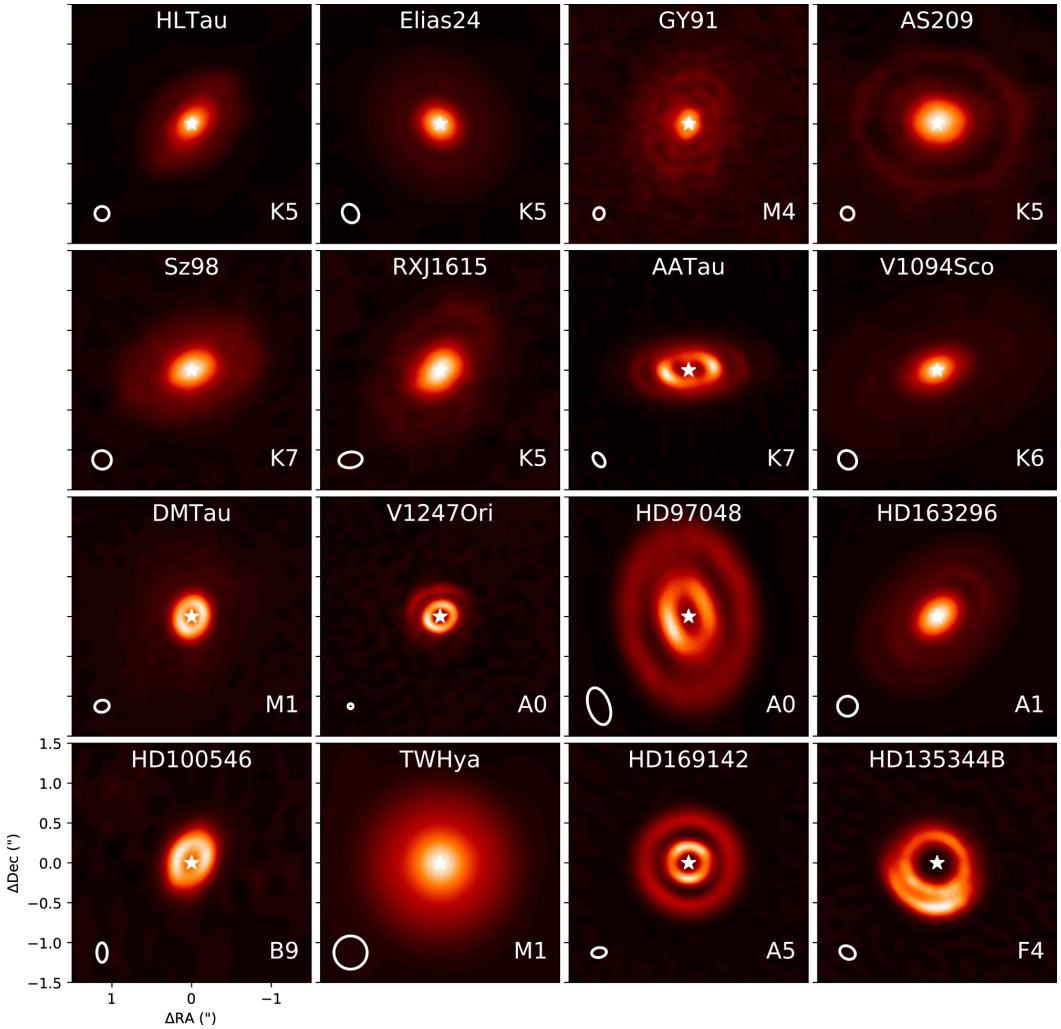


Fig. 1.8 Protoplanetary discs detected in continuum emission with ALMA around different pre-main sequence stars. The beam size (resolution) is indicated in the lower left corner. Figure from van der Marel et al. (2019).

growth processes. Hence, an important concept is the 'snowline', the radii where the different volatile elements ice-condensate on the surface of dust grains and hence are removed from the gas. Snowline locations depend primarily on temperature, although other factors might have an influence. See Andrews (2020) for a review for more details.

There is a large variety in the structure of PMS protoplanetary discs: spirals, rings, cavities, misaligned inner discs, etc. (Garufi et al., 2018; Andrews et al., 2018 and related papers). The causes of these structures are under debate and several

mechanisms have been proposed; for example dust traps (van der Marel et al., 2019), snow lines (Zhang et al., 2015) or gravitational instabilities (Dong et al., 2018; Booth and Ilee, 2020). However, the most accepted explanation is forming planets interacting with the disc (Dipierro et al., 2015; Liu et al., 2019). Some planets have been found within these discs in, for example, Quanz et al. (2015), Sallum et al. (2015), Haffert et al. (2019), and Pinte et al. (2019); although most have been refuted or questioned in later studies (see e.g. Mendigutía et al., 2018). Therefore, although most protoplanetary discs show sub-structures (Long et al., 2018), hardly any planet is detected in these discs. Note that detecting planets in such active environments needs complex techniques which are often open to interpretation (see e.g. Mendigutía et al., 2018 and references therein). However, there is a clear observational bias, as so far mostly long-lived, massive discs around low-mass stars have been observed. Manara et al. (2018) stated that discs as old as $1-10$ Myrs might not be massive enough to have active planet formation. As most planet detection methodologies trace the interaction of the disc with the forming planets, this implies that planets responsible for the observed sub-structures may have escaped searching campaigns so far. In addition the presence of planets is expected to be higher at high-mass stars (Kennedy and Kenyon, 2008; Pascucci et al., 2016; Panić and Min, 2017; Garufi et al., 2018).

All the previous discussion points towards the fact that what we are missing in these analyses is a homogeneous set of young Herbig Ae/Be stars. As explained in Sect. 1.4.2 to generate that set is one of the main motivations of this thesis.

1.5 Observables: Gaia and other telescopes

As explained in Sect. 1.4.3, protoplanetary discs are visible at sub-mm and radio wavelengths because they are irradiated by the central star. In fact, the disc is visible from the near-infrared, in addition to the stellar photosphere. This IR excess which

is re-radiated from the stellar emission can be seen in Fig. 1.5 in the different SEDs. It characterises the stage of stellar evolution: Class 0 objects are only visible in the IR, protostars might be optically visible but emit most of their energy as IR excess, PMS sources' SEDs peak in the optical but the disc is clearly visible from the near- or mid- IR. Class III objects only have small or negligible IR excesses.

The IR excess properties of Herbig Ae/Be stars have been classified into two groups differentiated by a flat or rising shape of the continuum (Meeus et al., 2001). It is believed that this difference has a geometric origin depending on the presence of flaring outer discs and puffed-up inner discs (Dullemond and Dominik, 2004a,b, 2005), and the presence of gaps in the disc (Maaskant et al., 2013; Honda et al., 2015). Indeed, near-IR scattered light images of HAeBe stars indicate large disc cavities which might be the base of the differences in IR SEDs. Whether this is circumstantial or an evolutionary effect is still under study.

Another characteristic I have not mentioned yet is the presence of spectral emission lines, one of Herbig's original conditions (Sect. 1.1). The material around the photosphere is excited and ionised, generating emission lines above the photosphere continuum (Kurosawa et al., 2006, 2011; Tambovtseva et al., 2014). Therefore, this emission mainly comes from the accretion shock, pre-shock, and accretion column (assuming magnetospheric accretion. In high-mass stars emission lines are observed which also arise from hot gas close to the star, but note that a boundary layer model for line emission is not yet available). At optical and IR wavelengths, the absorption lines might appear more filled than they should because of the underlying emission, this being called 'veiling'. However, it is in the UV where the excess from accretion is more evident, contributing to the continuum emission even orders of magnitudes over the photosphere. X-rays have been observed, although they do not seem strongly correlated with the accretion rates (Curran et al., 2011; Argiroffi et al., 2017).

Normally, the Hα line or other hydrogen lines are used to trace and characterise accretion. However, other lines can be used. For example, Cauley and Johns-Krull

(2014) found no infalling matter in HBe stars using He I λ10830. The correlation between accretion luminosity and emission line strength for different lines was studied in Mendigutía et al. (2011a, 2015), Donehew and Brittain (2011), and Fairlamb et al. (2017). However, the origin of this correlation is yet unclear. The line profile also carries information, and single-peaked profiles normally correspond to face-on discs whereas double-peaked profiles correspond to edge-on discs (see Chapter 2). P-Cygni profiles normally trace stellar winds or accretion. Similar emission for stars with and without magnetic fields has been detected, which means that some HAeBes have low magnetic fields but show strong accretion. Emission lines might be misleading though, as they might vary over short time-scales and be subjected to other stellar circumstances in addition to accretion (see Mendigutía et al., 2011b).

The last main observational characteristic of PMS objects is variability. Herbig Ae/Be stars are known to present irregular photometric variations, with a typical timescale from days to weeks (Oudmaijer et al., 2001; Eiroa et al., 2002) and of the order of one magnitude in the optical. This variability is typically caused by variable extinction ('dippers', detected in $\sim$ 33% of the objects, Bouvier et al., 1999, Bodman et al., 2017), which mainly arises from the rotating circumstellar discs. An extreme case of large photometric variations caused by discs is observed in UX Ori type stars (UXORs), with amplitudes up to $2-3$ mag. Other sources of variability are the effect of rotation on cold photospheric spots and pulsations due to internal instability (Marconi and Palla, 1998 and Zwintz et al., 2014, respectively). We add to this the episodic accretion events mentioned in Sect. 1.4 (EXor or FU Ori type sources, detected in $\sim$ 14% of the PMS objects, Cody et al., 2017). Infrared photometric variability related to disc structure variations is not always correlated with the optical variability (Eiroa et al., 2002) which implies that different mechanisms regarding both the disc structure and accretion underlie the final observed variability. Spectroscopic variability is also present in Herbig Ae/Be stars (Mendigutía et al., 2011b). PMS

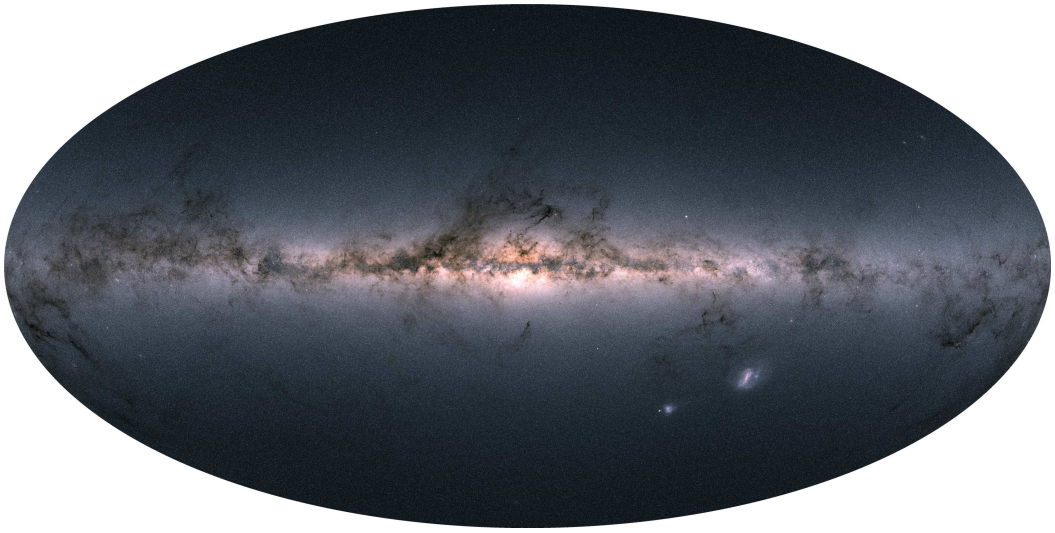


Fig. 1.9 Sky map of the total flux measured by Gaia for the almost 1.7 billion sources in Gaia DR2. The colour is obtained from the blue (G_{BP}) and red (G_{RP}) Gaia passbands. Note that this is not a picture, each point is an individual detection. Figure from Gaia Collaboration et al. (2018b).

sources have a particular footprint in variability curves (because of disc-variability but also because of asteroseismological oscillations; see Zwintz et al., 2015, 2011) and different surveys have been used to sample their variability light-curves such as CoRot, Kepler - K2, and the Transiting Exoplanet Survey Satellite (TESS).

As mentioned in Sects. 1.4.2 and 1.4.3, the purpose of this thesis is to perform a general study of the properties of Herbig Ae/Be stars and to find new objects of the class that allow to correct for the biases of the set of known objects. For this goal, I used the data from the Gaia mission as a cornerstone of the thesis.

Gaia is an astrometric all-sky mission, whose main goal is to homogeneously determine parallaxes, proper motions, radial velocities, and positions for about one billion celestial sources. The Gaia spacecraft is currently orbiting at the L2 point in a Lissajous orbit. Launched in 2013 for a nominal mission of five years, it carries two telescopes which register stars passing through a single focal plane composed of multiple CCDs. Gaia scans this way the whole sky while spinning, and it is totally independent of previous catalogues and surveys. After 22 months of observations its Data Release 2 (DR2, Gaia Collaboration et al., 2018b, 2016; map shown in

Fig. 1.9) provided a five dimensional astrometric solution and low resolution optical photometry in three bands for up to $G \lesssim 21$ mag (white G band, described in Evans et al., 2018) to over 1.3 billion objects (Lindegren et al., 2018). The accuracy of the parallax is of 22 micro-arcsecond at $G = 15$ mag. In addition, spectroscopy from Ca triplet provides radial velocities for 7.2 million objects on the brighter end.

Looking back, the first reliable stellar parallax was measured in 1838 by Friedrich W. Bessel (1784-1846) for 61 Cygni (Bessel, 1838). As discussed in Sect. 1.1, Russell produced a state-of-the-art catalogue of 55 parallax fields in 1906. He estimated that the error of a single exposure was ~ 0.05 arcsecond (Devorkin, 2000), which is three to four orders of magnitude above Gaia DR2 typical uncertainties. Hipparcos, the predecessor of Gaia, measured $\sim 120,000$ parallaxes down to magnitude $G \sim 11$. By the time this PhD thesis was written, 1.33 billion sources had a Gaia parallax.

1.6 Machine learning

In this thesis I have used different machine learning (ML) techniques to visualise and work with large datasets. ML is the branch of artificial intelligence which, by means of statistical techniques, designs programs with the ability to progressively improve at performing a specific task given some input data (i.e. to *learn*), without them being explicitly programmed for that particular problem. ML algorithms discover patterns in the data, achieve a good prediction performance in novel data and may also aim at *understanding*, if such a thing could be properly defined.

There are two main branches within ML, supervised learning and unsupervised learning. In supervised learning the input is a list of objects with measured properties (features) and labels (ground truth). Supervised algorithms are mostly used for regression or classification problems. The key concept of supervised learning is that prior knowledge is required, from which the algorithm learns. Some of the most famous supervised algorithms are: artificial neural networks (ANN), support

vector machines (Cortes and Vapnik, 1995) and random forests (Ho, 1995). In unsupervised learning the input is a list of objects with measured properties (no labels). Unsupervised algorithms are mostly designed to detect clusters, complex relations or outliers, or to reduce the dimensionality of a dataset. Prior knowledge is not required. Examples of popular clustering algorithms are the 'k-nearest neighbours' or the 'density-based spatial clustering of applications with noise' (DBSCAN, Ester et al., 1996). Examples of popular dimensionality reduction algorithms are 'principal component analysis' (PCA) or 't-distributed stochastic neighbor embedding' (t-SNE, van der Maaten and Hinton, 2008).

1.6.1 Loss and hypothesis functions

The output of a ML algorithm is called the hypothesis function, which is expressed by means of the internal parameters of the algorithm ($\theta_0, \theta_1, \theta_2, ..., \theta_m$). All machine learning algorithms are based on two main concepts. A loss or cost function and an algorithm to minimise (optimise) such function. The loss function measures of how good an algorithm does in terms of being able to predict the expected outcome. In other words, it measures the accuracy of the hypothesis function. A simple example of a hypothesis function is:

$$h_\theta(x) = \theta_0 + \theta_1 x. \tag{1.3}$$

If a set parameters (θ_0, θ_1) is chosen it is possible to make a prediction $h_\theta(x)$ for any value of the feature x. Selecting the best parameters θ_0 and θ_1 so $h_\theta(x)$ gives the best prediction according to some prior knowledge constitutes a *linear regression* problem, one of the simplest forms of ML algorithms. Say we have n known data pairs $[(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)]$ (i.e. some ground truth y_i for every object i with the feature value x_i). A perfect hypothesis would be the one with $h_\theta(x_i) = y_i$ for every x_i, y_i pair. Therefore, the way forward is to minimise the distance between the

$h_\theta(x_i)$ and the y_i values on the n pairs (x_i, y_i). That is, it is necessary to minimise the loss function J:

$$J(\theta_0, \theta_1) = \sum_{i=1}^{n} (h_\theta(x_i) - y_i)^2. \tag{1.4}$$

In order to minimise J we need of an optimisation algorithm that starts with some random θ_0, θ_1 and iteratively changes them to reduce the value of $J(\theta_0, \theta_1)$ until, hopefully, a minimum is reached. The parameters at the minimum define our trained algorithm, which makes the best hypothesis given the chosen ground truth. The (x_i, y_i) pairs receive the name of 'training set'. Loss function optimisation is exemplified in Fig. 1.10.

Other ML algorithms are not much different in their principles to this example, although they might be much more complex in their details. Hence, they can adapt to datasets which are beyond the scope of any polynomial fit. In particular, ML algorithms are characterised by the so-called 'hyper-parameters', which are the values the user may or must choose that define the fine details of the structure of the algorithm and constrain its behaviour. Thus, the selection of the hyper-parameters is of the utmost importance in any ML application. In the case of unsupervised learning, where there is no prior knowledge to aid the algorithm, the hyper-parameters totally dominate the output.

1.6.2 Overfitting and underfitting

Two paramount concepts in any supervised ML application are 'overfitting' and 'underfitting'. With the right hyper-parameters, most ML algorithms can get as complex as the user demands. However, in order to achieve optimal performance the complexity of the algorithm needs to match the one demanded by the specific problem. The more complex the algorithm is, the more it adjusts to the training data. If the algorithm is overly complex it gets too sensitive to the training data

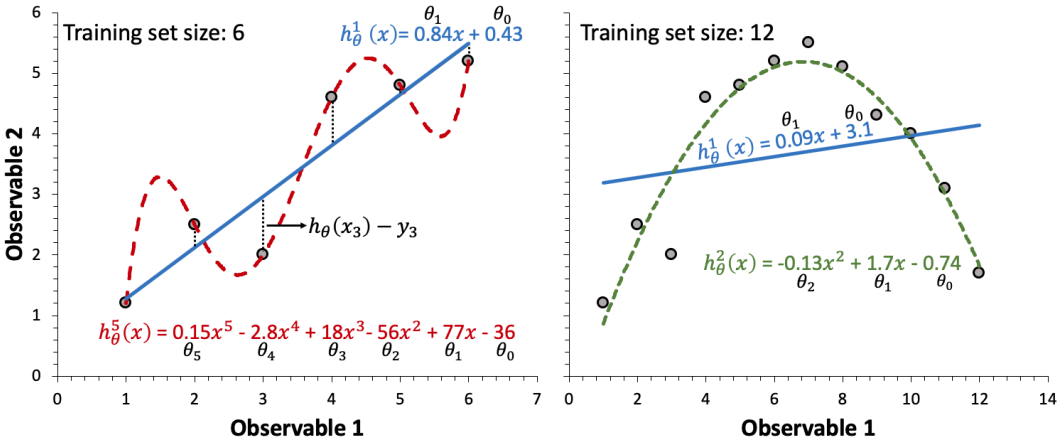


Fig. 1.10 *Left:* Two different hypothesis functions, one linear (h_θ^1) and one fifth degree polynomial (h_θ^5), result from training with the same training set of six data pairs. The training was performed by finding the parameters θ that minimised the total distance between the hypothesis predictions and the training labels (i.e. $h_\theta(x_i) - y_i$, see Eq. 1.4). After training, the more complex algorithm (h_θ^5) has achieved a much better fit to the training data, but it is unlikely that the resulting hypothesis function satisfactorily predicts novel data. This is a case of overfitting. *Right:* Six more data pairs were added to the training set at left, increasing the complexity of the problem. Now h_θ^1 is clearly underfitting the problem, whereas a more complex hypothesis like h_θ^2 is more capable of generating adequate predictions.

and no longer generalises properly to novel data outside the training set. This is called overfitting or high variance. On the other extreme, if a too simple algorithm is applied to a problem of greater complexity it will not achieve an accurate fit to the training set and, therefore, it will also fail when generalising. This is called high bias or underfitting. The performance of every supervised ML algorithm is mostly determined by the balance of these two concepts, which are exemplified in Fig. 1.10.

The loss function is defined in terms of the training set (see Eq. 1.4) so it does not inform about overfitting. This is why 'cross-validation' and 'test sets' are used. These sets are random subsamples of the training set that are not used for training, but are kept separate to evaluate the performance of the algorithm. Cross-validations sets are used at intermediate steps to estimate the best configuration of hyper-parameters or to assess when to 'early-stop' the optimisation of the loss function, whereas the

test set (typically $10-30\%$ of the original training set) is used only to evaluate the performance of the final trained algorithm.

There are different ways to approach underfitting, in addition to increasing the complexity of the algorithm, which is not always possible, desirable or convenient. It is often more useful to get more training data or to use additional features. If an algorithm suffers from overfitting it is possible to reduce its complexity or to use fewer features. Another approach is to use the hyper-parameters of the algorithm. For example, some common hyper-parameters used in ANNs to reduce overfitting are 'regularisation' or 'dropout'.

It is important to note that it has been widely tested that most supervised algorithms achieve similar performances when exposed to the same dataset (e.g. Banko and Brill, 2001). Therefore, it is not about the algorithm used, it is about the quality and quantity of the data. This also have been observed in multiple astrophysical applications (e.g. Pérez-Ortiz et al., 2017, Pashchenko et al., 2018 or Marton et al., 2019). Therefore, feature extraction and data selection constitute a necessary and important step in any ML application. In contrast, performance on unsupervised algorithms can vary heavily from algorithm to algorithm, and as happens with the hyper-parameters, the selection of the algorithm determines the output and it should be studied with caution for every particular case. Furthermore, there is a strong inverse correlation between algorithm interpretability and algorithm complexity. Linear regressions on one extreme are easy to interpret but cannot construct complex hypothesis functions. On the other end, ANNs and support vector machines can find highly complex non-linear relations in the data, but little or no information can be extracted about how they got to the output they provide.

Most of ML algorithms have been known for decades. However, only now that the Big Data era has arrived their use has generalised to almost all disciplines. This is because ML algorithms are very data-demanding. Therefore, large datasets as the ones produced by Gaia are excellent for exploitation with statistical learning

techniques (as done in e.g. Marton et al., 2019, Castro-Ginard et al., 2019, 2018; Galli et al., 2019; Cánovas et al., 2019; Kounkel and Covey, 2019 or Zari et al., 2019). There are some caveats; ML algorithms are not suitable for all problems, their outputs are hard to interpret and it is very easy to introduce unwanted or uncontrolled biases. In addition, they are computationally expensive and they struggle to deal with missing data and uncertainties. However, there are several problems and patterns that, either because of being too complex or too subtle, are unreachable for the human brain and can only be explored using these techniques. General references can be found in James et al. (2013) and Ng (2017). See also Baron (2019) for a general description of these techniques into astronomy, including the not mentioned but broadly used convolutional neural networks and other 'deep learning' algorithms. A more detailed description of the ML algorithms used in this dissertation is presented in Appendix B.

1.7 Thesis outline

Almost all of the known Herbig Ae/Be stars have parallaxes in Gaia DR2, which allows luminosities to be derived and 252 HAeBes to be placed in the HR diagram. This enables me to perform a homogeneous study of the general properties of these objects. This is done in Chapter 2 of this thesis.

Using the Gaia characterisation of known Herbig Ae/Be stars of Chapter 2 as starting point, I identify new Herbig Ae/Be candidates and create an unbiased and well defined catalogue of new objects of the class. This is done by means of machine learning techniques in Chapter 3. Classical techniques are not efficient for identifying new Herbig Ae/Be stars mainly because of their similarity with classical Be stars, with which they share many characteristics. By focusing on disentangling these two types of objects, our algorithm has also identified new classical Be stars.

Chapter 4 is dedicated to a study of the properties of the new PMS and classical Be candidates from the catalogues constructed in Chapter 3. To this I add a detailed description and analysis of the independent spectroscopic observations I performed for a subsample of the catalogues.

Finally, the general conclusions of the thesis are presented in Chapter 5, to which I add future prospects for research.

Chapter 2

Analysis of known Herbig Ae/Be stars

The geometry of innocence flesh on the bone
Tombstone Blues, 1965 – Bob Dylan

In this chapter I use Gaia Data Release 2 (DR2) to place 252 known Herbig Ae/Be stars in the Hertzsprung-Russell (HR) diagram and investigate their characteristics and properties. This chapter is organised as follows: In Sect. 2.1, I describe the data acquisition of not only the parallaxes, but also optical and IR photometry, effective temperatures, extinction values, Hα emission line information, and binarity. In Sect. 2.2, I derive the stellar luminosities and place the objects in a HR diagram, while I also present a method to derive a statistical assessment of the objects' variability in Gaia's database. In addition, I homogeneously derive masses and ages for all the sources, together with near-IR (NIR) and mid-IR (MIR) excesses. In Sect. 2.3, I carry out an analysis of the data and present various correlations and interdependencies, which I discuss in the context of intermediate-mass star formation in Sect. 2.4.

2.1 Data acquisition

2.1.1 Construction of the sample

I have gathered the majority of Herbig Ae/Be stars known and proposed to date from different works (272, see Chen et al., 2016 for a compilation of most of them). Chen et al. (2016) based their sample mostly on the work of Zhang et al. (2006) which in turn is based on the work of The et al. (1994) and Vieira et al. (2003). In addition, I included a few HAeBes from Alecian et al. (2013), Baines et al. (2006), Carmona et al. (2010), Fairlamb et al. (2015), Hernández et al. (2005), Manoj et al. (2006), and Sartori et al. (2010) that are not present in the aforementioned papers.

Although Herbig Ae/Be stars have long been considered, by definition, to be type A or B, there should be some flexibility in this constraint as the physical boundary between Herbig Ae stars and Intermediate Mass T Tauris is relatively poorly studied. This is because spectral types of T Tauri stars are typically K-M with some G-type objects, while Herbig Ae/Be stars are, quite unsurprisingly, limited to A and B spectral type. Therefore, PMS stars of intermediate spectral types have often been largely understudied. I therefore keep objects with F-type classification in Chen et al. (2016) in the sample. Similarly, no upper limit in mass was imposed, leaving the separation between MYSOs and HAeBes to the optical brightness of the sources [1].

Subsequently, I cross-matched the sources with Gaia DR2. Detections were considered to be matched with the catalogue when their coordinates agreed to within 0.5 arcsecond. If more than one match was found, I took the closest one. If no match was found within 0.5 arcsecond, successive cross-matches with larger apertures were performed up to 2 arcsecond. In these latter cases an individual inspection of the cross-match was applied. Finally, a comparison between the Johnson V band

[1]The MYSOs are typically IR-bright and optically faint (Lumsden et al., 2013). However, a number of optically visible objects are known to have passed all selection criteria such as the early-type objects PDS 27 and PDS 37 that are also classified as Herbig Be stars (Ababakr et al., 2015).

magnitudes and the Gaia filters was made for each source in order to discard possible incorrect matches. This provided me with parallaxes for 254 HAeBes.

As Lindegren et al. (2018) point out, not all Gaia DR2 parallaxes are of the same quality, and some values - despite their sometimes very small error bars - appear erroneous. I included the following constraint in astrometric quality following the indications in Appendix C of Lindegren et al. (2018) and what was applied in Gaia Collaboration et al. (2018a). This constraint will remove objects with spurious parallaxes from the sample:

$$u < 1.2 \times \max(1, e^{-0.2(G-19.5)}), \tag{2.1}$$

where G is the Gaia G band and u is the 'unit weight error', defined as the square root of the ratio of the 'astrometric_chi2_al' and ('astrometric_n_good_obs_al' − 5) columns (Lindegren et al. 2018, their Equation C.2). Of the 254 HAeBes, 228 satisfy this condition.

Some objects are found to be very close to this condition; PDS 144S, PV Cep and V892 Tau, and as I show below, they would appear significantly below the MS in the HR-diagram. Given that the Lindegren et al. (2018) condition is presented as a guideline rather than a rule by the Gaia astrometry team, I decided to treat these three objects as if they satisfy Eq. 2.1 as well.

I refer to the set of astrometrically well-behaved sources as the high-quality sample and to those that do not satisfy Eq. 2.1 as the low-quality sample. Two sources could not be placed in the HR diagram due to a lack of appropriate parameters (Sect. 2.1.2). In addition, I move five more sources to the low-quality sample in Sect. 2.2.1 for different reasons. Summarising, there are 218 objects ($228-3-2-5$) in the final high-quality sample and 34 in the low-quality one. Information about the objects in different samples is presented in separate tables in Appendix A (Tables A.1, A.2, and A.3 for the high-quality sample and Tables A.4, A.5, and A.6 for the

low-quality sample). The high-quality sample will be the one taken into account in further considerations unless otherwise specified.

Distances are not obtained by straightforwardly inverting the parallax. The conversion of one parameter to the other one is not strictly trivial because of the non-linearity of the inverse function (see for example Bailer-Jones, 2015). In addition, distances can be strongly affected by the selection of a prior for the conversion. In the case of Gaia DR2, Bailer-Jones et al. (2018) proposed distance values using a weak distance prior that follows a Galactic model. Their distances begin to differ from the distances obtained through simple inversion for sources with large errors, $\sigma_{\varpi}/\varpi \gtrsim 0.5$. Therefore, in the initial (high- and low-quality) sample only a small subset of 12 Herbig Ae/Be stars suffers substantially from these effects. Following the indications in Luri et al. (2018) on how to treat the Gaia parallaxes, I decided to apply a simpler exponentially decreasing prior to estimate distances. For completeness, I should note that the parallaxes provided by Gaia DR2 have a regional and not Gaussian systematic error as large as 0.1 mas and a global zero point error of about -0.029 mas which are not included in the Gaussian random errors provided in the Gaia archive (see Arenou et al., 2018 and Lindegren et al., 2018). The uncertainty in the parallaxes is therefore slightly underestimated. The final errors in the high-quality sample range from 0.016 to 0.37 mas.

Herbig Ae/Be stars have been historically confused with classical Be stars, with which they share many characteristics (Grundstrom and Gies, 2006; Rivinius et al., 2013; Klement et al., 2017). Indeed, the nature of some of the objects in the sample is still under debate. An interesting example in this respect is HD 76534, a B2Ve object that appears in listings of Be stars (e.g. Oudmaijer and Drew, 1997) and Herbig Be stars alike (Fairlamb et al., 2015). The latest dedicated study puts the object in the Herbig Be category (Patel et al., 2017). To assess the effect of ambiguous classifications in this study, next to the full sample, I also consider the subset of Herbig Ae/Be stars in Table 1 of The et al. (1994). This catalogue contains all

historically known, and best studied, Herbig Ae/Be stars. Ninety-eight of the 254 initial sources with parallaxes are present in this table (their best candidates). This The et al. (1994) subset is indicated in Table A.3 and Table A.6 for the high- and low-quality samples, respectively.

2.1.2 Atmospheric parameters, photometry, and extinction values

I obtained atmospheric parameters and photometric and extinction values for all the sources from the literature. These were mainly Alecian et al. (2013), Carmona et al. (2010), Chen et al. (2016), Fairlamb et al. (2015), Hernández et al. (2004), Hernández et al. (2005), Manoj et al. (2006), Montesinos et al. (2009), Mendigutía et al. (2012), Sartori et al. (2010), Vieira et al. (2003), and the APASS Data Release 9 (Henden et al., 2018). Whenever the effective temperature (T_{eff}) was not available it was derived from the spectral type with the effective temperature calibration tables of Gray and Corbally (2009). An uncertainty of 1 sub-spectral type was assigned in all cases. When not listed in the literature, A_V values were derived from the observed photometry and using the intrinsic colours of Pecaut and Mamajek (2013). An $R_V = 3.1$ was used in all cases in which A_V was derived; although other studies like Hernández et al. (2004) or Manoj et al. (2006) have suggested that a larger value of for example $R_V = 5$ could be more appropriate for HAeBes where local extinction dominates the total extinction. This is a topic for future investigations using diffuse interstellar bands (as done by e.g. Oudmaijer et al., 1997). The relevant data of each source is presented in Tables A.1 and A.4 for the high- and low-quality samples, respectively.

HAeBes usually show photometric variability. Thus, for objects with multi-epoch photometry available, I selected the brightest set to determine the extinction towards the objects and thus their intrinsic brightnesses. As I also show below, the variability

is often caused by irregular extinction; using those data with minimum extinction introduces the smallest errors in the determination of the stellar parameters. For this reason, I only used simultaneous photometry when deriving A_V values. All the photometric values were corrected for extinction using the reddening law of Cardelli et al. (1989).

Two sources, V883 Ori and GSC 1829-0331, do not have enough simultaneous photometry available to derive extinctions for them and therefore they were excluded from the sample. The total number of Herbig Ae/Be stars that can be placed in the HR diagram and for which I can derive stellar luminosities, masses, ages, IR excesses, and variabilities in Sect. 2.2 is therefore reduced to 252 objects.

2.1.3 Infrared photometry

All the sources were cross-matched with the Two Micron All-Sky Survey (2MASS, see Skrutskie et al., 2006) and with the Wide-Field Infrared Survey Explorer all-sky catalogue (hereafter WISE, see Cutri et al., 2013). Both of these surveys contain hundreds of millions of stars, guaranteeing a large overlap with Gaia. I used a three-arcsecond aperture for the cross-match. The few sources that did not lie within that three-arcsecond threshold were studied individually and, if present, their IR photometry was included. This provides values and uncertainties for the J, H, and K_s bands (1.24, 1.66 and 2.16 μm, respectively) and for the W1, W2, W3, and W4 bands (3.4, 4.6, 12 and 22 μm, respectively) for most of the HAeBes. I note that for some sources some of the bands may be missing or may just be upper limits. I double-checked all IR matches with the dereddened optical photometry and found no inconsistencies.

2.1.4 Hα equivalent width and emission line profile

I collected all the Hα equivalent widths (EW) I could find in the literature for the Herbig Ae/Be stars. Not only the intensity of the line but also the shape contains very useful information. Therefore, when possible, information about the shape of the Hα line was included. I have classified the Hα line profile as single-peaked (s), double-peaked (d) and showing a P-Cygni profile (P), both regular or inverse. Equivalent width and line shape information are presented in Tables A.2 and A.5 for the high- and low-quality samples, respectively. Many Herbig Ae/Be stars are quite variable in their Hα emission and their EWs may significantly change on short timescales (e.g. Costigan et al., 2014). This is also the case for the line shape, although spot checks on objects that have more than one Hα observation listed in the literature appear to indicate that there are not many changes in line profile classification (see also e.g. Aarnio et al., 2017), although changes between single-peaked and double-peaked profiles in a given star are also observed (Mendigutía et al., 2011b). I do note the additional complication that emission line shapes are often difficult to unambiguously classify.

Regarding the Hα EWs compiled, I note that the main references (Fairlamb et al., 2015 and Mendigutía et al., 2011b) provide the non-photospheric contribution of the EW, while most other authors state the observed EW, which includes the photospheric contribution. This photospheric absorption peaks for A0-A1 type objects, with EW values of $\sim +10$ Å (See e.g. Fig. 7 of Joner and Hintz, 2015) but is only $\sim +2$ Å for B0 objects. I used the Joner and Hintz (2015) results to correct those EWs that were not corrected for absorption.

I have Hα EWs for 218 of the 252 HAeBes and line profiles for 197 of these: 31% are single-peaked, 52% are double-peaked, and 17% are P-Cygni (of which the vast majority are of regular P-Cygni type). This is in agreement with Finkenzeller and Mundt (1984) who found that out of 57 HAeBes, 25% were single-peaked,

50% showed double-peaked Hα profiles, and 20% presented a P-Cygni profile (both regular and inverse). The main references for the EW values are Baines et al. (2006), Fairlamb et al. (2017), Hernández et al. (2004), Mendigutía et al. (2011b), and Wheelwright et al. (2010). The main references for the line profiles are van den Ancker et al. (1998), Baines et al. (2006), Mendigutía et al. (2011b), Vieira et al. (2003), and Wheelwright et al. (2010). The rest of the references can be found in Tables A.2 and A.5.

2.1.5 Binarity

More than half of the Herbig Ae/Be stars are known to be in binary systems (Duchêne, 2015). The true number is likely much larger, as there have been a small number of targeted surveys for binarity of HAeBe stars, the largest are Wheelwright et al. (2010) and Baines et al. (2006) who performed spectroastrometry of 45 HAeBes and 31 HAeBes, respectively, probing companions in the ~ 0.1 - 2 arcsecond range, and Leinert et al. (1997) who performed speckle interferometry of 31 objects, sampling separations of order 0.1 arcsecond. Eighty-one of the 252 HAeBes ($\sim 32\%$) of my set are catalogued as binary systems, a fraction that is in agreement with the Duchêne (2015) findings if the large number of faint Herbig Ae/Be stars which have never been studied for binarity is taken into account. The binary status of each HAeBe is presented in Tables A.1 and A.4 for the high- and low-quality samples, respectively. Main references were Baines et al. (2006), Leinert et al. (1997), and Wheelwright et al. (2010); I refer to Table A.1 and A.4 for a complete list.

Baines et al. (2006) found a typical wide (few hundred astronomical units) separation in the binary systems. Wheelwright et al. (2010) detected no binaries closer than 30 au and established a range of $\sim 40 - 4000$ au in their data.

2.2 Derived quantities

2.2.1 Luminosity and Hertzsprung-Russell diagram

Using the parallaxes, atmospheric parameters, and extinction values, I derived the luminosity for the 252 HAeBes with parallaxes employing a similar method to Fairlamb et al. (2015), which is similar to that of Montesinos et al. (2009) and van den Ancker et al. (1997). In short, it first consists of using values of T_{eff} and surface gravity (log(g)) to select an atmosphere model from Castelli and Kurucz (2003) (referred to as CK-models hereafter) for each star to be used for its intrinsic spectral energy distribution (SED). Solar metallicity CK-models were used in all cases but for BF Ori, RR Tau, SV Cep, XY Per and WW Vul for which the metallicities are known not to be solar from the spectroscopic work of Montesinos et al. (2009). When possible, the log(g) values were estimated from the luminosity class; otherwise they were taken as 4.00 (typical values range from 3.5 to 4.5). Uncertainties in log(g) and metallicity can be neglected in this study as their effect on the model SED and derived quantities is negligible.

I then scaled the model to the dereddened photometric Johnson V band. The energy distribution is then integrated over frequency to get the total flux. The final luminosities, presented in Tables A.1 and A.4 for the high-quality and low-quality samples, respectively, are then obtained by means of the total flux and the parallax. All sources of uncertainty were taken into account at this step including using different CK models for the different temperatures within the $T_{\rm eff}$ uncertainty range.

The 223 Herbig Ae/Be stars satisfying the Eq. 2.1 constraint are plotted in the resulting HR diagram in Fig. 2.1. This number is an increase of more than a factor of ten compared to the previous, Hipparcos-based study by van den Ancker et al. (1998). Pre-main sequence evolutionary tracks from Bressan et al. (2012) are also

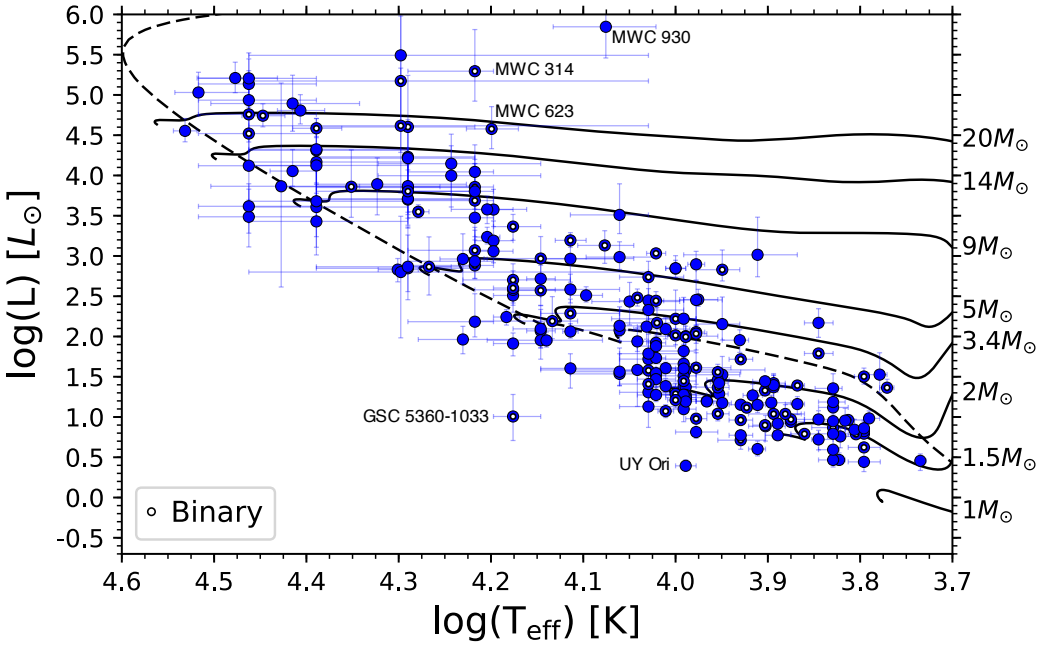


Fig. 2.1 223 Herbig Ae/Be stars in the HR diagram satisfying Eq. 2.1 constraint. In most cases vertical error bars are dominated by parallax uncertainties. Sources with a white dot have been classified as binaries. Explicitly named sources are discussed in the text in more detail. The mass of each PMS track (Bressan et al., 2012) is indicated on the righthand side. An isochrone (Bressan et al., 2012; Marigo et al., 2017) of 2.5 Myr is also shown for reference as a dashed line.

plotted in Fig. 2.1 in addition to a 2.5Myr isochrone (Bressan et al., 2012, Marigo et al., 2017), all of them with solar metallicities ($Z = 0.01$ and $Y = 0.267$).

Before I analyse this sample, I also plot the HR diagram for all 252 objects with parallaxes (the high- and low-quality samples together, hence including those which failed the Lindegren quality selection criteria) in Fig. 2.2 in the top panel. Many of these astrometrically badly behaved sources are located in unphysical positions, significantly below the MS, validating the approach of removing those from these analyses.

Returning to the HR diagram in Fig. 2.1, there are still several outliers that do not seem to be PMS objects. GSC 5360-1033 and UY Ori appear way below the MS, just like the lower-quality objects that were removed earlier. However, the Gaia DR2

data of these two objects appear to be of good quality. Regarding GSC 5360-1033, the situation is unclear, an ambiguous spectral type or photometry for this object or an incorrect estimation of the extinction may be the reason for the unexpected location of the object. For UY Ori, Fairlamb et al. (2015) assigned a spectral type of B9, but the photometry listed in SIMBAD indicates a large variability. Pending more certainty, I decided to move these two objects from the high- to the low-quality sample.

MWC 314, MWC 623, and MWC 930 on the other hand appear quite luminous and very much to the right of the MS (see Fig. 2.1), something very unusual for high-mass PMS objects. An individual inspection reveals that these objects are more likely to be evolved giants and they appear in the literature as such (e.g. for MWC 314: Carmona et al., 2010, for MWC 623: Lee et al., 2016, for MWC 930: Miroshnichenko et al., 2014). Deciding on the nature of the various Herbig Ae/Be stars in the master sample is beyond the scope of this chapter, which is essentially a statistical study (see Chapter 3 for further discussion on this topic). However, these objects occupy a special place in the HR diagram that is consistent with both a pre- and a post-MS nature, while there is much information regarding these objects supporting their post-MS nature. I therefore decided to err on the side of caution and exclude these from further analysis as well.

The final HR diagram without these $2+3$ problematic objects is presented in Fig. 2.2 in the bottom panel. In addition, in this graph, I highlight the sample of The et al. (1994) bonafide HAeBes in red. This final high-quality sample of 218 objects is the one used in the following plots and studies. The information concerning the 34 discarded objects in the low-quality sample can be found in Tables A.4, A.5 and A.6.

In this last high-quality HR diagram I see that there are many more low-mass HAeBes than high-mass HAeBes (69% of the sources are below 4 $M_{\odot}$). This is most likely because of the initial mass function (IMF). This trend of more objects for lower masses discontinues below ~ 2 $M_{\odot}$. This is roughly the mass corresponding

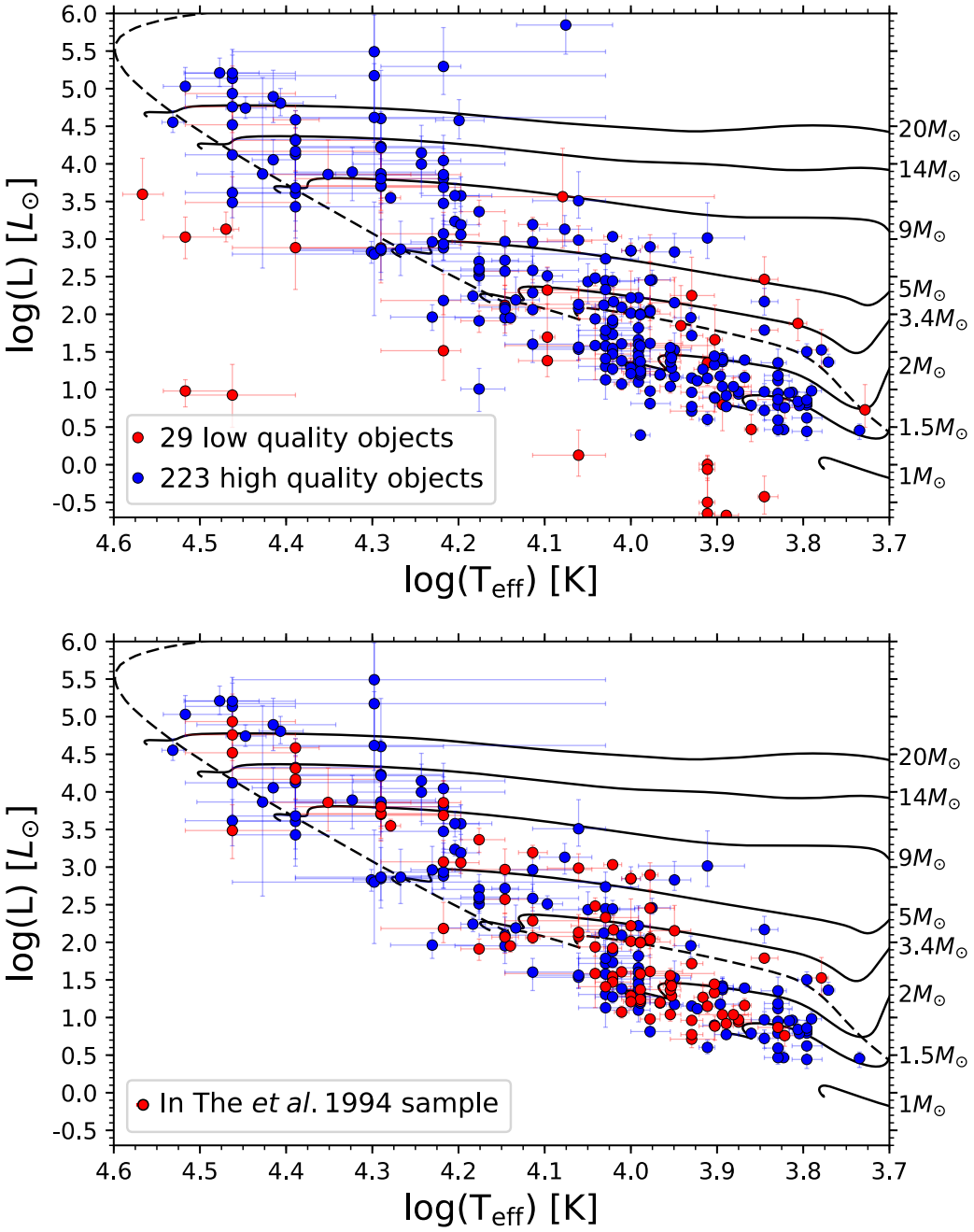


Fig. 2.2 *Top:* 223 high-quality and 29 low-quality Herbig Ae/Be stars in the HR diagram after the cut in astrometric quality described in Eq. 2.1. *Bottom:* 218 Herbig Ae/Be stars in the final high-quality sample after removing the five problematic objects described in Sect. 2.2.1. Those objects present in Table 1 of The et al. (1994) are shown in red. The mass of each PMS track (Bressan et al., 2012) is indicated on the right-hand side. An isochrone (Bressan et al., 2012; Marigo et al., 2017) of 2.5 Myr is also shown for reference as a dashed line.

to the boundary between MS A- and F-type stars, and thus the traditional lower-mass boundary at which the Herbig Ae/Be stars were originally selected.

For lower masses, the sources show a greater spread in temperature, occupying larger parts of the PMS tracks, while, instead, the high-mass objects tend to be predominately located close to the zero-age main sequence (ZAMS). This is likely because the higher the mass, the faster the PMS evolution. This fast evolution could explain why high-mass objects at low temperatures (and thus low surface gravities) are barely present in Fig. 2.1 or the sample. I encounter more examples below where high- and low-mass objects have different properties.

2.2.2 Mass and age

The masses and ages of the Herbig Ae/Be stars were estimated using isochrones. I used 100 PMS PARSEC (PAdova and TRieste Stellar Evolution Code) isochrones with solar metallicity (Bressan et al., 2012; Marigo et al., 2017) from 0.01 to 20 Myr. To each Herbig Ae/Be star I assigned the closest two isochrone points in the HR diagram; the solar metallicity isochrones did not match seven sources from the high-quality sample in the HR diagram and isochrones with lower metallicities were used in those cases. As each point is associated with a mass (M) and an age, for each HAeBe I computed an average of those values weighted by the distance to the points. The result is an estimate of age and mass for 236/252 HAeBes. These values are presented in Tables A.2 and A.5 for the high- and low-quality samples, respectively. Uncertainties were derived from the error bars in the HR diagram (Fig. 2.1 and Fig. 2.2) keeping a minimum error of 5%. Many of the derived masses and ages were compared with those of Alecian et al. (2013) and Reiter et al. (2018). I found that the determinations presented here of these parameters are consistent with the results of the previous authors.

2.2.3 Infrared excesses

An example of a dereddened SED with a CK-model scaled to the V band that allows for measuring luminosity is presented in Fig. 2.3. In the process of deriving the luminosity it is also possible to derive the IR excess. The different dereddened observed fluxes from the J band (1.24 μm) to the W4 band (22 μm) were logarithmically interpolated and the IR excess (E) was defined as:

$$E = \frac{(F_{\rm e} - F_*)_{[\lambda_1,\lambda_2]}}{F_*}, \tag{2.2}$$

where $F_{\rm e}$ is the total flux underneath the *observed dereddened photometry* (the IR photometry has also been dereddened) and F_* is the total photospheric flux below the CK model (see Fig. 2.3). λ_1 and λ_2 define the range of wavelengths of interest and the total fluxes in the numerator simply refer to that range. This measure expresses the excess in terms of the total luminosity of the object. For example, all things being equal, if two stars have the same amount of dust surrounding them, and one of them is brighter, the IR re-radiated emission will be larger but the IR excess, as defined here, would be the same, as it is a relative measure. The same or a very similar indicator was used by Cote and Waters (1987, their Eq. 8), Waters et al. (1987, Eq. 3), and more recently by Banzatti et al. (2018) in their Sect. 2.3.

Uncertainties in the IR excesses were derived using the uncertainties in the observed fluxes and the uncertainties in the temperature (which affect the CK models) of each object.

I have split the total IR excess into two: a NIR excess ($1.24 - 3.4$ μm, roughly the 2MASS region) and a MIR excess ($3.4 - 22$ μm, the WISE region). The values for these excesses are presented in Tables A.2 and A.5 for the high- and low-quality samples, respectively. The total IR excess ($1.24 - 22$ μm) is the sum of the two.

In addition, I also computed the IR excess at each individual band (J, H, K_s, W1, W2, W3, and W4) as the flux ratio between the dereddened observed monochro-

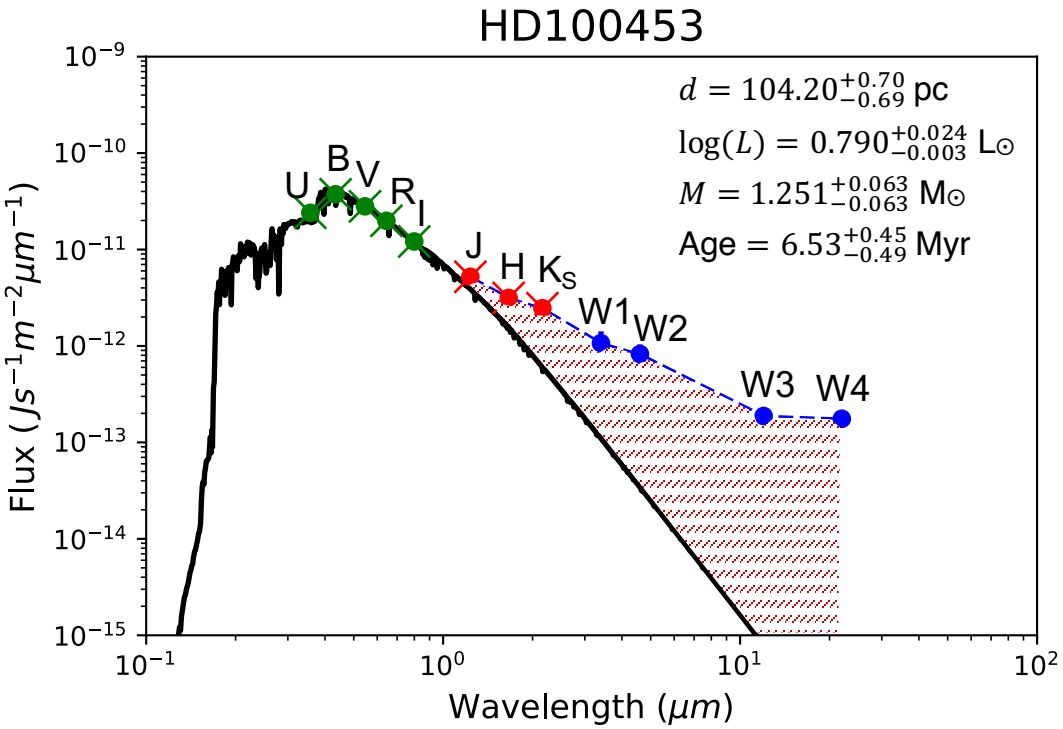


Fig. 2.3 Dereddened SED of HD 100453 Herbig Ae/Be star with a Castelli and Kurucz (2003) atmosphere model of the literature temperature in black scaled to the V band. 2MASS (J, H, K_s) and WISE (W1, W2, W3, W4) infrared bands are shown. Shaded area indicates the IR excess region considered at the numerator of Eq. 2.2. The derived parameters in Sects. 2.2.1 and 2.2.2 are shown on the top right. This Herbig Ae/Be SED is presented as an example of the 252 obtained in this chapter.

matic flux and the expected flux according to the CK model. The values for these excesses are presented in Tables A.3 and A.6 for the high- and low-quality samples, respectively

2.2.4 Variability information

Gaia DR2 does not provide a general variability indicator for all sources. Here, Gaia's repeated observations are used to extract photometric variability information. Gaia DR2 used a total of 22 months of observations and each source was observed repeatedly in a non-periodic fashion. Data Release 2 provides the average photometry, the uncertainty on this value, and the number of observations. All things being equal,

the photometric *error* will be larger for a photometrically variable object than for a stable object. Here I aim to quantify the variability of the objects. I start with the 'variability amplitude' (A_i) for a certain source i as presented in Deason et al. (2017):

$$A_i = \sqrt{N_{\mathrm{obs},i}}\, e(F_i)/F_i, \tag{2.3}$$

where N_{obs} is the number of CCD crossings, and F and $e(F)$ are the flux and flux error, respectively. This quantity is powerful in identifying objects that show larger flux variations than expected for a stable star. However, in order to statistically assess the level of variability, I introduce a variability indicator V_i, which quantifies how much more variable an object is compared to stable objects of the same brightness. In short it compares the variability amplitude from Eq. 2.3 of a given object (i) to that of all Gaia objects in a brightness interval of ± 0.1 magnitude around the G band value of the object (i.e. to $A_{a,G_a\in(a_1,a_2)}$, with a indexing the Gaia catalogue and being $a_1 = G_i - 0.1\mathrm{mag}$ and $a_2 = G_i + 0.1\mathrm{mag}$). The equation is as follows:

$$V_i = \frac{A_i - \overline{A_a}_{,G_a\in(a_1,a_2)}}{\sigma[A_a]_{G_a\in(a_1,a_2)}}, \tag{2.4}$$

where G is the Gaia white G band magnitude and σ is the standard deviation. In essence, I subtract the error to flux ratio of each HAeBe, weighted by the number of observations, to the mean of the same expression (A_a, Eq. 2.3) for the sources in the Gaia catalogue within ± 0.1 mag of the Herbig star in the G band. I then divide by the standard deviation of $A_a s$ of that Gaia subset. This results in a variability indicator which measures the variability (in standard deviations, σ) for each Herbig Ae/Be star compared to the mean of field objects of the same brightness.

For completeness, I note that it is necessary to impose more constraints to exclude the cases in which a larger error is not due to intrinsic variability. Following Deason et al. (2017), Appendix C of Lindegren et al. (2018) and what was done in Gaia Collaboration et al. (2018a) I require $N_{\mathrm{obs}} \geqslant 70$ and more than eight visibility periods

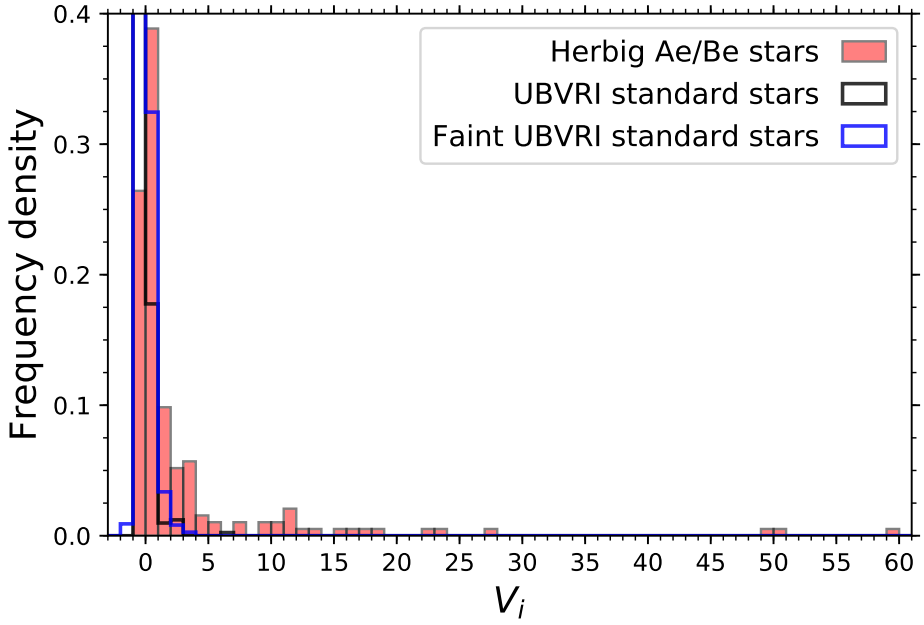


Fig. 2.4 Distribution of the variability indicator for Herbig Ae/Be stars and two catalogues of photometric standards; one of bright sources (Landolt, 2009) and one of faint sources (Clem and Landolt, 2016). As a class, the Herbig Ae/Be stars are more variable than the photometric standards.

(i.e. groups of observations separated by at least four days), in addition to the Eq. 2.1 constraint that limits the astrometric quality (and hence the variability indicator can simply be derived for sources in the high-quality sample). In order to also limit the photometric quality, I included the following criterion presented in Gaia Collaboration et al. (2018a):

$$1.0+0.015(G_{\rm BP}-G_{\rm RP})^2 < E_{\rm F} < 1.3+0.06(G_{\rm BP}-G_{\rm RP})^2, \qquad (2.5)$$

where E_F is the 'flux excess factor' and G_{BP} and G_{RP} the Gaia blue and red passbands, respectively. I note that these constraints may inevitably exclude many of the very variable HAeBes as they also trace larger errors and hence variability. These constraints will also be biased toward discarding binaries and faint sources in crowded areas (Lindegren et al., 2018; Gaia Collaboration et al., 2018a).

The variability indicator values for the 193 sources satisfying the previous conditions are presented in Table A.2.

In Fig. 2.4, I show the V_i distribution of Herbig Ae/Be stars and compare it to the V_i distribution of bright photometric standards from Landolt (2009) and faint photometric standards taken from Clem and Landolt (2016). If Eq. 2.4 had not been used these two latter samples would have had a different mean in the distribution of A_i. The Herbig Ae/Be stars appear to show, on average, a larger variability indicator value than the standard stars, which have typical values of $V_i \lesssim 2$. I performed a two-sample Kolmogorov-Smirnov (KS) statistical test to decipher whether or not Herbig Ae/Be stars can be drawn from those two samples of standard stars. The result shows that I can reject that hypothesis to within a 0.001 significance and hence this variability indicator differentiates them as a group.

In order to assess the relation between the variability indicator (G band variability) and variability in the V band I compared the magnitude variations in the V band as presented in the International Variable Star Index (VSX, Watson et al., 2006) with the variability indicator values. I found that variabilities as small as ~ 0.5 mag are being traced with the $V_i = 2$ cut-off. In Eiroa et al. (2002) 7/23 (30%) PMS objects homogeneously observed for variability have variabilities above 0.5 mag. In the case presented here, 48 out of 193 sources have values above $V_i = 2$ (25%) and hence can be considered as strongly variable. Of those 48, 17 are catalogued as UXOR type (Oudmaijer et al., 2001; Poxon, 2015). There are 5 other UXORs in the sample with V_i values, and 4 of them have reported optical variabilities smaller than 0.5 mag in the V band. The other one is BO Cep. This object has been reported to have a periodic variability with a single prominent peak with a period of ~ 10 days (Gürtler et al., 1999). The regular non-periodic variability of the object is smaller than 0.5 mag which explains why this UXOR has not been detected by the variability indicator. Supporting this, it appears as UXOR in Poxon (2015) but not in Oudmaijer

et al. (2001). Known UXORs are indicated in Tables A.2 and A.5 for the high- and low-quality samples, respectively.

To put the variability indicator into perspective, I find that 6 out of 411 photometric standards from Landolt (2009) have variability indicator values larger than 2. I would therefore expect only 3 of the 193 Herbig Ae/Be stars for which I could determine V_i to be strongly variable, at amplitudes of 0.5 magnitudes in the V band or higher. However, I find 45 more, indicating that a large fraction of Herbig Ae/Be stars exhibit strong variations.

In addition, it is interesting to compare the variability indicator with the variability catalogues published alongside the Gaia DR2 general catalogue (Holl et al., 2018). Just 1 every 3000 objects passed the Gaia DR2 stringent selection criteria for variability. Ten of the 252 objects in my list fall in this category and appear as variable in those catalogues. Of the five of those that have derived V_i values, they are larger than $V_i = 5$.

2.3 Data analysis

2.3.1 Infrared excesses

In Fig. 2.5 the total IR excess ($1.24 - 22$ μm) versus the estimated mass and age of the sources is plotted. There appears to be a difference in IR properties between high- and low-mass stars. Whereas low-mass stars show a range of IR excess, the higher-mass stars in general only present very low levels of excess. A similar behaviour is seen when the excess is plotted as a function of age; the excess for the youngest objects is smallest. This is probably readily explained by the fact that the more massive PMS objects in the HR diagram have the lowest ages by virtue of their rapidly evolving isochrones, so trends in mass will automatically also be present in those with age. To study trends as a function of age, it would be necessary to

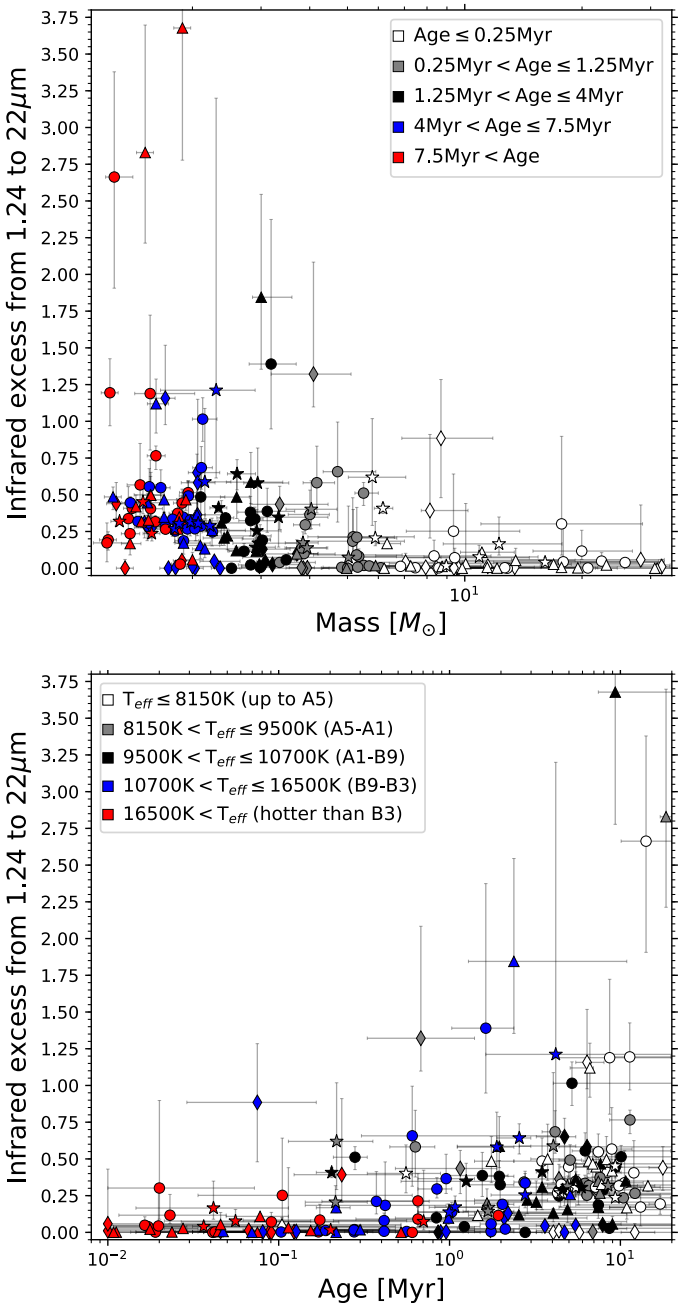


Fig. 2.5 *Top:* IR excess in the range $1.24 - 22$ μm vs. estimated mass of the objects. The most massive objects (more massive than ~ 7 $M_{\odot}$) barely show an IR excess. *Bottom:* IR excess in the range $1.24 - 22$ μm vs. estimated age. Ages and effective temperatures are respectively colour coded in the legend. The symbols stand for the Hα line profiles: circles (double-peaked), triangles (single-peaked), stars (P-Cygni profile), and diamonds (no information). I note that although it is not necessarily a one-to-one correlation, lower ages correspond to higher masses.

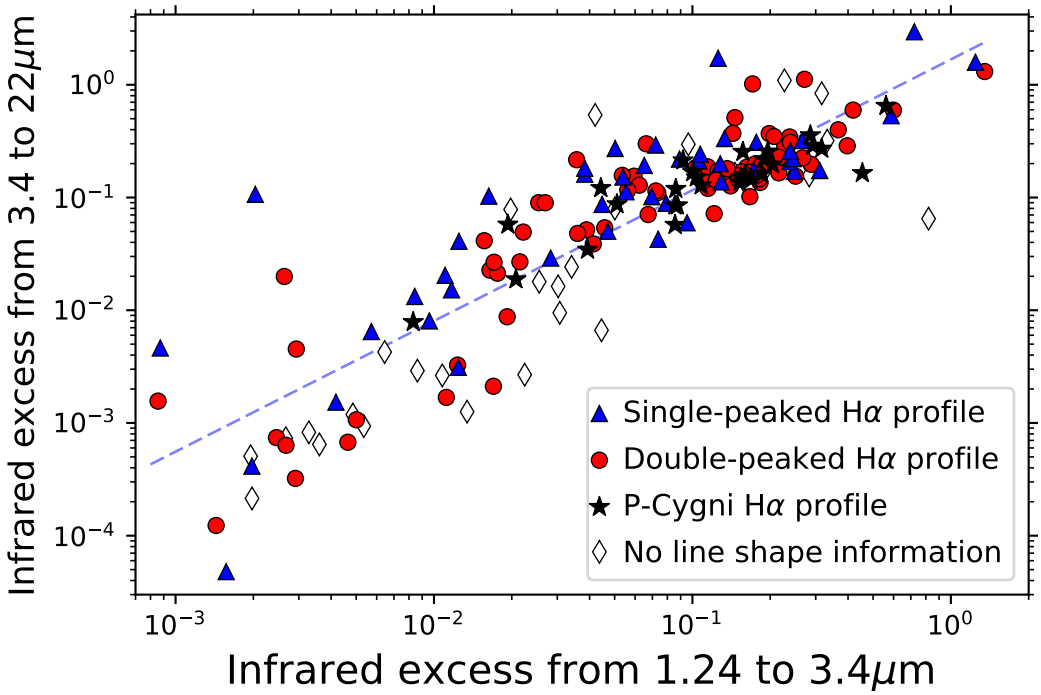


Fig. 2.6 IR excess in the range $3.4-22$ µm (MIR excess) vs. IR excess in the range $1.24-3.4$ µm (NIR excess). The symbols stand for the Hα line profiles: circles (double-peaked), triangles (single-peaked), stars (P-Cygni profile), and diamonds (no information). A linear fit in the log space is shown in blue ($\log(\text{Mid IR}_{\text{excess}}) = 1.16\log(\text{Near IR}_{\text{excess}}) + 0.23$, $r = 0.88$).

consider subsamples with a narrow range in mass. I therefore consider that the main result of this exercise is that high-mass objects have a very low IR excess, and that there appears to be a break at ~ 7 M$_{\odot}$ from where almost no sources with significant excess appear.

Figure 2.6 splits the total IR excess into two, a NIR and a MIR part. This demonstrates that the excesses at both wavelength ranges are highly correlated with each other (the linear fit in logarithmic space that can be seen in the plot has a correlation coefficient of $r = 0.88$). Therefore, it is not unexpected that the ~ 7 M$_{\odot}$ break is also present at NIR and MIR.

2.3.2 Hα equivalent width

Figure 2.7 shows the Hα EW as a function of mass and age, respectively. As the definition of a Herbig Ae/Be star includes the presence of emission, which is mostly from the Hα line, it may not come as a surprise that essentially all measurements are negative (i.e. tracing emission).

The EWs show a large range of values, which appears to increase with increasing mass and decrease with increasing age (studied by Manoj et al., 2006). The older objects typically have lower EWs than younger objects. It is tempting to read an evolutionary effect into this finding - after all it would be expected that the accretion (and therefore emission) would decrease when the PMS objects are closer to the MS. However, I should recall that there is a strong correlation between the age and the mass of the stars, so this may well be a mass effect instead. As the EW is a relative measurement with respect to the stellar continuum, a larger EW for otherwise similar objects indicates a stronger emission line. The observed trend towards higher temperatures/masses and thus higher luminosities implies that the lines become even stronger than the EW alone would seem to imply.

2.3.3 Hα equivalent width and infrared excess

The correlation between Hα emission, measured by its EW, and NIR and MIR excess is studied in Table 2.1 and Fig. 2.8 for each one of the IR bands (J, H, K_s, W1, W2, W3, and W4). In this case, I computed the IR excess as the flux ratio between the dereddened observed monochromatic flux and the expected flux according to the CK model at each band (see Sect. 2.2.3 and Fig. 2.3; the values for these excesses are presented in Tables A.3 and A.6 for the high- and low-quality samples, respectively). In all cases, there is a general and consistent increase of the Hα EW from sources with very little IR excess to those with higher IR excess.

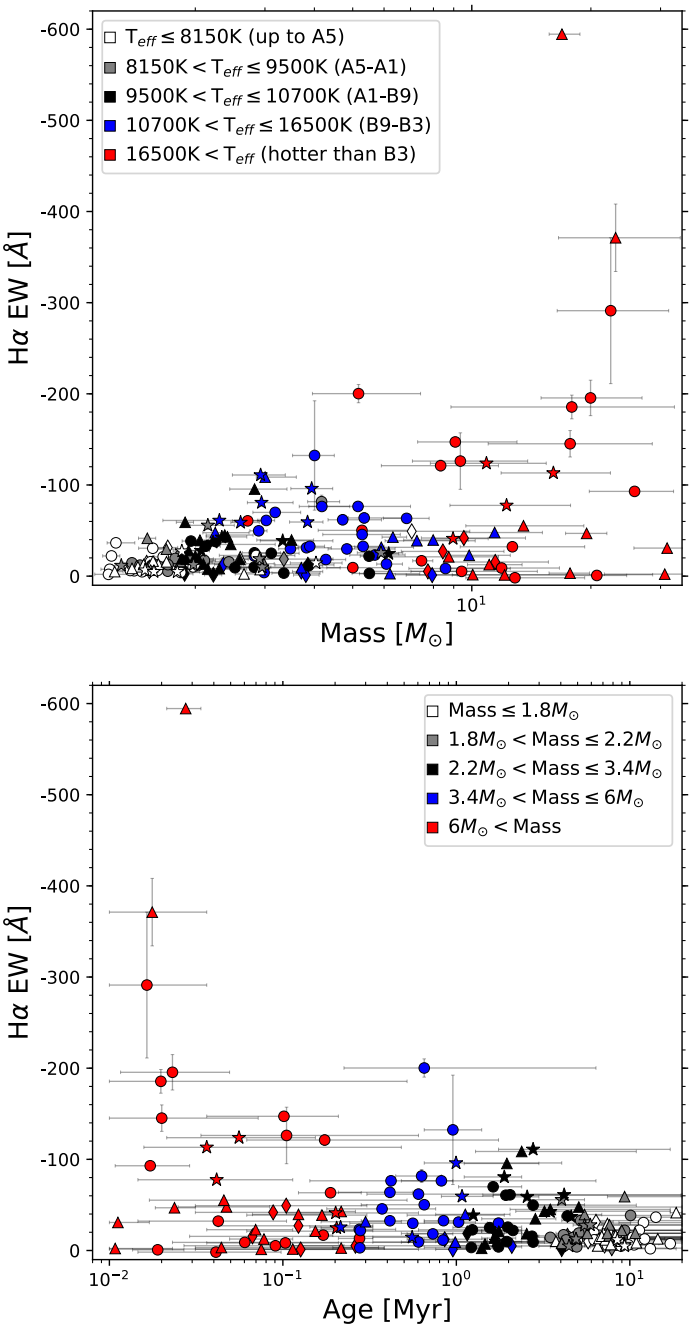


Fig. 2.7 *Top:* Hα EW vs. estimated mass. *Bottom:* Hα EW vs. estimated age. Effective temperatures and masses are respectively colour coded in the legend. The symbols stand for the Hα line profiles: circles (double-peaked), triangles (single-peaked), stars (P-Cygni profile), and diamonds (no information).

Table 2.1 Correlation between IR excess and Hα EW at different wavelengths.

Band	Correlation coefficient (r)	A	B
J (1.24 μm)	0.41	0.15 ± 0.03	0.025 ± 0.034
H (1.66 μm)	0.56	0.32 ± 0.03	0.0024 ± 0.0478
K_s (2.16 μm)	**0.60**	$\mathbf{0.48 \pm 0.05}$	$\mathbf{0.046 \pm 0.066}$
W1 (3.4 μm)	0.57	0.64 ± 0.07	0.15 ± 0.10
W2 (4.6 μm)	0.57	0.78 ± 0.09	0.24 ± 0.12
W3 (12 μm)	0.52	0.93 ± 0.12	0.79 ± 0.16
W4 (22 μm)	**0.41**	$\mathbf{0.71 \pm 0.12}$	$\mathbf{2.05 \pm 0.17}$

Notes: Correlation between IR excess (defined as a flux ratio, $F_{observed}/F_{CK}$) and Hα EW at different wavelengths for all the sources. The coefficients are defined by: $\log(F_{observed}/F_{CK}) = A \log(|EW|) + B$. The K_s band, with the higher correlation, and the W4 band are in bold; both are shown in Fig. 2.8.

In Table 2.1 I show that the Hα emission line EW is more correlated with the IR excess at shorter wavelengths than at larger wavelengths, with the correlation peaking at 2.16 μm (K_s band).

An obvious question might be whether there is a causal correlation between the Hα emission and presence of emission due to dust around these objects. The various excesses at various wavebands are correlated with each other (Fig. 2.6), and as a consequence the IR excesses at many wavelengths also correlate with the EW. However, the correlation with Hα is strongest at the K_s band which traces the hot dust in the inner disc, suggesting that the accretion mechanism or wind activity as traced by Hα is related to the inner parts of the dusty disc (see also Manoj et al., 2006). As presented in Table 2.1, the correlation rises from a minimum at 1.24 μm (effectively tracing the stellar photosphere) up to 3.4 μm and then goes down again to the same minimum at 22 μm (W4 band), where dust in the outer disc is found. In fact, Mendigutía et al. (2012) discovered the same correlation between IR excess and accretion rate and they found that it is no longer present beyond 20 μm.

For comparison purposes, in Fig. 2.8 the K_s band is plotted in the upper panel and the W4 band in the lower. It is noteworthy that for the K_s band, where the strongest

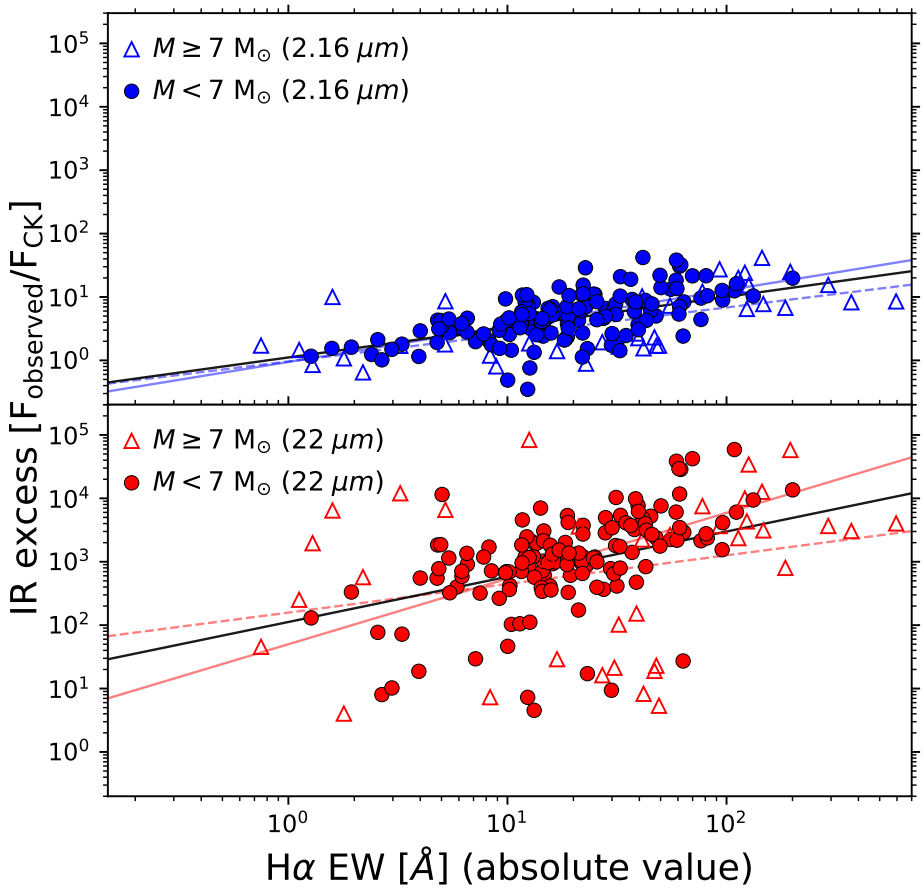


Fig. 2.8 2.16 µm (blue markers in top panel) and 22 µm (red markers in bottom panel) IR excesses defined as $F_{observed}/F_{CK}$ vs. Hα EW (absolute value). I note that this IR excess indicator is a flux ratio and not the one described in Eq. 2.2 where I integrated under the SED. Dots are Herbig Ae/Be stars with $M < 7$ M$_\odot$ and triangles are Herbig Ae/Be stars with $M > 7$ M$_\odot$. Lines are linear fits to the data, dashed for HAeBes with $M > 7$ M$_\odot$ and in solid colours for HAeBes with $M < 7$ M$_\odot$; black solid lines are the linear fits for all the sources (equations and correlation coefficients for these fits to all the sources for all the IR bands can be seen in Table 2.1).

correlation is found, small EWs are almost only present in sources with little IR excess and, in consonance with Sect. 2.3.1, for a given Hα EW value low-mass stars ($M < 7$ M$_\odot$) tend to have higher IR excesses. However, these trends are weaker or non-existent in the case of the W4 band, where I find the weaker correlation. This reinforces the idea that the Hα emission is correlated with the inner parts of the disc.

I note that in both panels the average excess is still lower for the higher-mass objects. The emission line strengths were subject of a follow-up study using accretion rates in Wichittanakom et al. (2020).

2.3.4 Variability

I conclude this section by studying the variability of the objects and its correlation with the various properties discussed so far, including the Hα line profiles taken from the literature.

The top panel of Fig. 2.9 presents the variability indicator as a function of the total (near- plus mid-) IR excess. As described in Sect. 2.2.4, the variability indicator states the number of standard deviations separating a certain source from the mean of the Gaia objects of the same brightness. No, or hardly any variability is present at the lowest IR excesses but sources can be both variable and non-variable at the higher IR excesses, consistent with van den Ancker et al. (1998) based on a smaller sample.

The bottom panel of the same figure shows the variability as a function of mass. As high-mass stars in this sample generally do not have a strong IR excess, I find that mostly the lower-mass and cooler objects display high variabilities, with the break also around 7 $M_{\odot}$, corresponding to a MS spectral type of around B1.5-B2. Although cooler objects tend to have larger variabilities (also observed by van den Ancker et al., 1998), I can observe how the range in temperatures for variable sources is wide in the bottom panel of Fig. 2.9, and that there are in fact many Herbig Be stars with very strong variabilities. Therefore, this is more likely a trend with mass and not with temperature. I note that although I detect photometric variability from the $V_i = 2$ value, the $V_i = 5$ value is a better separation boundary for the observed trends in both panels of Fig. 2.9.

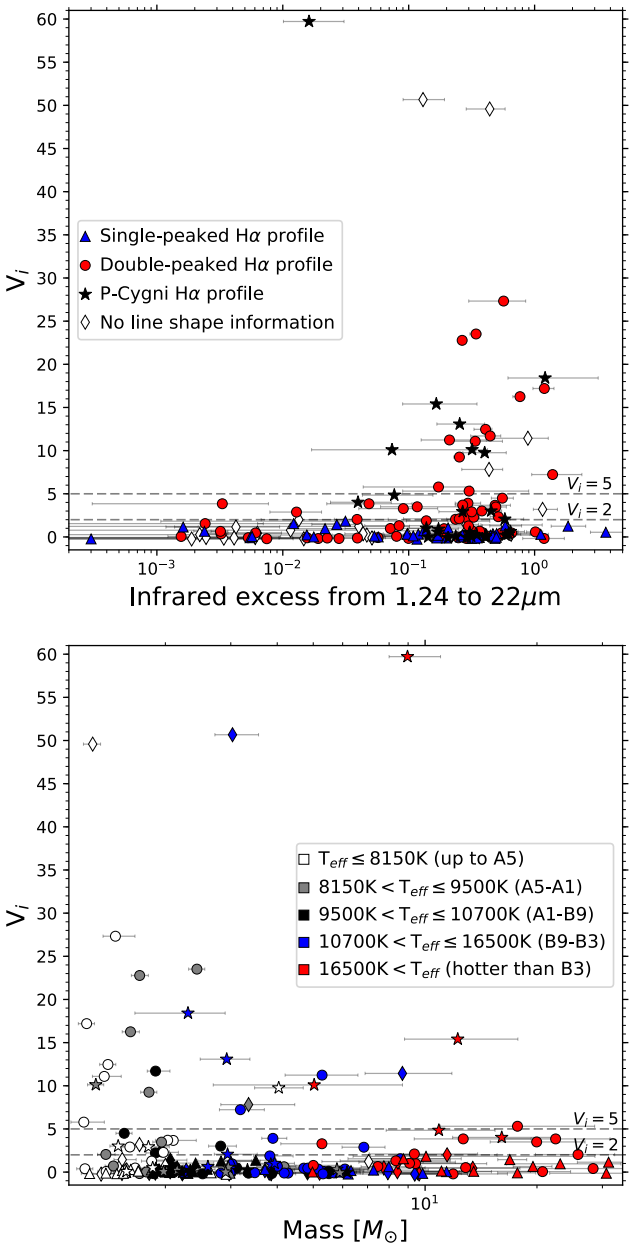


Fig. 2.9 *Top:* Variability indicator vs. IR excess in the range $1.24 - 22$ μm. It can be seen how objects with the lower IR excess do not show high variability. *Bottom:* Variability indicator vs. estimated mass. It can be seen how the most massive objects (more massive than ~ 7 $M_\odot$) barely show any variability. Line profiles and temperatures are colour coded in the legend in the top and bottom panels, respectively. The symbols stand for the Hα line profiles: circles (double-peaked), triangles (single-peaked), stars (P-Cygni profile), and diamonds (no information). The $V_i = 2$ and $V_i = 5$ values are stressed for clarity.

The challenge is to identify which property lies at the cause of the variability; is it the mass of the objects, their age, or IR excess emission or something else? An important clue is that many objects with strong variability (above $V_i = 2$) and line shape information have doubly peaked Hα profiles (31 out of 43; 72±7%, 68% confidence interval). In general, double-peaked emission line profiles are due to rotating discs, so the data are suggestive of an edge-on disc-type orientation and structure (from the remaining 12 objects they all have a P-Cygni profile and none have a single-peaked profile). The number of variable objects with doubly peaked line profiles is significantly different from the full sample, in which only half of the targets with known line classifications have a double-peaked profile (of the sources with derived variability indicator and known line profile, 79 out of 155 or 51±4% are double-peaked and 48 out of 155 or 31±4% are single-peaked). These fractions are significantly different, and I therefore suspect that the variable sources are mostly oriented edge-on, and that the line-of-sight inclination to the objects could be a decisive factor in the cause of the variability. This is in agreement with the trend observed in the top panel of Fig. 2.9. Sources with large amounts of circumstellar material show large IR excesses and high or low levels of variability depending on the inclination of their disc whilst sources with little material around have low IR excesses and low variabilities in all cases (also discussed in van den Ancker et al., 1998).

2.4 Discussion

2.4.1 General findings

In the above I have determined fundamental parameters such as temperature, mass, age, IR excess, variability, and luminosity for a large sample of Herbig Ae/Be stars which was made possible due to the more than a factor of ten increase in available

distances to these objects compared to Hipparcos. With the Gaia DR2 data, the majority of known Herbig Ae/Be stars could be placed in the HR diagram. I found the following:

- There are more low-mass objects than high-mass objects, with the high-mass objects mostly located close to the MS.
- High-mass objects have in general very small IR excesses and low variability; the properties appear to differ around 7 $M_{\odot}$.
- Hα emission is generally correlated with IR excess, with the correlation stronger for IR emission at wavelengths tracing the hot dust closest to the star.
- More massive and younger objects have higher Hα EWs.
- When split at 7 $M_{\odot}$ into 'low'- and 'high'-mass samples, the Hα - IR excess correlations hold for both mass ranges, with the average excess being lower for the higher-mass objects.
- Photometric variability can be traced back to those objects with double-peaked Hα emission and large IR excesses.
- All catalogued UXORs in the sample with detected variabilities above 0.5 mag in the V band appear as strongly variable (above $V_i = 2$) with the exception of BO Cep (discussed in Sect. 2.2.4).

Below, I discuss these findings and their implications for the formation of intermediate-mass stars.

2.4.2 Selection effects

Let me first investigate the various selection effects and biases that could potentially affect the results.

Quality parallaxes. It could be argued that the quality of the astrometric data has an effect on the findings. The parallax errors occupy a comparatively small range, from $\sim 0.016 - 0.37$ mas, but because of the large spread in distances, the relative uncertainties can be very large. To investigate whether this has a detrimental effect on the results, I repeated the analysis with only the objects with the very best parallaxes ($\varpi/\sigma_\varpi > 10$). This, of course, limits the sample and 182 out of 218 objects remain in the high-quality sample. These 182 objects are less luminous, which may be expected as in general they have larger parallaxes and are therefore closer. As a result they will be less massive and have greater ages than the objects in the entire sample. This, as a consequence of the trends described in previous sections, implies that these objects also show larger IR excesses and variabilities as well as smaller Hα EWs (see Figs. 2.5, 2.7 and 2.9). However, I find that essentially all correlations also hold for the higher-quality parallax sample, and if anything, they appear stronger. For example, almost all of the high-mass sources that have large IR excesses and variabilities in Figs. 2.5 and 2.9 have $\varpi/\sigma_\varpi < 10$. The inclusion of lower-quality parallaxes induces an extra scatter in the results, but the larger sample and wider coverage in luminosity aids in reinforcing them.

Quality identification as Herbig Ae/Be star. Another potential source of error is source misclassification. I have used the largest sample of Herbig Ae/Be stars known before this thesis (Chen et al., 2016 with some added from Alecian et al., 2013; Baines et al., 2006; Carmona et al., 2010; Fairlamb et al., 2015; Hernández et al., 2005; Manoj et al., 2006 and Sartori et al., 2010). The defining characteristics of HAeBes are not unique to the class, and can often also be found in other types of stars such as classical Be stars, which display Hα emission and a NIR excess (e.g. Rivinius et al., 2013), and evolved stars, which can have spectral types A and B, display hydrogen recombination emission, and be surrounded by dusty shells and discs such as the luminous blue variables and B[e] stars (Davies et al., 2007; Oudmaijer et al., 1998 on

HD 87643). It is therefore inevitable that some sources will have been misclassified (see Chapter 3 for a more detailed description of contaminants). It would be fair to say that the more 'classical' Herbig Ae/Be stars going back to the Herbig (1960) and The et al. (1994) papers have been studied in more detail and are better established as young PMS stars.

I therefore studied the The et al. (1994) sample of objects (their Table 1, 85 sources out of the 218) separately and find that all correlations do hold for this 'gold standard' sample as well. I do find that on average these objects have a larger Hα EW and have larger IR excesses than the full sample. These properties are the defining characteristics of a Herbig Ae/Be star, and it may not be surprising that the first objects to be proposed as Herbig Ae/Be stars are on average more extreme in these properties. Yet, again, as with the higher-quality parallax sample, the trends are still present in this sub-sample.

Mass distribution of the sample. The known Herbig Ae/Be stars have mostly been found serendipitously, and the first large-scale systematic search for them is done in Chapter 3. Yet, an interesting question is how representative the present sample is for the class. To this end, I consider the mass distribution of the objects. There are more or less the same number of low-mass, A-type objects as there are higher-mass B-type objects. There are more Herbig Be stars than might be expected from the IMF; however, the B-type objects are brighter and are sampled from a larger volume, as also attested by their smaller parallaxes. A larger fraction of Herbig Be stars in the sample should therefore be expected. When limiting the sample in distance, I obtain a Herbig Ae/Herbig Be ratio that is close to the IMF. As far as the mass distribution is concerned, I may say that the current sample is representative of the class. In Chapter 3 I draw an increased and well-selected sampled of Herbig Ae/Be stars from the Gaia catalogue.

Binarity. One may think that binarity may affect the observed photometry and for example produce fake levels of variability in the variability indicator. This is because binary sources tend to be more astrometrically and photometrically irregular. I studied the group of binaries against the group of isolated sources and overall I find that the known binaries are slightly brighter than the objects that have not been reported to be a binary. This is probably a selection effect in that brighter objects were more likely to be included in the binary surveys. I compared the brightnesses of binaries and non-binaries in the Baines et al. (2006), Wheelwright et al. (2010), and Leinert et al. (1997) studies separately and find that within the surveys there are indeed no brightness differences between binaries and non-binaries.

Returning to the Gaia sample; all other properties but IR excesses, including variability, are similar. I do find that binaries have in general slightly larger IR excesses. With the benefit of hindsight, this is perhaps something that could have been expected. Most of the binaries are distant binaries with separations larger than 0.1 arcsecond (Gaia's angular resolution is 0.4 arcsecond). Indeed, no binaries are found closer than 30 au and therefore binarity is not expected to play a significant role in the optical photometry. At the same time, companions could potentially contribute to the IR emission whose fluxes have been measured with apertures larger than the typical separations. Given that I do detect slight differences in IR excess between binaries and non-binaries, a preliminary inference would be that the companions may contribute to the IR flux in some cases.

2.4.3 Infrared excess as a function of mass

Figure 2.5 shows the IR excess as a function of mass and of age. There is a marked difference in the IR excess observed towards high- and low-mass objects. Herbig Be stars more massive than ~ 7 $M_{\odot}$ in general appear to have little to no excess, while the lower-mass objects show a wide range of excesses. There is also a trend

with age with the youngest objects having the smallest IR excess. Although it would be tempting to assume a causal relation between age and presence of dust, and try to explain why the youngest objects have the smallest amount of dust around them, I suspect the stellar mass is the dominant factor. The durations of the PMS evolutionary tracks are progressively shorter for higher masses, and an underlying relation between mass and IR excess would therefore also appear as a correlation between age and IR excess. This inverse correlation of infrared excess with stellar mass was also found in Ribas et al. (2015) and Arun et al. (2019).

Either way, the lack of dusty emission from high-mass objects is puzzling, as it would be expected the more massive objects to be formed in more massive clouds and therefore be more embedded. A natural conclusion would be that at any time of their PMS evolution, these young objects would be surrounded by more dust than their lower-mass counterparts, and therefore, at any stage, they would have a stronger IR emission. A counterargument is that the Herbig Be stars are predominately found closer to the ZAMS and are therefore more evolved, having dispersed their circumstellar material. Supporting this idea, Alonso-Albi et al. (2009) found, from their compilation of millimetre observations of 44 objects, that Herbig Be stars have much weaker millimetre emission than their later-type counterparts. In addition, they found that the masses of the discs around Herbig Be stars traced at millimetre wavelengths are usually five to ten times lower than those around lower-mass stars, with the boundary also around 7 $M_{\odot}$. These authors suggest that the disc dispersal is more efficient and faster in high-mass objects above 7 $M_{\odot}$. Indeed, the disc dispersal times are a steep, declining function with stellar mass, from millions of years for the lower-mass stars to tens of thousands of years for the highest-mass young stars of 10 $M_{\odot}$ and higher (Gorti et al., 2009).

The latter timescales are comparable to the evolutionary timescales as for example computed by Bressan et al. (2012) for these massive objects. Therefore, the observation here is consistent with the classical scenario that the Kelvin-Helmholtz

contraction timescale is much smaller for massive objects compared to the free-fall timescale of the collapsing parental cloud. In this scenario, the massive young stars only become visible once they are on, or close to, the MS - the so-called birthline. I discuss this further below, but note that with this interpretation one would still expect a range of IR excesses in any sample. This is consistent with what I find for massive objects (larger than 7 $M_{\odot}$); a large number of low-excess stars, but still a few with noticeable excess (see Fig. 2.5).

Moving to the lower-mass objects, which do display a large range of IR excess emission, an immediate question to ask is whether it is possible to detect any evolutionary effect in the sense that objects that are further evolved have smaller IR excesses, as one expected from the progressive dust dispersal, and as suggested by Fuente et al. (1998). For example, if the inside-out clearing model of disc evolution is correct, a trend at each PMS track from high excess to little excess should be observed.

However, it appears that Herbig Ae/Be stars do not show any consistent evolution of the IR excess from high to low excess at any mass range. There are many objects appearing younger than 2.5 Myr or even 1 Myr at all mass ranges with little IR excess. Arguably the lack of an evolutionary effect can be explained by the size of the error bars on for example the luminosity. The evolutionary timescales vary strongly with mass (and thus luminosity), masking any trend of IR excess emission with age. Here, I would highlight that many young Herbig Ae stars show little excess. By looking at these objects in the bottom panel of Fig. 2.2 it is not difficult to find sources with error bars small enough to discard the contribution of uncertainty to the problem. Finally, the contamination by binaries as discussed in Sect. 2.4.2 can play a role here as many HAeBes can still remain as undetected binaries.

I should also note that the underlying assumption of the evolutionary calculations is that the conditions under which the stars form are uniform, and the accretion rates are a smooth function of time, resulting in an overall similar evolution for all

stars. However, the final configuration is undoubtedly affected by inhomogeneities, varying accretion rates, and even the masses of the initial clouds. Nevertheless, looking for real evolutionary effects in the SEDs requires selecting subsamples of objects that are located at or close to the same mass tracks. In many cases this may require even more precise parallaxes than can presently be provided by Gaia. It also requires precise determinations of the atmospheric parameters and extinction values. A proper statistical study with high-quality parameters of the evolutionary properties of the HAeBes as they move towards the MS is therefore pending but is planned for the future.

2.4.4 Variability in terms of the UXOR phenomenon

An extreme case of large non-periodic photometric and polarimetric variations caused by discs is observed in UX Ori type stars (UXORs). Many of them are catalogued as HAeBes and their extreme variability is explained by eclipsing dust clouds in nearly edge-on sources and the scattering radiation in the circumstellar environment (see Grinin, 2000 and references therein; Natta et al., 1997; Natta and Whitney, 2000).

The variability indicator that was developed in this chapter (Eq. 2.4) specifically for the Gaia data demonstrates that the class of Herbig Ae/Be stars is more variable than the general population of stars. Figure 2.9 shows that the lower-mass objects are much more photometrically variable than those of higher mass, for which the variability appears to cease beyond ~ 7 $M_{\odot}$. The photometrically variable objects contain most of the so-called UXOR variables reported in the literature. Using the compilation of UXOR variables by Oudmaijer et al. (2001) and Poxon (2015), I find that 17 out of the 48 strongly variable objects - those with variability indicator values larger than 2, representing variations of 0.5 magnitudes (in the V band) or higher - are classified as UXORs. The remaining 5 UXORs with variability indicator values

present in the sample have documented variabilities below 0.5 magnitudes with the exception of BO Cep (discussed in Sect. 2.2.4).

The defining characteristic of the UXOR phenomenon is not only the photometric variability but also the reddening and blueing associated during the variations. The explanation put forward for this behaviour is the obscuration of the star by a rotating, inhomogeneous, dusty, edge-on disc. The objects first become redder when dust obscures the object, and can even become blue at their faintest phases, when the direct light from the stars is blocked and, predominately blue light is scattered into the line of sight. As the polarization - resulting from scattered light - also peaks during the faintest phases (e.g. Grinin 2000), the obscuring disc hypothesis is favoured. Interestingly, observational evidence other than the polarization supporting this conclusion has been relatively sparse.

With the large sample of Herbig Ae/Be stars, and the large number of UXORs among them, I can repeat a similar experiment using the Hα line as a proxy for the inclination of the circumstellar discs. I consider the line profiles of the Hα emission in tandem with the variability indicator. Figure 2.9 shows that all but twelve of the strongly variable objects with documented line profiles (above $V_i = 2$, those with $\Delta V > 0.5$ mag) have double-peaked Hα emission. The five objects for which no line profile is listed have, to my knowledge, no reported profiles. The occurrence of double-peaked profiles in the highly variable sample is significantly higher than for the other objects (see Sec. 2.3.4). It is significant that the other twelve objects have P-Cygni profiles and none of them show a single-peaked profile. The P-Cygni profile is often related to episodic energetic phenomena and it is not unexpected that it is also traced by the variability indicator. Given that doubly peaked line profiles are most easily explained by at least part of the emission originating in a rotating disc leads me to conclude that the photometrically variable objects are seen edge-on and surrounded by a disc-like structure. It is true that outflows or winds not limited to the disc can produce double-peaked Hα profiles (Kurosawa et al., 2006;

Tambovtseva et al., 2014). Supporting the hypothesis of edge-on discs being the main cause of photometric variability, I find in variability the same separation at ~ 7 $M_{\odot}$ between low- and high-mass objects that I found when studying IR excesses, which suggests that photometric variability and IR excess have the same cause. In addition, sources with high IR excesses have both high and low variability levels, which can be understood as depending on the disc inclination, while sources with lower IR excesses show little variability in all cases (top panel of Fig. 2.9, discussed in Sect. 2.3.4). This would also explain the few high-mass strongly variable objects that can be seen in the bottom panel of Fig. 2.9; they are mostly the ones with high IR excess in the top panel of Fig. 2.5 (discussed before in Sect. 2.4.3). Given that an edge-on orientation is the major and main ingredient of the dust obscuration hypothesis, these results lend very strong support to it using a large sample of Herbig Ae/Be stars.

The large fraction of objects with double-peaked line profiles or variability is in agreement with the model predictions by Natta and Whitney (2000) who worked out how many Herbig Ae/Be stars would undergo the UXOR phenomenon considering the scale heights of dusty discs and under which inclinations the photometric variability would still be visible. They conclude that around half of the Herbig Ae stars could be UXORs. In the high-quality sample I have 85 A-type stars with variability indicator values and just 16 of them were previously listed as UXORs; however, again, most of them have been largely unstudied. Nevertheless, of the 25 A-type stars with variabilities above $V_i = 2$, 13 are known UXORs. This means that for the Herbig Ae stars for which I detect variability at the $V_i = 2$ level, $\sim 52\%$ are known UXORs (and just two have P-Cygni profiles). Moreover, this implies that I am retrieving $\sim 81\%$ of known A-type UXORs with the variability indicator and hence, assuming that all the 25 A-type stars with variabilities above $V_i = 2$ are of UXOR type; 31 of the 85 A-type stars with variability indicator values in the sample should be UXORs. In turn this would imply that $\sim 37\%$ of all Herbig Ae stars belong to the UXOR class. If I also take into account that I have potentially removed some

UXORs from consideration, possibly the most variable ones, when applying the constraints described in Sect. 2.2.4, I get to values close to the 50% predicted by Natta and Whitney (2000). In general, there are 31 Herbig Ae/Be stars above the $V_i = 2$ level which have not been classified as UXOR and can be considered as new candidates of the class.

Finally, Davies et al. (2011) recently studied the UXOR object CO Ori in detail, which has single-peaked Hα emission. Consequently, they found that the inclination of its disc is of $\sim 30^\circ$ (i.e. it is nearly face-on). In this particular case, whether the disc is still causing the UXOR phenomenon or if it is caused through fluctuations in the circumstellar material outside the disc is still uncertain. I could not derive a variability indicator value for this object to assess its variability. Inspired by this example, I took a look at the other UXORs in the sample with single-peaked profiles; they all have variabilities below $V_i = 2$ in the variability indicator (HD 100546, HD 142527, HD 98922, and IL Cep), suggesting a category of low-variability UXORs with nearly face-on discs. Nonetheless, the results presented in this section strongly support the idea that most UXORs are caused by edge-on discs, which are responsible for large photometric variabilities.

2.4.5 Missing objects in the HR diagram

When inspecting the bottom panel of Fig. 2.2, it appears that most Herbig Be stars are located relatively close to the MS, whereas the lower-mass Herbig Ae stars occupy a larger part of their evolutionary tracks, contracting to higher temperatures at constant luminosity. In other words, the late-type Herbig Be and Herbig Ae stars at high luminosities (and low surface gravities) that would occupy the tracks towards the locations of B-type stars on the MS are missing. It is only due to the use of Gaia parallaxes, expanding the number of Herbig Ae/Be stars with well established luminosities, that this observation can be made.

In the discussion above, I mentioned the fact that these objects could still be heavily embedded in their parental clouds, preventing them from being optically visible when evolving on their way to the MS. There is evidence for optically invisible but IR-bright objects at locations in these regions of the HR-diagram. For example, Pomohaci et al. (2017) were the first to spectrally type an IR-bright MYSO based on the rare absorption spectrum at NIR wavelengths (higher-order Brackett lines are in absorption for this object, while Brγ is in emission). They found that the object could be fitted with the spectra of an A-type giant star. Had this object been optically visible, it would have occupied the empty region in the bottom panel of Fig. 2.2. To this, I add the early B-type Herbig Be stars/IR bright MYSOs PDS 27 and PDS 37 (Ababakr et al., 2015). They are found in the upper regions of the HR diagram, slightly off the MS. They are optically visible, but not overly bright at $V \sim 13$ mag, and have not been included in many (optical) magnitude-limited catalogues. Therefore, there are several examples that might lead to the conclusion that the - implicit - optical brightness limit of any catalogue of Herbig Ae/Be stars would prevent the inclusion of massive PMS stars on the horizontal portions of the evolutionary PMS tracks. However, many of these objects are present in Gaia DR2, due to its ~ 21 mag faint end, although they are yet uncatalogued as HAeBes. Numerous new young massive PMS stars that were so far missing from the catalogues are identified in Chapter 3.

In a sense this is a situation similar to that outlined for the low IR excesses observed toward the Herbig Be stars that are mostly located close to the MS. This could be explained by the fact that the objects would be embedded and thus optically invisible or faint in earlier phases of their evolution. Further observations of optically fainter objects are necessary to settle this issue (see Chapter 4). Additional progress can be made by connecting the PMS evolutionary tracks with radiative transfer codes to provide synthetic observations (as e.g. Davies et al., 2011, or Zhang et al., 2014

for Massive Young Stellar Objects) extended to optical wavelengths in the Herbig Be mass range.

Related to the 'missing' high-mass stars in the HR diagram, it will be important to fill the historic, and entirely man-made, gap between the Herbig Ae stars and the T Tauri stars. The latter are confined to have spectral types G-K-M, and typically Herbig Ae/Be stars, in this case by definition, have spectral types A and B. This implies that historically most PMS studies have missed out the F-type stars, resulting in an incomplete coverage of the HR diagram for PMS stars. See Chapter 4 for a further discussion of this topic.

2.4.6 The difference between Herbig Ae and Herbig Be stars

From the above discussion it appears that the dusty discs surrounding Herbig Ae and Herbig Be stars are different, with the break in IR excess occurring at 7 $M_{\odot}$ (around B1.5-B2 spectral type), a value which was also found by Alonso-Albi et al. (2009) from their compilation of millimetre emission tracing the outer parts of the dusty discs. As discussed, given the much stronger radiation field from B-stars, both in intensity and photon-energies, the most straightforward explanation for the much less massive discs of higher-mass objects is a more efficient disc dispersal mechanism (see e.g. Gorti et al., 2009). This also explains why the same 7 $M_{\odot}$ break is seen in variability (Fig. 2.9). As described in Sect. 2.4.4, the high levels of variability in some sources are caused by edge-on dusty discs. A more efficient disc dispersal mechanism beyond 7 $M_{\odot}$ would result in these sources showing no strong variability in the indicator. It also explains why the objects with the lower IR excesses are not strongly variable while the rest can have both high and low variability values.

Other studies of large samples of Herbig Ae/Be stars indicate a break in properties at a much lower mass of $3-4$ $M_{\odot}$. Fairlamb et al. (2015) studied the accretion rates, which are proportional to the mass of the objects, and found a different slope for

lower-mass than for higher-mass objects. Similarly, Wichittanakom et al. (2020) placed at $3.98^{+1.37}_{-0.94}$ $M_{\odot}$ the change in accretion properties (around the B3 spectral boundary). Ababakr et al. (2017), extending the work of Mottram et al. (2007), found a distinct difference in spectro-polarimetric properties across the Hα line between the Herbig Ae and late Be-type stars on the one hand and earlier Herbig Be-type objects on the other hand. These authors also point out the similarity in the Hα spectro-polarimetry of the Herbig Ae stars and T Tauri stars. Finally, Mendigutía et al. (2011b) noted the difference in Hα variability; Herbig Ae and late Be stars are largely variable, whereas early type Herbig Be stars are not. Later, Fang et al. (2013) showed that T Tauri stars display even more variable Hα emission - again hinting at a similar accretion mechanism for the T Tauri stars and Herbig Ae stars.

How can I reconcile the fact that some studies show a break in properties that is different from that of others? It is worth pointing out that the latter investigations consider regions much closer to the star than the dusty emission. Fairlamb et al. (2015) derive accretion rates from the UV excess which trace the shocked material on the stellar surface, and Ababakr et al. (2017)'s spectropolarimetry traces the free electrons in ionized material at distances of the order of stellar radii from the stars. The spectro-polarimetric properties of the B-type stars can be explained by stable circumstellar discs, while the line properties for T Tauri and Herbig Ae objects are consistent with magnetically controlled accretion. Likewise, the Hα emission (used by e.g. Wichittanakom et al., 2020) traces the ionized zones close to the star, such as the accretion columns and circumstellar discs, and the variability is explained by the accretion columns orbiting the central star (e.g. Kurosawa et al., 2008).

Earlier, I showed that the IR fluxes and Hα properties are largely correlated, especially at the wavelengths that better trace the hot dust of the inner disc, but that the IR fluxes are smaller for the earlier-type objects because of them being subjected to a more efficient disc dispersal mechanism. I therefore conclude this section with the observation that the IR and millimetre emissions trace the circumstellar discs and

originate much further from the stars than the UV, hydrogen recombination emission, and free electrons, which themselves trace the accretion onto the stars. The break in accretion mechanism appears to occur around 4 $M_{\odot}$, whereas the disc dispersal becomes significant at higher masses of 7 $M_{\odot}$.

I conclude this chapter with this discussion. Summarising, in this chapter I have homogeneously characterised all previously known and proposed Herbig Ae/Be stars with parallaxes in Gaia DR2. Their established parameters were compiled and masses, ages, luminosities, distances, photometric variabilities, and infrared excesses were derived. The HR diagram presented in this chapter contains over ten times more objects than was possible previous to Gaia. All this information is used in the next chapter to find new objects of the Herbig Ae/Be class.

Chapter 3

New Herbig Ae/Be and classical Be candidates

This emerged from my conviction, that if stars form at 1–2 solar masses, then more massive stars must also form and ought to be recognizable in some way.
Autobiographical notes, 1993 (as quoted by Reipurth, 2016) – George Herbig

In the introduction (Chapter 1) and Chapter 2 of this thesis it was shown the usefulness to obtain a large homogeneous and low biased catalogue of new Herbig Ae/Be stars. In this chapter I use an algorithm based on an artificial neural network (ANN) to identify new Herbig Ae/Be stars within Gaia DR2. Herbig Ae/Be stars share many characteristics with other types of objects, of which classical Be stars stand out. The ANN-based algorithm spotlights on disentangling these two types of objects, and as a consequence new classical Be candidates are also found.

This chapter is organised as follows: in Sect. 3.1, I describe the observables, features, and the metrics used for evaluating the performance of the algorithm as well as the sources that the algorithm classifies once it is trained. In Sect. 3.2 I present the labelled sources used for training the ANN. The algorithm itself is detailed in Sect. 3.3. In Sect. 3.4 I describe and evaluate the output of the algorithm which I analyse in Sect. 3.5, describing its flaws and biases.

3.1 Observables, features, and data

ANNs are supervised learning classifiers, this means that they need to be trained with a list of known sources (training set) that have a set of characteristics (features) and a label (ground truth) that assign them to a certain category (e.g. a stellar class). The features are the individual properties or characteristics that are used by the ANN to learn how to classify new sources. Feature selection is important, as the use of useless features or the lack of very relevant ones for differentiating the categories can heavily affect the performance of the algorithm. Once trained, ANNs assign probabilities of belonging to every one of the chosen categories to each input source. A brief introduction to artificial neural networks is presented in Appendix B.

The known HAeBes constitute a small, biased, and contaminated set (described in Chapter 2). In order to achieve a good training performance the strategy adopted was to include T Tauri stars in the training and use an algorithm focusing on the high-mass end. In the resulting catalogue of new PMS candidates, the most massive ones can be further selected by means of the Hertzsprung-Russell (HR) diagram.

3.1.1 Observables

The features that feed the ANN need to be relevant for identifying PMS sources. Hence, I want the features to trace the main observational characteristics of PMS sources. As described in the introduction (Chapter 1), the main observational characteristics of PMS sources are: IR excesses, emission lines and photometric variability. If HAeBes were unique in these properties, a simple linear separation in the parameter space would suffice for identifying more objects of the class (e.g. in a colour-colour plot). However, HAeBes share these characteristics with other types of objects, of which classical Be (CBe) stars stand out, as their outwardly diffusing gaseous discs generate very similar observables (Grundstrom and Gies, 2006; Rivinius et al., 2013; Klement et al., 2017).

To feed the algorithm with these characteristics, in this thesis I used observables belonging to different surveys. These are mainly Gaia DR2 for variability, 2MASS (Skrutskie et al., 2006) and WISE (Wright et al., 2010) for near- and mid-IR excess respectively, and IPHAS (Drew et al., 2005; Barentsen et al., 2014) and VPHAS+ (Drew et al., 2014) for Hα emission.

I used the following passbands: from Gaia DR2, the broad white G band (0.59 µm), and the blue (G_{BP}) and red (G_{RP}) bands (0.50 µm and 0.77 µm respectively). A description of the Gaia filters can be found in Evans et al. (2018). From IPHAS and VPHAS+, I used the SDSS passband *r* (0.62 µm) together with the Hα narrow filter (0.66 µm). A description of the IPHAS passbands and associated footprints can be found in Drew et al. (2005) and Barentsen et al. (2014) (for the second data release that I am using) and in Drew et al. (2014) for VPHAS+. Finally, from 2MASS I used the three passbands J, H, and K_s (1.24 µm, 1.66 µm and 2.16 µm respectively) and from WISE the four passbands W1, W2, W3, and W4 (3.4 µm, 4.6 µm, 12 µm, and 22 µm respectively). These passbands of 2MASS and WISE were obtained from the AllWISE catalogue, which is described in Cutri et al. (2013).

It is important when setting up the features to be cautious about introducing unwanted bias regarding the selection to be performed. An example of an unwanted bias is, for example, to introduce distance as a feature. Most of the known PMS objects are close-by because it is easier to study bright objects. If I introduce a distance dependent feature the algorithm would work with the idea that being close is an intrinsic property of PMS objects, and it would be biased to find PMS objects that are nearby. In addition, if I introduce position dependent features any posterior analysis about the clustering properties of Herbig Ae/Be stars would be biased towards the selected preferred positions of the training data. Therefore, I set the features to be distance and position independent, which implies that most of the observables used are colours. Of course, there are unwanted biases in the resulting catalogues because of the selected features, and they are addressed in Sect. 3.5.

For example, interstellar extinction results in colours that are not strictly distance-independent, and by demanding to have detections in all the WISE bands I am biasing myself to the most extreme IR-bright sources.

In total, I chose 48 observables from Gaia DR2, 2MASS, WISE, and IPHAS and VPHAS+ data. The colours used are $r-H\alpha$ plus all combinations of the passbands of Gaia DR2, 2MASS, and WISE (i.e. $G_{BP}-G$, $G_{BP}-G_{RP}$, $G_{BP}-J$, $G_{BP}-H$, $G_{BP}-K_s$, $G_{BP}-W1$, $G_{BP}-W2$, $G_{BP}-W3$, $G_{BP}-W4$, $G-G_{RP}$, $G-J$, $G-H$, $G-K_s$, $G-W1$, $G-W2$, $G-W3$, $G-W4$, $G_{RP}-J$, $G_{RP}-H$, $G_{RP}-K_s$, $G_{RP}-W1$, $G_{RP}-W2$, $G_{RP}-W3$, $G_{RP}-W4$, $J-H$, $J-K_s$, $J-W1$, $J-W2$, $J-W3$, $J-W4$, $H-K_s$, $H-W1$, $H-W2$, $H-W3$, $H-W4$, K_s-W1, K_s-W2, K_s-W3, K_s-W4, $W1-W2$, $W1-W3$, $W1-W4$, $W2-W3$, $W2-W4$, $W3-W4$). The idea behind using all these combinations is that it is not entirely known which colours are ideal for selecting PMS objects, so I let principal component analysis (PCA) facilitate this (see Sect. 3.1.2). The reason why neither r nor Hα passbands are combined with the other passbands is explained in Sect. 3.2.1.

In addition, I constructed two observables, G_{var} and V_{htg}, that trace optical photometric variability and are based on the Gaia passbands. I define G_{var} as:

$$G_{\rm var} = \frac{F'_{\rm G}\, e(F_{\rm G})\sqrt{N_{\rm obs,G}}}{F_{\rm G}\, e'(F_{\rm G})\sqrt{N'_{\rm obs,G}}}, \tag{3.1}$$

where $F_{\rm G}$ and $e(F_{\rm G})$ are the Gaia G band flux and its associated uncertainty for a certain source and $N_{\rm obs,G}$ the number of times that source was observed in the G band. The idea is that variable sources have larger uncertainties (weighted with the square root of the number of observations) than non-variable ones. $F'_{\rm G}/e(F'_{\rm G})\sqrt{N'_{\rm obs,G}}$ refer to the median value of Gaia DR2 sources of the same brightness. This denominator is necessary as non-variable objects of different brightness show different median uncertainties (see Fig. 3.1, the sources presented in that figure are the ones used to construct this denominator). A similar indicator was used in Chapter 2 to

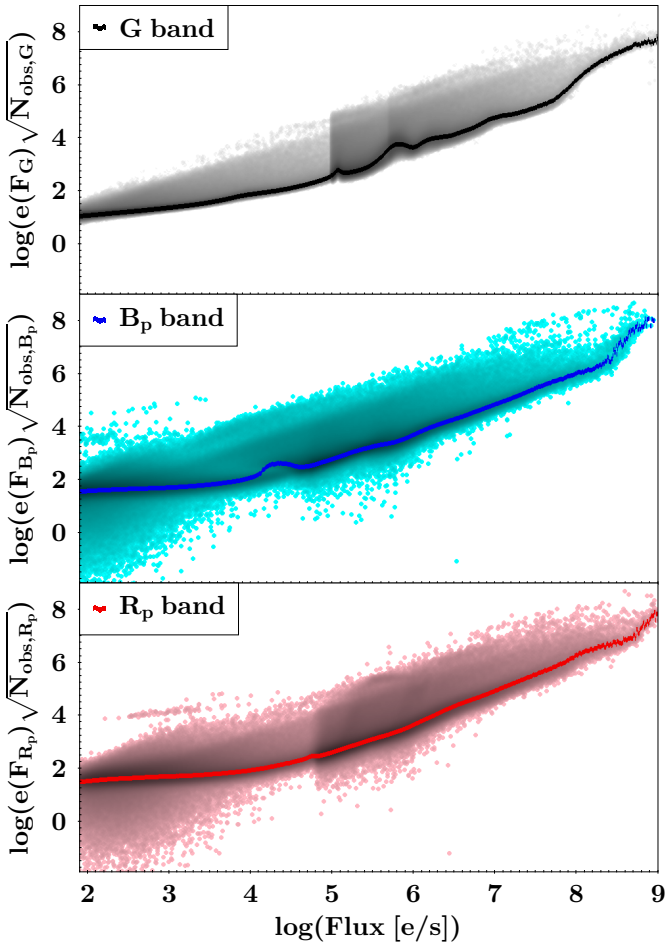


Fig. 3.1 Uncertainty in the Gaia G, G_{BP}, and G_{RP} band fluxes scaled to the square root of the number of observations in each band versus the associated Gaia flux for a random subset of about ten million Gaia DR2 objects. These random sources were selected from different flux bins in order to cover the whole flux space and not only the faint end. The lines trace the median value for each flux.

study the variability of known Herbig Ae/Be stars. In that chapter it was evidenced that this variability proxy mostly traces irregular (i.e. non-periodic) variabilities caused by material on the line of sight, so I expect it to be efficient in separating CBes from HAeBes. I define the heterogeneous variability (V_{htg}) as:

$$V_{\rm htg} = \frac{F'_{\rm B_p}\, e(F_{\rm B_p})\sqrt{N_{\rm obs,B_p}}}{F_{\rm B_p}\, e'(F_{\rm B_p})\sqrt{N'_{\rm obs,B_p}}} - \frac{F'_{\rm R_p}\, e(F_{\rm R_p})\sqrt{N_{\rm obs,R_p}}}{F_{\rm R_p}\, e'(F_{\rm R_p})\sqrt{N'_{\rm obs,R_p}}}. \tag{3.2}$$

This V_{htg} observable is based on the same idea as G_{var} but it evaluates the heterogeneous variability that may be present among the blue (G_{BP}) and red (G_{RP})

filters. This may arise, for example, by circumstellar material causing irregular extinction episodes (as is the case in the reddening and blueing associated during the variations of UX Ori type stars, Grinin, 2000) or by variable accretion.

3.1.2 Features

I use PCA to select the optimal set of features for this problem. When applying PCA to the complete set of 48 observables I obtain 48 principal components. However, in my pipeline only 12 of those principal components carry 99.99% of the variance (see Sect. 3.3.3). These principal components that carry almost all of the variance of the space of observables constitute the set of features. In other words, these principal components are the features used by the ANN. PCA also removes any linear dependency between the observables. A brief description of the PCA algorithm is presented in Appendix B.

3.1.3 Evaluation metrics

I use two correlated metrics, precision (P) and recall (R). They are defined as follows:

$$P = \frac{TP}{TP + FP}, \tag{3.3}$$

where TP is the number of true positives, that is, the number of sources of a certain category correctly catalogued, and FP is the number of false positives, this is the number of sources of the same category wrongly classified[1]. In other words, of all objects for which I have predicted a certain category, P describes what fraction was correctly classified. Separately:

$$R = \frac{TP}{TP + FN}, \tag{3.4}$$

[1]Other metrics, like the Area Under the Curve or the F_1 score, were discarded because FP is over-measured in this classification problem due to contamination in the training data, see Sect. 3.3.1.

where FN is the number of false negatives, that is, the number of sources that belong to a certain category but were not classified as such. In other words, of all objects that are actually of a certain class, R describes what fraction have I detected as belonging to that class, introducing a notion of completeness. These metrics are defined independently for each category.

3.1.4 Data

Before describing the training data, it is necessary to assess how many sources exist with all the observables I am using (Sample of Study, SoSt hereafter). The first step for generating this SoSt was to cross-match the catalogues that contain the required observables (Gaia DR2, AllWISE, IPHAS, and VPHAS+). Examples of works where this was done to a high level of accuracy are Scaringi et al. (2018) for Gaia DR2 with IPHAS and Marrese et al. (2019) for Gaia DR2 with AllWISE, among others. However, these cross-matches arrived at a high level of accuracy by sacrificing completeness (see Wilson and Naylor, 2018). PMS sources in particular, because of their variability and preferred location in extincted and crowded regions, tend to be excluded in those general cross-matches (e.g. only $\sim 52\%$ of the known HAeBes of Chapter 2 are present in the AllWISE 'BestNeighbour table' of Marrese et al., 2019). Instead, I perform a more generous cross-match accepting that I may generate some incorrect associations.

I first cross-matched Gaia DR2 (using epoch 2000 adapted coordinates) with IPHAS and VPHAS+ independently with a 1 arcsecond aperture because that is approximately the angular resolution of VPHAS+, IPHAS being slightly worse. 95% of the sources are found to be within 0.25 arcsecond. These two catalogues present a further complication. They present different observations of the same source as different entries and hence produce duplications in the cross-match. Therefore, in those cases I chose the observation with data in all the passbands, if any. If none or

more than one of the observations have information in all the passbands I chose the one with a higher quality flag and, in the case of having the same flags, I chose the object with the smaller angular distance to the Gaia DR2 source. Similarly, whenever a Gaia DR2 source was present in both IPHAS and VPHAS+ I gave priority to the observation with all the passbands, followed by the one with a higher quality flag and, in the case of having the same flags, to the object with the smaller angular distance to the Gaia DR2 source. Then, I performed another cross-match using Gaia DR2 coordinates with AllWISE, using a cross-match aperture of 2 arcsecond. This cross-match aperture, though large, was chosen after the experience in Chapter 2 where even a 3 arcsecond aperture was still not sufficient for some HAeBes. 95% of the sources are found to be within 1.12 arcsecond. This last cross-match provides me with a set of 51,548,230 sources. However, missing values are not allowed in ANNs and only 4,151,538 sources (8% of the original set) have all the 48 observables (see Sect. 3.1.1). This constitutes the SoSt (Fig. 3.2), the master sample of all the objects with the data necessary to enter the ANN. This set has a mean of $G = 16.7 \pm 2.0$ mag (error is 1σ of the mean) so 98% of the sources are in the range $12.3 < G < 20.3$ mag. The mean parallax is $\varpi = 0.36 \pm 0.75$ mas. I note that the Gaia parallax is not available for all the sources. The sky footprint of the SoSt is not homogeneous as it is limited by the combined footprint of the surveys used. Primarily, IPHAS and VPHAS+ are limited to the galactic plane ($5.5° > b > -5.5°$) and VPHAS+ footprint ($29° > l > -145°$, see Fig. 3.2) is largely incomplete at the time of writing. In addition, spurious WISE photometric detections in the galactic plane are a known issue (Marton et al., 2019 and references therein). Furthermore, due to the Gaia scanning law, Gaia DR2 itself presents a heterogeneous footprint completeness. Finally, demanding proper detections up to W4 (22 μm) and in the Hα passband excludes many objects and it may be expected to have overdensities of SoSt sources around star forming regions. The impact of this footprint of Fig. 3.2 in the final catalogues is addressed in Sect. 3.5.2.

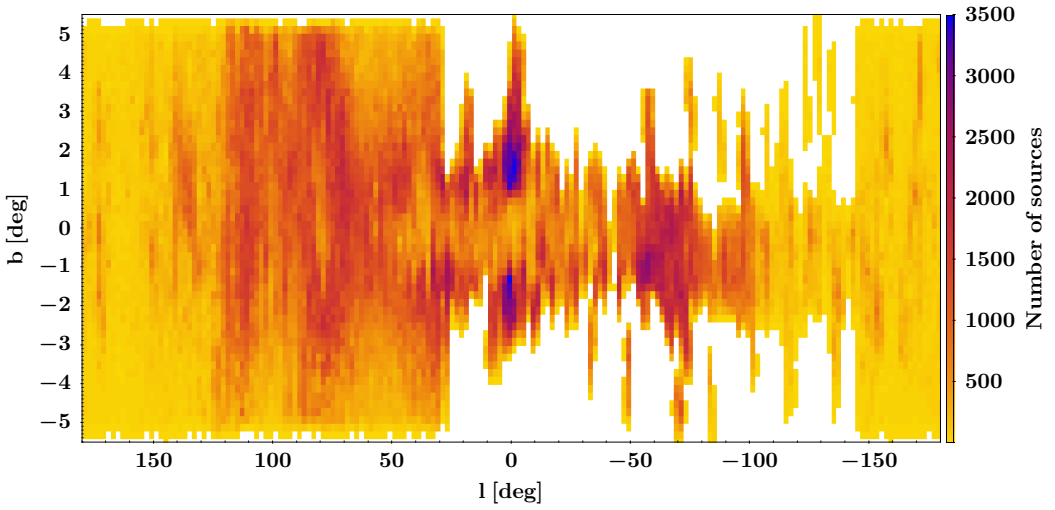


Fig. 3.2 Sky footprint of the SoSt in galactic coordinates, colour-coded by number density. Note the heterogeneity of the footprint. The scarcity of sources between $29^\circ > l > -145^\circ$ is due to the incompleteness of VPHAS+ at the time of writing. Each pixel is $2^\circ \times 0.2^\circ$.

As the beams of IPHAS, VPHAS+, and AllWISE are larger than Gaia's, different Gaia sources could have been assigned to the same IPHAS, VPHAS+, or AllWISE source. This can be the case if various Gaia sources are present within the same beam or if a wrong assignation was done in the generous cross-match. Indeed, 4.9% of the AllWISE sources are repeated and 0.31% of the IPHAS or VPHAS+ objects. These do not affect the classification, as the values are too small to have a significant impact on the training or the final catalogues (see Sects. 3.2.4 and 3.5.2 respectively). However, this implies that on average 1/42 (regarding AllWISE) and 1/625 (regarding IPHAS or VPHAS+) sources of the SoSt are fake, in the sense that its associated photometry does not belong to them, or it is a mixture of all the Gaia sources within the same beam. Another way of estimating the number of purely incorrect cross-matches is by comparing the Gaia passbands and colours with the AllWISE and IPHAS or VPHAS+ ones. In the case of AllWISE I compared $G_{RP} - J$ vs. $J - H$, which are strongly linearly correlated, and labelled as potential incorrect cross-matches those sources that were beyond 0.5 mag of the best linear fit. This results in about 2.3% bad matches for AllWISE. In the case of Hα I compared G_{BP}

vs. r (there is no linear relation between colours) and labelled as potential incorrect matches those sources that were beyond 1 mag (to account for variability) of the best linear fit. This results in a contamination of 1.3% for IPHAS or VPHAS+. Therefore, I conclude that the cross-matches are good to the $\sim 98\%$ level.

I did not take into account the quality flags of the catalogues. This decision was made for two reasons. First, IPHAS and VPHAS+ have very stringent quality indicators, and by limiting myself to sources with a good flag in these catalogues I reduce the size of the training set significantly (e.g. the SoSt previously described, shown in Fig. 3.2, would be reduced to 47% the size of the original set). Similarly, the mid-IR colours W3 and W4 tend to have very poor quality flags. Only $\sim 10\%$ of the sources within Gaia and AllWISE with information in all passbands have reliable mid-IR measurements (Marton et al., 2019). However, in this work the mid-IR is of paramount importance and cannot be excluded, as it is where the discs around HAeBes start to differ from the dust-free discs around classical Be stars (Waters et al., 1988; Rivinius et al., 2013). Second, because introducing cuts in the training data based upon quality criteria can introduce uncontrolled biases in the subsequent selection. This is because these quality flags are a result of a combination of very different factors. It is preferable to let the ANN deal with bad quality photometry as well as contaminants. Nonetheless, these quality flags are added to the final catalogues of new PMS and CBe candidates (Sect. 3.4 and Tables 3.3 and 3.4). The consequences of using low-quality data are discussed in Sect. 3.5.2.

3.2 Labelled sources

As introduced in Sect. 3.1, it is necessary to select which categories the ANN should learn to classify. Then I need to label a set of sources as belonging to these categories and use them to train the ANN. These labels are considered as ground truth and any bias, trend or contamination of this sample inevitably results in a bias in the

final classification. In this section I describe the construction of this set of Labelled Sources, which is a subset of the Sample of Study. The complementary subset of the SoSt that is not labelled (Input Set) is the one classified by the trained ANN (see Sect. 3.3 and Fig. 3.5 for further details).

I use one category of PMS sources and another category of classical Be stars, as telling the difference between these two groups is the main goal of the algorithm. In addition of learning from the characteristics of PMS and CBe objects I need the algorithm to learn from the existence of other similar or distinct sources that do not belong to these categories. This includes the erroneous or spurious data present in every catalogue. In other words, I need to construct a representation of what the algorithm finds when being applied to the Input Set. Hence, I use a third category of other objects, which comprises all types of sources present within the catalogues used that are neither a PMS source nor a CBe star. Therefore, the set of Labelled Sources contains already known PMS sources (Sect. 3.2.1), already known CBes (Sect. 3.2.2), and other objects (Sect. 3.2.4). In the following sections the construction of these three categories is described.

All known PMS and CBe sources considered with a good astrometric solution appear on the Gaia HR diagram (colour vs. absolute magnitude diagram) in Fig. 3.3. I define as sources with a 'good astrometric solution' those with a re-normalised unit weight error (RUWE parameter of Gaia DR2) of smaller than 1.4 and $\varpi/\sigma(\varpi) \geq 10$. Only these astrometrically well behaved sources have trustworthy positions in the HR diagram, as astrometry carries most of the uncertainty (see Chapter 2). However, those with a bad astrometric solution are still used by the algorithm as the observables are astrometry-independent (see Sect. 3.1.1). In this chapter I use the parallax to distance conversion of Bailer-Jones et al. (2018). In order to achieve the most accurate HR diagram positions I also needed to correct for extinction. Unfortunately, often it cannot be totally taken into account as in general the intrinsic extinctions are unknown. However, I corrected for interstellar extinction by using the dust map of

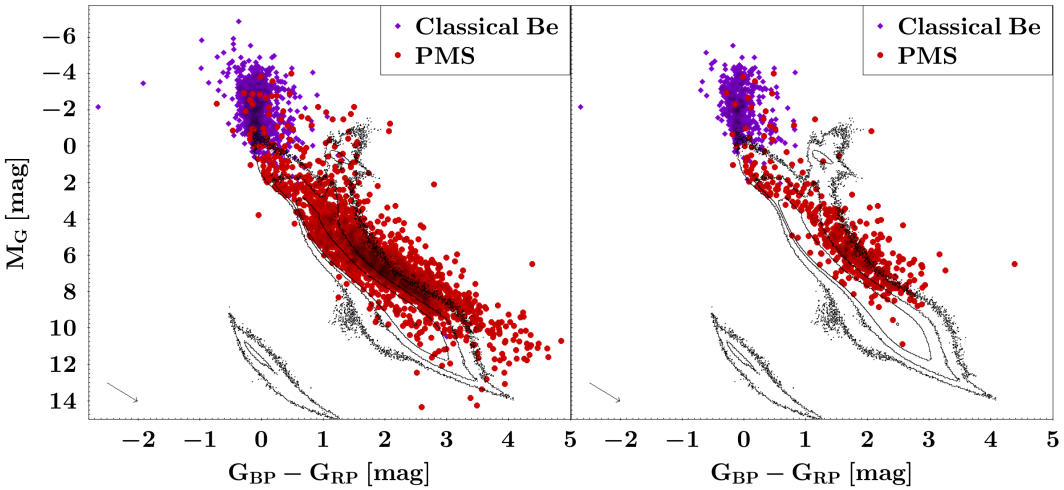


Fig. 3.3 Gaia colour vs. absolute magnitude diagram. Known PMS (in red circles) and classical Be stars (in violet diamonds) with good astrometric solution and corrected from interstellar extinction are plotted. An extinction vector corresponding to $A_G = 1$ is shown at the bottom left of each plot. Black contours trace Gaia sources within 500 pc with good astrometric solution. *Left:* All known sources. *Right:* The subset of sources with all the observables that are used for training. The very blue classical Be star is ω CMa and it probably has a spurious G_{RP} magnitude because of being brighter than the bright limit of Gaia DR2.

Lallement et al. (2019) and the extinction coefficients of Gaia of Casagrande and VandenBerg (2018). This interstellar extinction shall only be considered as a lower limit to the total extinction. This procedure for generating HR diagrams is standard throughout the chapter, so all the HR diagrams presented can be directly compared.

3.2.1 PMS object category

Although for the algorithm there is just a single class of PMS objects, I create that class by combining intermediate-mass Herbig Ae/Be stars and lower mass T Tauris, in order to cover the whole mass range.

Herbig Ae/Be stars

Regarding the Herbig Ae/Be stars, I start with the compilation of Chapter 2 where most known HAeBes could be matched with Gaia DR2 data. The main issue with

Herbig Ae/Be stars is that almost all of them are brighter than the bright limit of IPHAS and VPHAS+ (12 – 13 mag). Using Hα equivalent widths (EWs) I derived the IPHAS and VPHAS+ like colour $r - H\alpha$ using the synthetic tracks of Drew et al. (2005, see their Fig. 6, extinctions and effective temperatures are present in Chapter 2 and references therein). Combining the Hα EWs of Chapter 2 and Wichittanakom et al. (2020) with the few sources present in IPHAS or VPHAS+ provided the $r - H\alpha$ colour for 215 HAeBes. This is why neither r nor Hα passbands are combined with the rest in Sect. 3.1.1, as they are not available for many sources. There is a bias in this conversion from Hα EWs to $r - H\alpha$ colour because it can only be applied to those objects with observed Hα in emission above the continuum. Hence, it could not be applied to the many HAeBes with intrinsic emission filling in the underlying absorption but below the continuum level. This bias also appears later for T Tauri stars and CBes in Sects. 3.2.1 and 3.2.2 and its impact is addressed in Sect. 3.5.2.

The cross-match with AllWISE to obtain 2MASS and WISE passbands was already performed in Chapter 2. The final number of Herbig Ae/Be stars considered is 255, of which 163 have all observables. I did not include Massive Young Stellar Objects (Lumsden et al., 2013) in this sample as in general they are not optically visible so they are not present in Gaia DR2 (except those that are already in Chapter 2 list which were also included in this chapter).

T Tauri stars

To the set of intermediate-mass Herbig Ae/Be stars I add a set of T Tauri stars to complete the low-mass regime. If I use those objects catalogued as T Tauris in the SIMBAD database (around 3500 objects at the time of writing) I end up, after the cross-matches, with most of the objects having being catalogued by a few papers dedicated to very specific regions (e.g. Venuti et al., 2014 on NGC 2264 open

cluster or Sicilia-Aguilar et al., 2013 on Tr 37). In order to minimise the possible implications due to this I add the sources of the Herbig-Bell (HB) Catalogue (Herbig and Bell, 1988) which, although focused in the Orion region, has sources distributed all over the sky. I cross-matched the set of T Tauris with Gaia DR2 with a 0.5 arcsecond aperture (close to the 0.4 arcsecond angular resolution of Gaia DR2). I double checked that the cross-matched sources have a similar V and G band (within $\pm$2 mag, the range is rather generous to avoid biasing to exclude very variable sources) when possible to discard bad cross-matches. Then, I cross-matched the Gaia source identifications with those of the SoSt (see Sect. 3.1.4) to obtain the T Tauri stars with all the observables.

In addition, the HB catalogue provides me with Hα EWs and spectral types that allow me to derive $r - H\alpha$ colour for 297 more T Tauris. To this end, I used the HB B-V colour, which come from simultaneous passbands at maximum brightness, and the spectral types provided by the HB catalogue to derive extinctions for these T Tauris. Whenever B-V colours were not available I used those of the APASS survey (Henden et al., 2018) with a 3 arcsecond cross-match. A small error is introduced for objects colder than roughly a G2 V star which are typically given slightly smaller $r - H\alpha$ magnitudes than those that correspond to them (see Drew et al., 2005 for further details). The overall result is a sample of 3171 T Tauri stars, of which 685 have information in all the observables.

3.2.2 Classical Be stars

For the classical Be stars, I use the Be Star Spectra Database (BeSS Database, Neiner et al., 2011) which comprises 2264 CBes. This includes the candidates of Raddi et al. (2013, 2015). To these I add 35 more CBes from Shokry et al. (2018, those they claim as secure detections). I cross-matched that catalogue with Gaia DR2 using a 0.5 arcsecond aperture. Again, I double checked that the cross-matched sources

have a similar V and G band photometry (within ±2 mag) when possible, in order to discard bad cross-matches. Then, I cross-matched the Gaia source identifications with those in the SoSt (Sect. 3.1.4) to assess how many CBe stars are there with all the observables.

In order to increase the number of stars in this category, I complemented it with Hα EWs from the spectra available in the BeSS database. I estimated an uncertainty measuring EWs of 15%, which is probably within the intrinsic EW variations of these objects. Then, I used again the synthetic tracks of Drew et al. (2005) to transform Hα EWs to IPHAS and VPHAS+ $r - H\alpha$ colour for 442 sources. In order to do this, I used the spectral types of the BeSS database to estimate effective temperatures and, if undetermined, I estimated them from the positions in the HR diagram (Fig. 3.3). I assumed no extinction, which is roughly safe for this kind of object (only the faint ones from Raddi et al., 2015 suffer significantly from interstellar extinction). To assess whether this is a valid assumption I studied the extinction in the G band provided by Gaia DR2 for all the CBe stars for which it is available. If I take the central values I found that 94% of the sources have an A_G lower than 1.55, which is roughly the value beyond where the extinction becomes significant for the colour conversion of Drew et al. (2005). The final number of classical Be stars considered is 1992 of which 775 have information in all the observables.

3.2.3 Disentangling Herbig Ae/Be, CBe stars, and B[e] stars

There is some inevitable contamination between categories. For example, the set of known PMS objects is contaminated in its massive end by classical Be stars and vice-versa. Indeed, there were 15 sources that appeared both as PMS and CBe star in the previous selections. Therefore, I needed to take decisions on how to catalogue them, even though in many cases there is no clear answer in the literature. These

objects were not excluded as they are the most interesting ones for the algorithm to learn from.

In the following I specify which classification decision was made regarding the sources that appear both as Herbig Ae/Be and classical Be in this section. Not all of them have all the observables.

- BD+41 3731 - Classical Be - It appears as a Herbig Ae/Be star in Alecian et al. (2013) and Reiter et al. (2018). However, Labadie-Bartz et al. (2017) consider it a classical Be star and Cauley and Johns-Krull (2014) suggest not to treat it as a PMS object and so did I.

- GU CMA - Herbig Ae/Be - It is generally considered as a Herbig Ae/Be star (e.g. Ababakr et al., 2017; Fairlamb et al., 2015; Reiter et al., 2018; Mathew et al., 2018).

- HBC 7 - Herbig Ae/Be - It is a bit doubtful but Hernández et al. (2004) argue that it shows charecteristics of PMS objects.

- HD 114981 - Classical Be - It appears as Herbig Ae/Be in many papers (Reiter et al., 2018; Fairlamb et al., 2015) but as CBe in Labadie-Bartz et al. (2017). Cauley and Johns-Krull (2014) found evidence for it to be a CBe star.

- HD 130437 - Classical Be - Although it appears in The et al. (1994) as a Herbig Ae/Be star, the situation is very unclear. I decided to follow the intuition of Acke and van den Ancker (2006).

- HD 158643 - Herbig Ae/Be - Shokry et al. (2018)

- HD 174571 - Herbig Ae/Be - It displays a doubtful nature in many papers (Reiter et al., 2018; Ababakr et al., 2017; Cauley and Johns-Krull, 2014; Labadie-Bartz et al., 2017) but there is a general consensus that it is a Herbig Ae/Be star.

- HD 36408 - Herbig Ae/Be - Donehew and Brittain (2011) and Cauley and Johns-Krull (2014)
- HD 37490 - Classical Be - Cauley and Johns-Krull (2014) and Cochetti et al. (2019)
- HD 50083 - Herbig Ae/Be - Reiter et al. (2018), Alecian et al. (2013), Sartori et al. (2010), Wheelwright et al. (2010), and Cauley and Johns-Krull (2014)
- HD 76534 - Herbig Ae/Be - Patel et al. (2017)
- HD 94509 - Herbig Ae/Be - Fairlamb et al. (2015)
- LkHA 350 - Herbig Ae/Be - Hernández et al. (2004)
- MWC 655 - Herbig Ae/Be - Ababakr et al. (2017) and Wheelwright et al. (2010)
- V1493 Cyg - Herbig Ae/Be - It has been little studied in the recent years but appears as a Herbig Ae/Be star in The et al. (1994) and in a few papers since then (e.g. Mathew et al., 2018) although Hernández et al. (2004) was unable to classify it.

In addition, within the sets of known sources there were many 'unclassified B[e]' stars (also known as FS CMa stars, Miroshnichenko, 2017; Arias et al., 2018). FS CMa objects are an inhomogeneous group of B stars with forbidden lines and a very unclear nature. These forbidden lines and the dust-type infrared excess exclude them from being PMS or CBe sources (Rivinius et al., 2013) and I removed them from the sets of known objects in order to not bias the results. As a word of caution, independently, around half of the HAeBes display the B[e] phenomenon (see Oudmaijer, 2017).

Almost all the confirmed FS CMa objects are listed in Lamers et al. (1998), Miroshnichenko (2007), Miroshnichenko et al. (2007), Miroshnichenko et al. (2017),

and Khokhlov et al. (2018) and they add up to 53 objects (around 70 proposed in total, Miroshnichenko and Zharikov, 2015). A total of 17 FS CMa stars from this list were discarded from the sets of known PMS and CBe stars: BD+23 3183, CD-24 5721, CD-49 3441, AS 119, HD 328990, HD 45677, HD 50138, HD 85567, Hen 3-847, LkHA 348, MWC 1055, MWC 342, MWC 657, PDS 021, PDS 211, V2211 Cyg, and V669 Cep.

Separately, but related, in Chapter 2 it was found that because of their positions on the HR diagram: MWC 314, MWC 623, and MWC 930 were not very likely to be PMS objects. Indeed, MWC 314 seems to be a supergiant B[e] star (Frasca et al., 2016), MWC 930 looks like a luminous blue variable (Martayan et al., 2016; Jiang et al., 2018) and MWC 623 seems clear to be a FS CMa star (Miroshnichenko, 2007; Polster et al., 2018). Therefore, I also removed these three objects from the set of known HAeBes (Sect.3.2.1).

3.2.4 Other objects

I construct the category of other objects by randomly sampling sources from the Sample of Study. I would like to have a representative set of whatever else might be present in the SoSt that is not a PMS object or a classical Be star. The question is how large this category should be in order for the algorithm to generalise properly. In other words, I want to know how many random sources from the SoSt are necessary so all populations present in the cross-matched catalogues have been represented in this category of other objects.

This can be estimated by training an ANN with different sizes of this third category, and studying how well it generalises in each case. The size of the previously described categories is kept constant. Using an ANN (3 fully connected hidden layers of 300 neurons each) I evaluate how the precision and recall of the network on the PMS group behave on a test set (sized 20% of the training set) for different sizes of

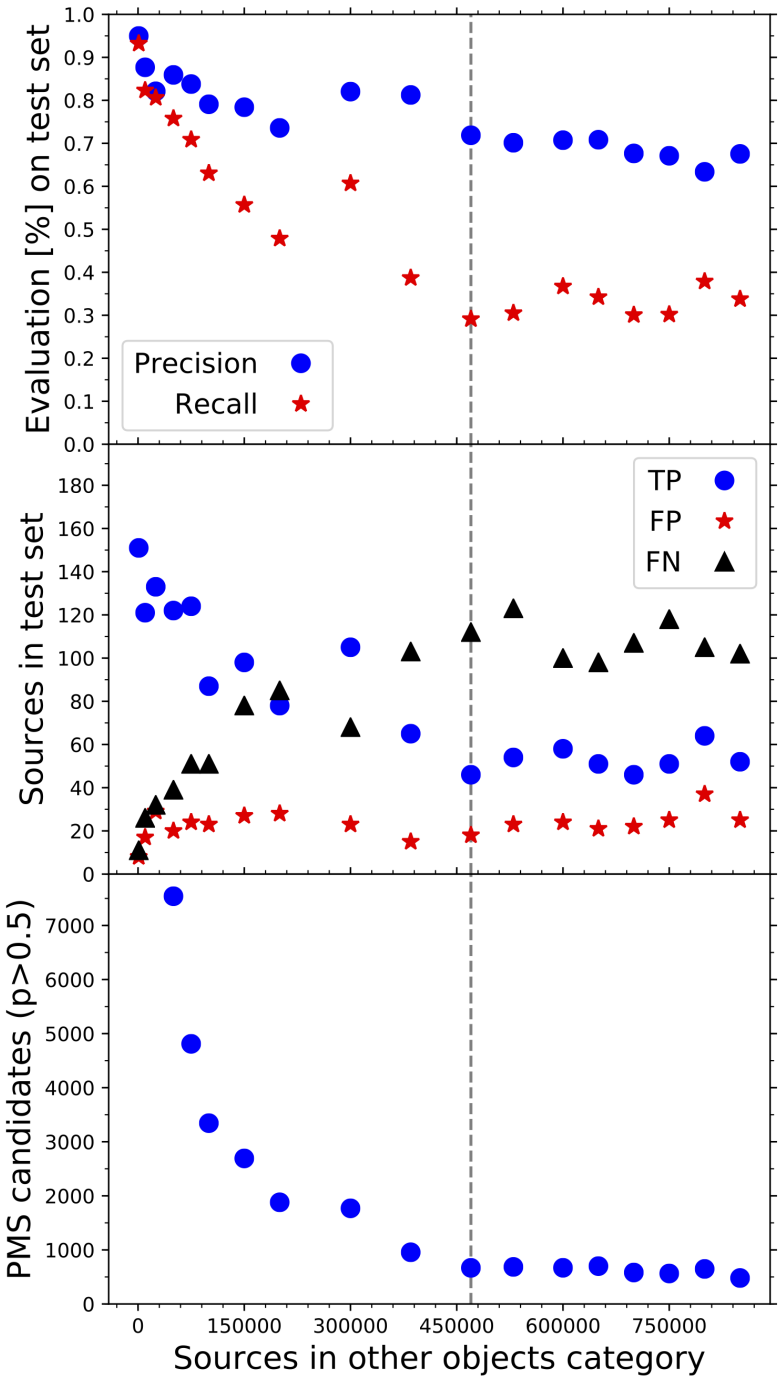


Fig. 3.4 Different metrics of the ANN on the PMS category vs. different sizes of the other objects category. Stabilisation point is marked with a vertical grey line. *Top:* Precision and recall. As the size of the category gets larger the recall drops drastically up to the stabilisation point whereas the precision is roughly stable at all sizes. *Middle:* TP, FP, and FN. Similarly, TP and FN have equal stabilisation point whereas FP is stable for all sizes. *Bottom:* Number of PMS candidates obtained when generalizing the trained ANN, I note the same stabilisation point.

this other objects category (Fig. 3.4). The architecture of this ANN is a bit more sophisticated than the complexity demanded by the problem (as can be seen by the chosen architecture in Sect. 3.3.2), but I wanted to be sure to not underfit in any case so the ANN is always sensitive to new data. If the category of other objects is very small the algorithm is very precise and has a high recall (Fig. 3.4 on top); few other objects appear in the regions of the feature space where PMS and CBe stars tend to be placed, so they have little impact in the classification (it also indicates that the ANN is good in telling the difference between PMS and CBe stars, although I note that a large fraction of the PMS category are low-mass T Tauri stars). The more I populate the feature space with other objects the algorithm is less able to recognise PMS stars (the false negatives rise, Fig. 3.4 at middle) as the regions of the feature space with the more common PMS sources start to be highly populated by objects similar to PMS stars and undiscovered PMS objects. The number of false positives stays the same as the algorithm is still being efficient in the less populated regions. In other words, the PMS candidate region in the feature space gets smaller and localised around the less common PMS sources. The number of true positives drops as a consequence of the increase of false negatives. This causes the precision to barely change (Eq. 3.3) but the recall to drop (Eq. 3.4, Fig. 3.4 on top) up to a stabilisation point (grey line in Fig. 3.4) where most of the different types of objects that populate the feature space differently have appeared, and hence adding more sources does not further constrain significantly the locus of PMS candidates in the feature space.

This stabilisation point can also be found if I study the number of PMS candidates retrieved after generalizing the trained ANN to the unlabelled sources of the SoSt (Fig. 3.4 at bottom, selecting as candidates those with a probability $p \geq 50\%$ of belonging to the PMS category). This number drops quickly from a very high value, where the algorithm does not know about the existence of anything but PMS stars and CBes, up to the stabilisation point where it becomes roughly stable. Of course,

adding more other objects always diminish the region of PMS and CBe candidates in the feature space, but this stabilisation point constitutes the optimal size for the category of other objects, as larger sizes do not compensate the amount of extra information for the contamination they introduce.

Therefore, for constructing the category of other objects, I randomly sample sources from the SoSt (excluding the sources in the PMS and CBe categories) so that they are in a proportion of 99.82% with the number of sources in the category of PMS objects (848), this being the observed stabilisation point. This scales to 470,263 objects. Some of these sources might have been classified previously by different catalogues, although most remain unclassified.

I can approximate what is the proportion of other objects to PMS objects in the SoSt from a simulation. Robin et al. (2012), using the Gaia Universe Model Snapshot (GUMS) simulation estimated that the percentage of PMS objects within $G < 20$ mag in Gaia is 0.18%. The real proportion of PMS sources in the SoSt is somewhat larger as I am demanding detections up to 22 μm and in Hα. This implies, theoretically, that roughly there are as many undiscovered PMS stars in the other objects category as known PMS stars in the PMS category.

3.3 Algorithm and methodology

The pipeline used can be seen in Fig. 3.5. Most of the algorithm described in it is available on a GitHub repository[2] under the name of YODA (Young Object Discoverer Algorithm). In this section I describe this pipeline in detail.

3.3.1 Class weights

As explained in Sect. 3.2.4, the set of other objects is contaminated by undiscovered PMS and CBe stars. This causes the ANN used in Sect 3.2.4 to obtain a decent

[2] https://github.com/MVioque/YODA

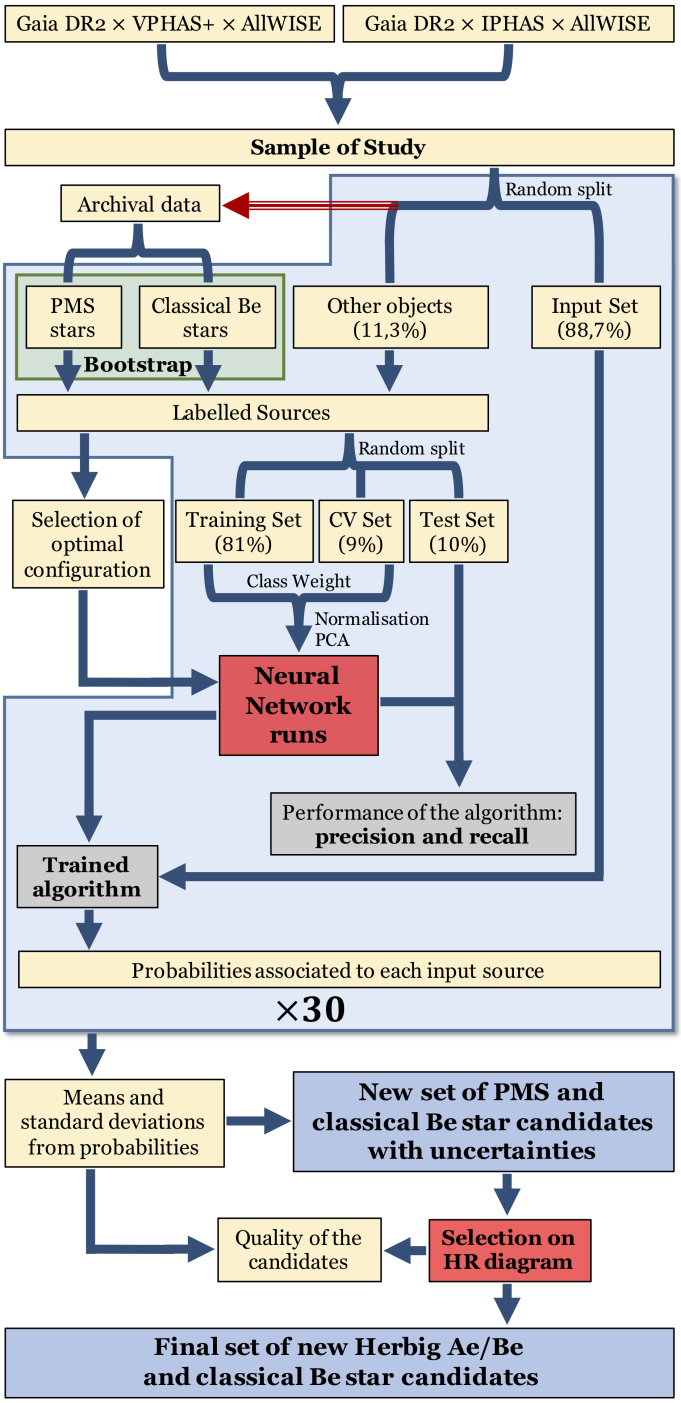


Fig. 3.5 Pipeline of the whole algorithm, from the cross-match of Gaia DR2, AllWISE, IPHAS, and VPHAS+ to the set of new HAeBes and classical Be stars. The light blue area indicates the set of processes that are repeated in a loop 30 times, each time generating a different set of probabilities associated to each input source. The green area shows the bootstrapped sets. The red arrow indicates that the Archival data is partially contained within the Sample of Study, but was also constructed using external information like the Hα EWs.

precision but a terrible recall (see Fig. 3.4). As discussed in Sect. 3.2.4, this is because I am only retrieving the less common sources. Any pre-classifications performed with the observables used by the algorithm would artificially bias the results. One option is to use simulations to generate this well defined category of other objects without PMS and CBe stars (as done for example by Castro-Ginard et al., 2018) but there is none that lists PMS and CBe objects and contains IR and Hα information.

I can address this issue by changing the weights of the sources used in the training, in a way that they are balanced for the different category sizes. Hence, during training the ANN is heavily penalized when failing at categorising PMS or CBe sources, but lightly affected by mistakes on the other objects category, which is much larger. Therefore, the training is not dominated by contaminants or undiscovered sources, although they still are considered. This weighting technique produces a decent recall and a very low precision, but this precision is just a lower limit as the candidate regions of the feature space contain many undiscovered PMS and CBe stars. In other words, FP is over-measured (Eq. 3.3). In addition, these class weights stress the algorithm to focus on the differences between PMS sources and CBe stars, which also bias the selection of PMS sources towards the high-mass end.

3.3.2 Architecture selection

Different machine learning algorithms were considered for this classification problem. A variety of them have been used so far for similar matters. For example: random forests (Hedges et al., 2018; Marton et al., 2019; Rimoldini et al., 2019), support vector machines (Małek et al., 2013; Marton et al., 2016; Solarz et al., 2017; Ksoll et al., 2018), or artificial neural networks (Snider et al., 2001; Hampton et al., 2017). Akras et al. (2019b) used the classification tree, linear discriminant analysis, and K-nearest neighbour algorithms to distinguishing Hα emitters with infrared information

in photometric surveys. However, similar performances are achieved with most of the algorithms and it is evident that the output is mainly dominated by the quality of the training data (Pérez-Ortiz et al., 2017; Pashchenko et al., 2018, or Marton et al., 2019 in a similar problem of identifying Young-Stellar Objects). Therefore, I decided to use a shallow artificial neural network as it has the advantage of flexibility and non-linearity, being able to describe very complex and subtle relations. In addition, its output can be a probability vector, which eases the catalogue construction. Cons are the number of hyper-parameters required, which are normally hard to interpret.

Therefore, I needed to find the architecture or optimal configuration of the ANN for my particular problem. This means choosing the hyper-parameters of the ANN (e.g. layers, neurons per layer, regularization, see Appendix B). Ideally, this architecture would be selected with cross-validation (CV) sets that are not used for the training. However, the sample of known PMS and CBe stars is too small to have the number of CV sets necessary to test a large enough grid of ANN configurations. Instead, I used the set of Labelled Sources to select the optimal hyper-parameters and then, independently, used those hyper-parameters and that same set of Labelled Sources for the training (see Fig. 3.5). I ran an ANN over the Labelled Sources set 100 times (each time with a 10% test set random split); evaluating at each training iteration on a CV set (sized 10%) and early-stopping whenever the precision of the algorithm on the PMS category (selecting as candidates those with a probability $p \geq 50\%$ of belonging to such category) stopped increasing over 250 iterations. In addition, I imposed that the recall had to be at least 90%. In each run a different grid of hyper-parameters was used. After the 100 runs, the best architecture resulted in two fully connected hidden layers of 580 neurons, with a dropout rate of 50% and 0.01 L2 regularization. Batch Normalisation was applied after every layer, though no batches were used and the whole training data was evaluated in each training iteration (as it is a very skewed training set, see Sect. 3.2).

The activation functions used were 'ReLU' for the hidden layers and 'softmax' for the output layer. This is because softmax output can be interpreted as a probability distribution. The loss function used was 'categorical crossentropy' with the 'AdaMax' optimiser (Kingma and Ba, 2014). See Appendix B for a brief introduction to these concepts. To construct the ANNs of this project I used Keras (Chollet et al., 2015), a high-level neural networks application programming interface.

3.3.3 Training, cross-validation, and test set

I shuffle the Labelled Sources set and randomly split it into two subsets (see Fig. 3.5). One contains 90% of the sources and is used to train the algorithm (training set). The other, containing 10% of the sources is used to evaluate its performance (test set).

The first step is to perform feature scaling and mean normalisation to the observables, so they all have the same mean and standard deviation. Then I apply PCA to the scaled observables to get the set of features used by the ANN (12 principal components of the 48 carry 99.99% of the variance, see Sect. 3.1). Next, I train the ANN, which has the architecture chosen in Sect. 3.3.2, with the training set and use a CV set (sized 10% of the training set) to evaluate the ANN performance after every training iteration. Early-stopping finishes the training whenever the precision on the PMS category (with $p \geq 50\%$) stops increasing over 50 iterations. I note that, as discussed in Sect. 3.3.1, the precision retrieved is just a lower limit. Once the ANN is trained, I run it on the test set, that needs to be scaled and feature extracted as done for the training set. Evaluation on test set gives a value of precision and recall for each probability threshold for classification p (i.e. the performance of the algorithm, see Fig. 3.6). Finally, I can apply the trained ANN to the Input Set, giving a probability for every source of belonging to each of the chosen categories.

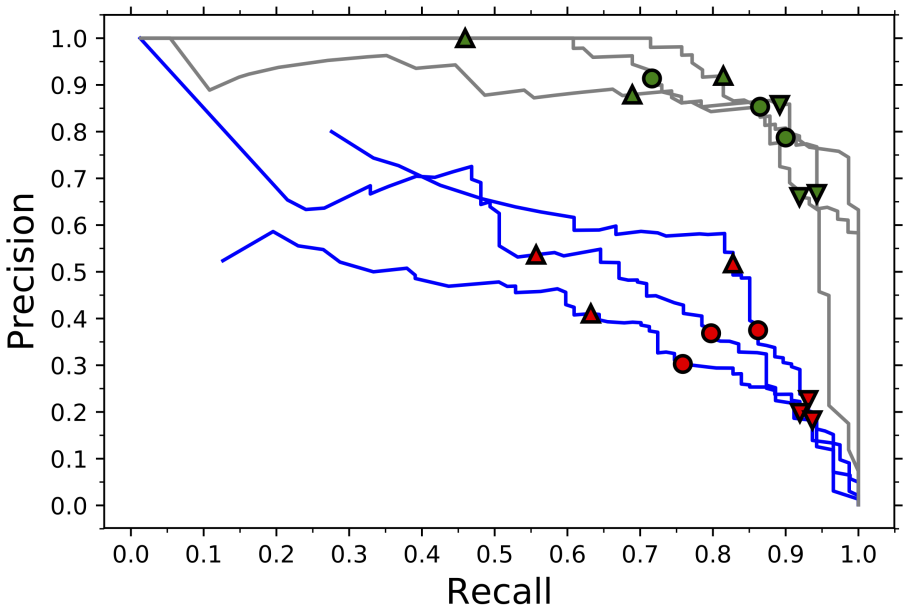


Fig. 3.6 Precision vs. recall trade-off plot resulting after evaluation on test set for three different bootstrapped iterations. Blue lines correspond to PMS classification and grey lines to classical Be classification. Different probability thresholds (p) for selecting candidate objects correspond to different locations on the line. Upwards arrows ($p \geq 75\%$), circles ($p \geq 50\%$), and downwards arrows ($p \geq 25\%$) are examples of such probability thresholds. Some lines do not cover the whole metric space because evaluation stops whenever there are no longer true positives in the corresponding test set. The precision values are lower limits to the real precisions.

3.3.4 Bootstrap

A major issue is the small size of the PMS and classical Be categories in the training data (see Sect. 3.2). Small training sets imply that outliers and contaminants have a very strong influence and might dominate the posterior generalisation. In addition, the training might be biased to any hidden trend or pattern.

One way to minimise the impact of this is by means of the bootstrap. The key idea is to fake the construction of new training sets. It works by repeatedly sampling the original training data and randomly substituting sources with others of the same data set. If the same algorithm is run over two bootstrapped sets similar, but slightly different metrics are obtained as a result. If I repeat this bootstrapping process a large enough number of times I end up with a distribution of precisions and recalls characteristic of my method, which allows me to estimate the uncertainty of the

metrics for each probability threshold. Bootstrapping has another advantage, which is to better represent the distribution of the categories on the feature space and minimise the impact of outliers.

Therefore, I run the processes described in Sect. 3.3.3 (blue area of Fig. 3.5) 30 times in a loop. In each iteration, I create a bootstrapped version of the combination of the categories of known PMS and classical Be stars (so the number of objects in each group is not conserved). In the case of the other objects category, I just withdraw another random set of sources from the Sample of Study. Once the algorithm is trained with a certain Labelled Sources bootstrapped set, I obtain by evaluating on the corresponding test set values for the precision and recall at different probability thresholds (see Fig. 3.6). When I run the trained ANN over the Input Set, I retrieve probabilities associated to every source of belonging to each of the three categories. Hence, after the bootstrapped iterations I end up with 30 values for precision and recall at different thresholds (in Fig. 3.6 only three bootstrapped iterations are shown for clarity) and 30 sets of probabilities associated to each source of the Sample of Study. This is because as the category of other objects has been randomly sampled 30 times, the whole SoSt has been covered eventually. I can average those 30 repetitions and take the standard deviation of the mean as the uncertainty of each measurement. This gives a solid estimate of the final precisions and recalls for different p thresholds as well as final probabilities and uncertainties for all the sources in the SoSt.

Summarising, the set of Labelled Sources described Sect. 3.2 and the Sample of Study of Sect. 3.1.4 are introduced in a pipeline of algorithms which core is an ANN. As an output, the pipeline gives probabilities and associated uncertainties for all the sources in the SoSt of belonging to either the PMS, classical Be, or other objects category (excluding the known PMS and CBe sources included in the categories of the Labelled Sources set). These probabilities, that sum up to one in each source, are presented in Fig. 3.7. The uncertainties for the PMS category are shown in Fig. 3.8.

3.4 Results

The Sample of Study with probabilities is available in electronic form in its entirety (4,150,983 sources, see Sect. 3.6). This data is made available so the user can choose their own probability threshold (p) to select PMS and classical Be candidates. Choosing p implies fixing a precision (P) and a recall (R). The pipeline also gives a solid estimate of the precisions and recalls for different p thresholds. However, due to the nature of the pipeline the values for the precision are only lower limits (see Sect. 3.3.1). Ideally these two metrics should be as high as possible but there is a trade-off between them. This is shown in Fig. 3.6, where the precision and recall for both the PMS and CBe category are plotted for different p probability thresholds. Raising the threshold to $p \geq 99\%$ maximises the precision to almost 1, but as a consequence the recall lowers to almost 0. The opposite also applies and neither of both extremes is close to be representative of a good selection; as it would be either largely incomplete or largely imprecise. The general shape of the curves is determined by the architecture of the algorithm and the peculiarities of the classification problem (see Sect. 3.3.2).

In practice, using probability thresholds below 50% is possible, but entering the regime where the algorithm assigns larger probabilities to other categories is not advisable as p does not correlate linearly with the precision and recall (see Fig. 3.6). At $p \geq 50\%$ the resulting catalogues are: new PMS candidates (8,470 sources, $P = 40.7 \pm 1.5\%$, $R = 78.8 \pm 1.4\%$), new classical Be candidates (693 sources, $P = 88.6 \pm 1.1\%$, $R = 85.5 \pm 1.2\%$). I note that the precisions are lower limits. These catalogues of new candidates are presented in Sect. 3.6 in Tables 3.3 and 3.4 respectively and highlighted in Fig. 3.7 (full tables available in electronic form, see Sect. 3.6)[3]. In those tables, together with the probabilities (and associated

[3] As a word of caution, these recalls do not imply that the presented catalogues contain $\sim 80\%$ of the existing PMS and CBe stars within any region. They imply that $\sim 80\%$ of the known PMS and CBe stars in the test set are recovered by the algorithm. This is for example affected by what the different surveys used are probing and the distribution of the SoSt (see Sect. 3.1.4 and Fig. 3.2). As

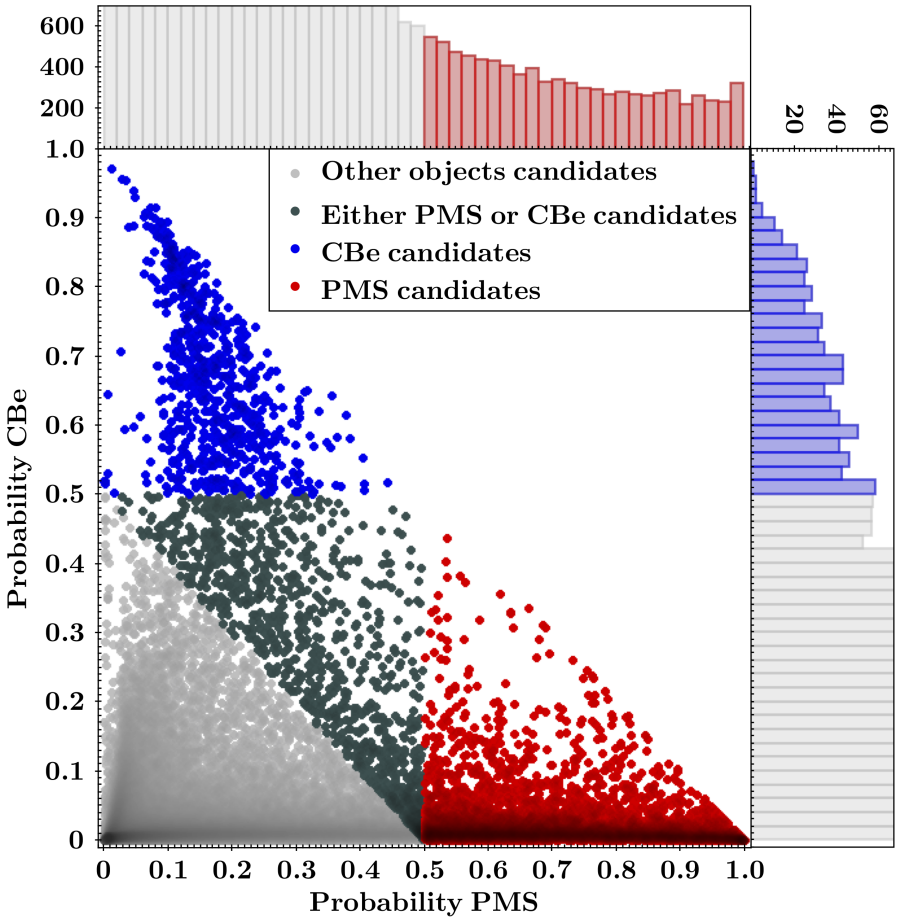


Fig. 3.7 Output probability map of the Sample of Study. A probability threshold of $p \geq 50\%$ is used to select the PMS (in red) and classical Be candidates (in blue). On top and right number histograms of the candidates for different probabilities. In dark grey the sources which belong to either category $(p(PMS) + p(CBe) \geq 50\%$ but $p(PMS) < 50\%$, $p(CBe) < 50\%)$. Uncertainties are not indicated for clarity. Numbers are: PMS candidates (8,470), classical Be candidates (693), either (1,309), other objects (4,140,511).

uncertainties, see Fig. 3.8) I present the observables used for the training (Sect. 3.1.1) and Gaia astrometric information. In addition, I included the derived interstellar extinction (A'_G) and A'_G corrected M_G (Gaia G absolute magnitude) and $G_{BP} - G_{RP}$ for those sources with RUWE<1.4 and $\varpi/\sigma(\varpi) \geq 5$. These allow for a better positioning in the HR diagram (see Sect. 3.2 and Figs. 3.3, 3.9, and 3.13).

explained in Sect. 3.5, probably some of the less extreme objects in the observables used have not been classified. Similar reasoning can be applied to the precision values.

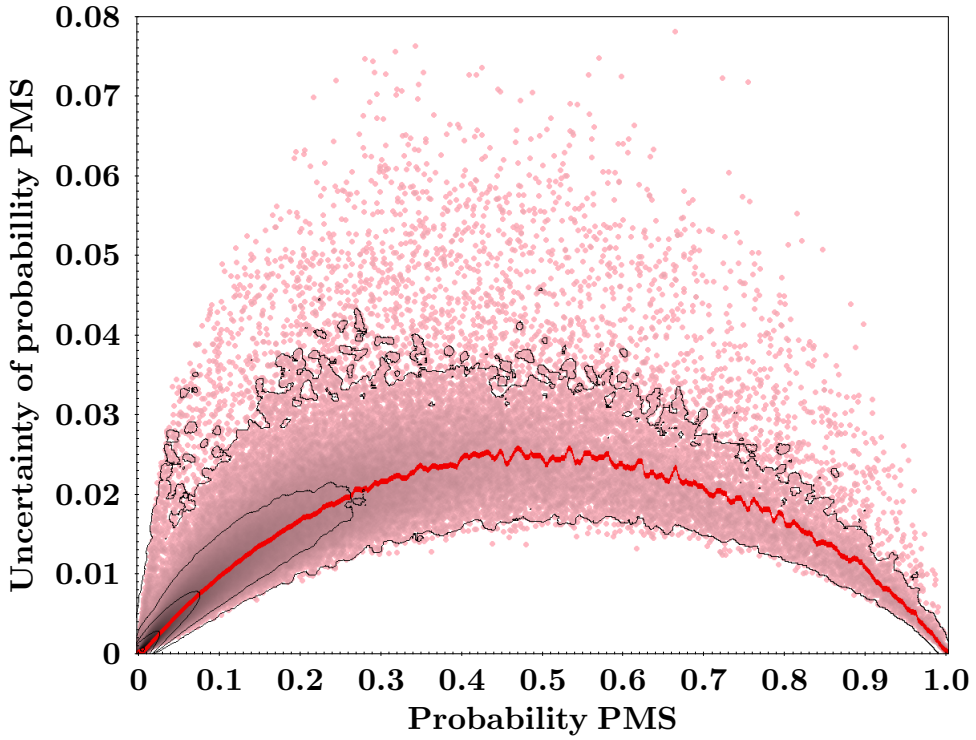


Fig. 3.8 PMS probabilities of all Sample of Study sources vs. the uncertainty in that probability given by bootstrap. Logarithmic contours trace number density and the red line traces the median uncertainty value for each probability. Naturally, as the uncertainty is the standard deviation of the mean of 30 bootstrapped iterations (see Sect. 3.3.4), central values tend to have larger uncertainties than extreme values. This does not imply that this distribution is random.

In the CBe case the precision does not drop drastically (see Fig. 3.6). This implies that for the algorithm it is easier to find CBe stars than PMS stars as their locus in the feature space is less prone to contaminants but mostly because there are fewer unclassified CBes in the SoSt (see Sect. 3.3.1). A consequence of this is that I retrieve an order of magnitude less CBe candidates than PMS candidates.

Following the discussion of Sect. 3.2.4, the size of the other objects category roughly coincides with the point where there is approximately one undiscovered PMS source per known PMS source in the training and test sets. Taking this into account, the lower limit on the precision of $P \sim 40\%$ for the PMS group obtained with $p \geq 50\%$ is an adequate enough result (i.e. the real precision is roughly double). However, as the precision is a lower limit, it is hard to assess whether a higher

probability threshold is better to retrieve a stronger catalogue of PMS candidates. In order to decide this, I need to use parameters and observables that have not been used in the training, and are hence independent of the selection. As explained in Sect. 3.1, the set of features (and hence the classification) is distance and position independent, at least at first order. This means that I can use the HR diagram and the sky locations to assess this issue.

Before analyzing these catalogues, I first remove the sources brighter than the typical bright limit of IPHAS and VPHAS+ that show significant differences between their IPHAS or VPHAS+ magnitudes and their Gaia magnitudes (marked in Tables 3.3 and 3.4 with a 'X-mtch' flag). These objects did not affect the training as they barely account for 0.5% of the other objects category. There are 18 PMS and 57 CBe candidates with this flag. These sources are likely to be incorrect cross-matches and they are left out in the following analyses.

3.4.1 Evaluation using the HR diagram

The HR diagram is not entirely selection independent, as I used different colours in the classification and I do not correct for the unknown intrinsic extinctions. However, the location in the HR diagram, which carries information about evolutionary status, is almost independent of the classification.

The Gaia HR diagram of the PMS and classical Be candidates (those with $p >$ 50%) can be seen in Fig. 3.9 at left panels. In Fig. 3.9 at right panels I also distinguish those with a good astrometric solution (RUWE<1.4 and $\varpi/\sigma(\varpi) \geq 10$) to which interstellar extinction corrections have been applied. These well behaved sources have trustworthy positions in the HR diagram. The candidates have been colour-coded according to their membership probability to the corresponding category. In this HR diagram I can evaluate the quality of the retrieved catalogues. Regarding the PMS candidates, the majority are placed to the right of the main sequence, as

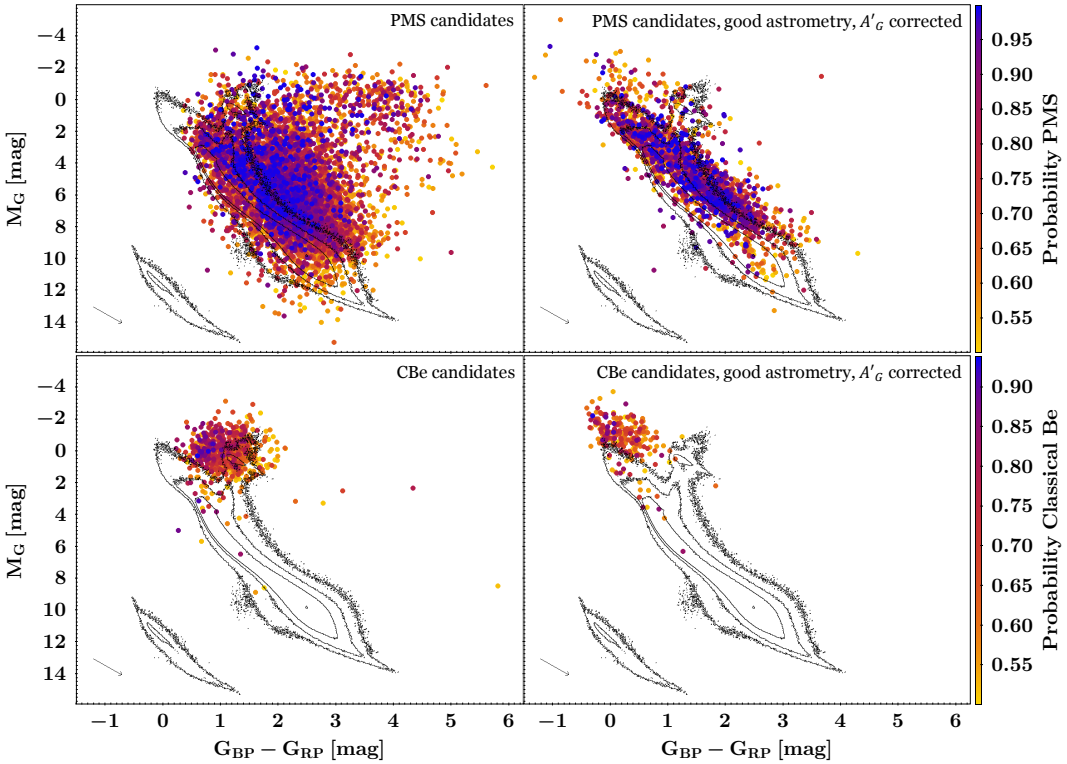


Fig. 3.9 Gaia colour vs. absolute magnitude HR diagram. An extinction vector corresponding to $A_G = 1$ is shown on the bottom left of each plot. Black contours trace Gaia sources within 500 pc with good astrometric solution. *Top left:* HR diagram of PMS candidates ($p \geq 50\%$) colour-coded by their associated membership probability. *Top right:* HR diagram of PMS candidates ($p \geq 50\%$) with good astrometric solution colour-coded by their associated membership probability and corrected from interstellar extinction. *Bottom left:* HR diagram of classical Be candidates ($p \geq 50\%$) colour-coded by their associated membership probability. *Bottom right:* HR diagram of classical Be candidates ($p \geq 50\%$) with good astrometric solution colour-coded by their associated membership probability and corrected from interstellar extinction.

expected for PMS sources. Moreover, if higher probability thresholds or only those sources with a good astrometric solution are used the selection of PMS candidates is constrained to those sources that are located in the more likely PMS positions. Something similar happens with the CBe candidates, as they are placed where CBe stars are supposed to be. This can be better appreciated when comparing these candidates with the locus of the known PMS and classical Be stars in Fig. 3.3. I note that $\sim 6\%$ of PMS candidates and $\sim 1\%$ of CBe candidates lack parallax information.

This, in addition of evidencing the quality of the selection, allows to select a higher probability threshold by looking at the retrieved candidates in the HR diagram of astrometrically well behaved sources corrected from interstellar extinction (Fig. 3.9 at top right). This threshold can be adapted to the requirements of different studies or situations. Here, I stick to the probability threshold of $p \geq 50\%$ for constructing the catalogues of candidates. This is because the candidates with the higher probabilities are the easier ones to find. Hence, as can be seen in Fig. 3.9, most of the high-mass PMS candidates do not have high associated probabilities as the algorithm struggles more to differentiate them from classical Be stars and vice-versa. In addition, Fig. 3.7 shows that there are very few CBe candidates with a negligible PMS probability. Therefore, a more conservative selection of the probability threshold would exclude many of the high-mass objects (see histograms of Fig. 3.7).

Finally, although in the rest of the paper I discuss the catalogues as of $p \geq 50\%$, the user can construct their own catalogues by means of Tables 3.3, 3.4, and the SoSt with probabilities. As any new catalogue is likely to have a higher probability threshold, the discussion and analysis that follows holds true. Hence, from here onward I refer to PMS and CBe candidates as those with a $p > 50\%$ in their respective categories.

3.4.2 Evaluation using sky locations

The selection has been totally independent of coordinates. This, though true, is limited by the combined footprint of the surveys used; for example, it is limited to the Galactic plane, see Sect. 3.1.4 and Fig. 3.2. Now, in Fig. 3.10 at top, I plot the contours of the catalogue of new PMS candidates on the Sample of Study footprint. The PMS candidates trace some of the overdensities of the SoSt. This is because any random selection of sources traces the footprint of the SoSt but might also be because

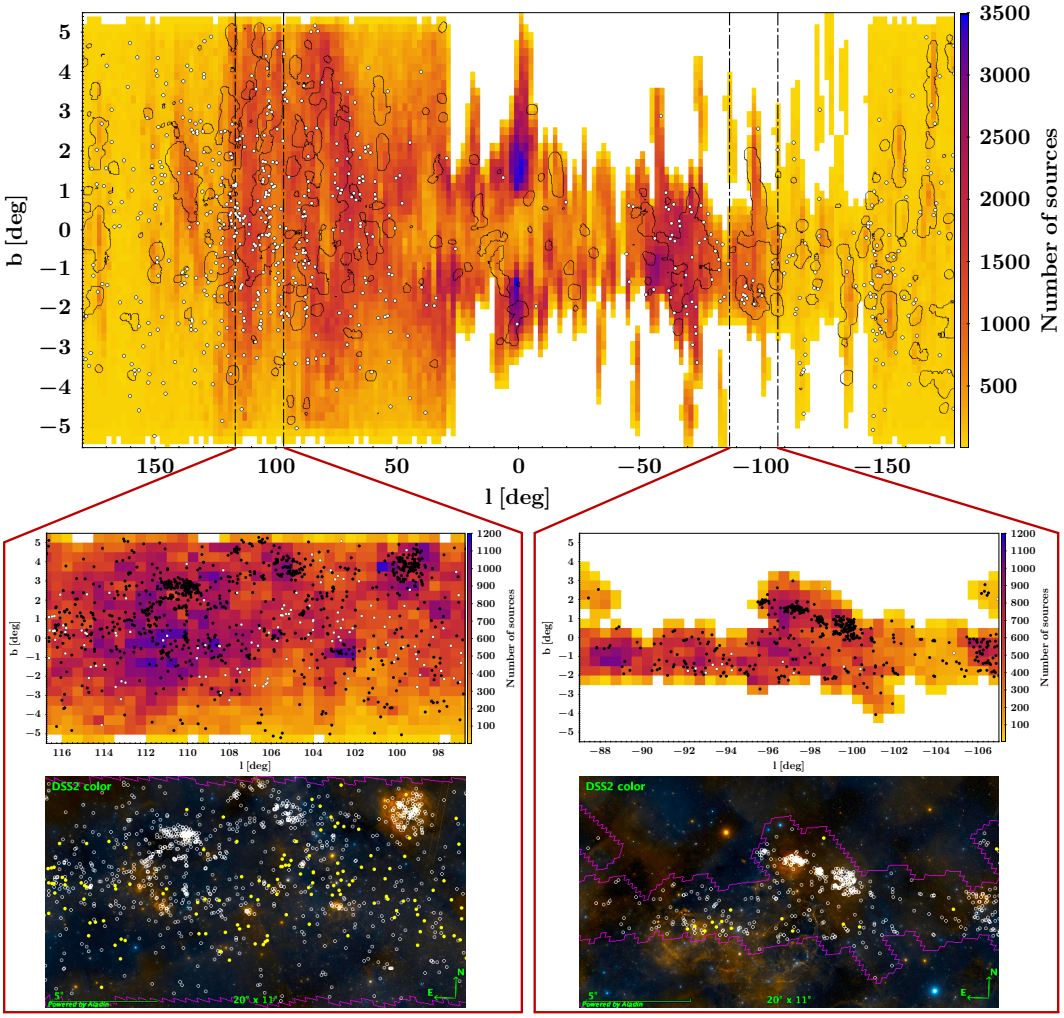


Fig. 3.10 *Top:* Sky footprint of the Sample of Study in galactic coordinates, colour-coded by number density. I note the heterogeneity of the footprint. The scarcity of sources between $29^\circ > l > -145^\circ$ is due to the incompleteness of VPHAS+ at the time of writing. Each pixel is $2^\circ \times 0.2^\circ$. PMS candidates overdensities appear as black contours. There are ten time more candidates inside than outside the contours. CBe candidates appear as white dots. Expanded regions at bottom panels appear between dashed lines. *Middle:* Expanded regions. Each region is $20^\circ \times 11^\circ$. PMS contours are replaced by PMS candidates (black dots). Each pixel is $0.5^\circ \times 0.5^\circ$. *Bottom:* Same expanded regions in DSS2 colour with the PMS candidates as white circles and the CBe candidates as yellow dots. Contours trace the footprint of the Sample of Study.

the new PMS candidates are mostly appearing in star forming regions, which would be a strong assessment of the selection.

In this respect, it is noticeable that some overdensities are not particularly populated by PMS candidates. Moreover, when looking at the small scale (examples

in Fig. 3.10 at middle and bottom panels) it can be seen that the PMS candidates are not strictly following the SoSt overdensities but are more likely associated to nebulosities. In addition, in Fig. 3.11 I plot all the PMS candidates in the sky. They appear distributed all over the Galactic plane but there are associations of candidates, regions of around ~ 0.5 to 1 squared degrees where there are ten to a hundred more PMS candidates than normally distributed. These associations also appear if I include the distances (Fig. 3.12).

This means that the Gaia coordinates are assessing the efficiency of the algorithm, as the retrieved PMS candidates are prone to appear around nebulosities and star forming regions, even though these regions are not over-represented in the input data. Further evidence in this respect can be seen in Fig. 3.11, where the new classical Be candidates are also plotted in the sky. These candidates are distributed all over the Galactic plane but they are not tracing the associations of PMS candidates or nebulosities (see e.g. Fig. 3.10 at bottom panels), which implies that they are indeed of an independent nature for the algorithm. Moreover, if I include the distances (Fig. 3.12), CBe candidates also appear decoupled from the PMS candidates. CBe candidates are typically further away, something expected from bright B-type stars.

Although the clustering properties of this new set of PMS sources is beyond the scope of this study (see Chapter 4 for a more detailed insight), I can make some remarks. Firstly, on a global scale, PMS candidates trace some of the regions with more data available. This is likely because these zones contain star forming regions, as not all the regions with more data available are overpopulated with PMS candidates (Fig. 3.10 at top panel). However, on a local scale, PMS candidates do not trace overdensities in the available data, and the associations of candidates appearing are normally not related to those overdensities (e.g. Fig. 3.10 at middle panels). Secondly, the PMS candidates seem to trace nebulosities, and the large sky associations obtained are mostly related to them (e.g. Fig. 3.10 at bottom panels). There are also a few smaller associations of PMS candidates unrelated to

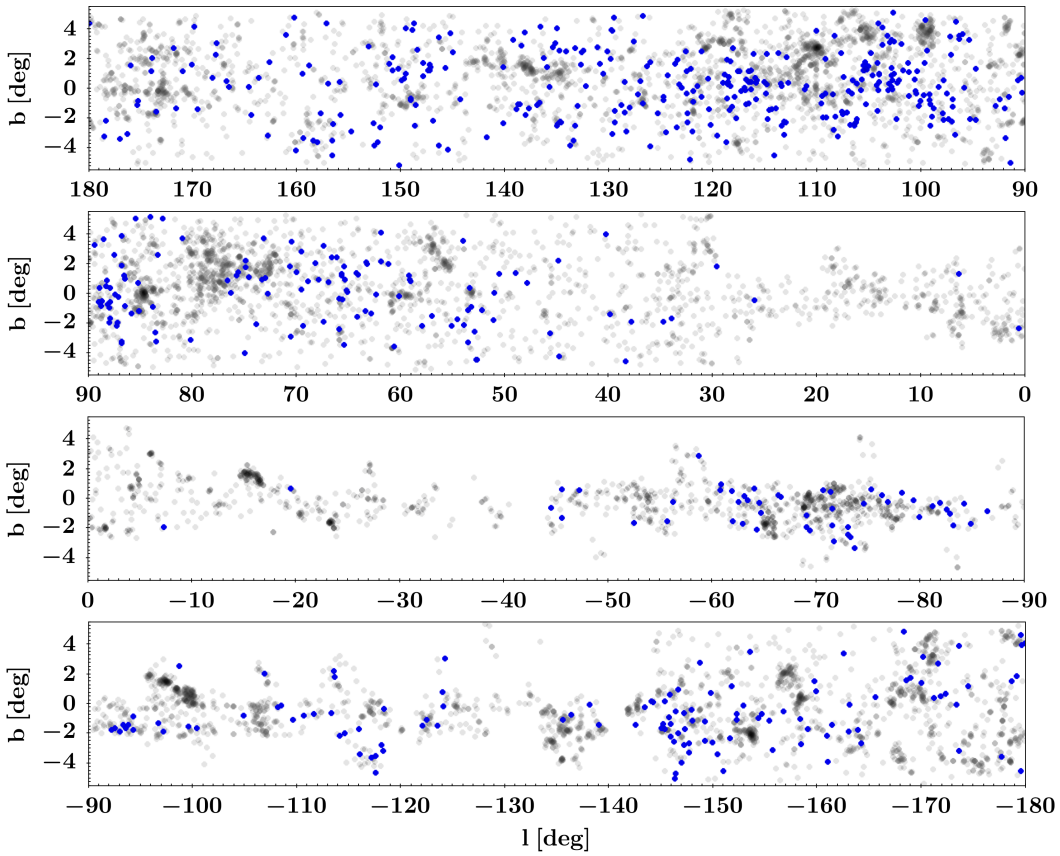


Fig. 3.11 Classical Be candidates in blue distributed in the sky in galactic coordinates plotted on top of the PMS candidates in light grey. The densest regions of PMS candidates appear darker. The scarcity of sources between $29^\circ > l > -145^\circ$ is due to the incompleteness of VPHAS+ at the time of writing.

footprint overdensities that seem to trace dark nebula and are placed on their edges. Lastly, among the PMS candidates there is no significant correlation between PMS probability ($0.5 \geq p \geq 1$, see Fig. 3.7) and coordinates.

3.4.3 Herbig Ae/Be candidates

I have constructed a new catalogue of PMS candidates, of which some can be plotted accurately in the HR diagram (Fig. 3.9 at top right), a selection-independent plot. Therefore, I can further select the Herbig Ae/Be stars. In order to do this I study where the known HAeBes (Sect. 3.2.1) are placed in the HR diagram using the same quality constrains and extinction corrections. This is done in Fig. 3.13. There, I

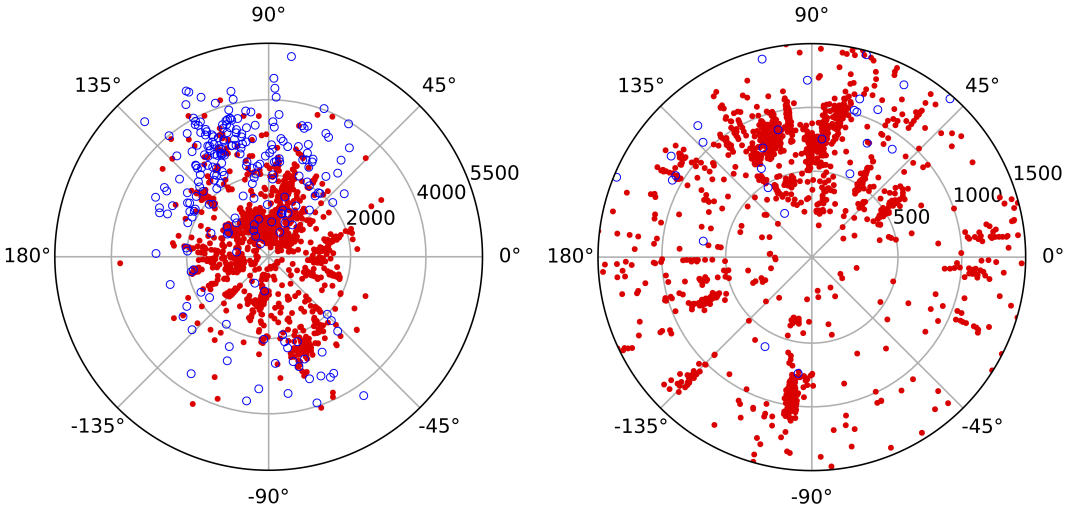


Fig. 3.12 Galactic longitude vs. distance (in parsec [pc], from Bailer-Jones et al., 2018) of PMS (red dots) and classical Be (blue circles) candidates with good astrometric solution. *Left:* All candidates, *Right:* Candidates up to 1500 pc.

estimate that PMS candidates with absolute magnitude $M_G < 6$ are possible HAeBe candidates. This is taking into account that the intrinsic extinction, typically large for these objects, has not been considered (most of these sources do not have measured spectral types). This way, I retrieve 1361 new Herbig Ae/Be candidates which are marked in Table 3.3 (end of the pipeline, Fig. 3.5). This constitutes an improvement of one order of magnitude with respect to the $\sim$ 272 previously known HAeBes (see Chapter 2). The new Herbig Ae/Be candidates are shown in Fig. 3.13.

By construction, these HAeBe candidates are astrometrically well behaved sources (RUWE<1.4 and $\varpi/\sigma(\varpi) \geq 10$). Hence, I expect many more HAeBes among the PMS candidates, as most of them do not satisfy these conditions ($\sim$ 60%, see Fig. 3.9 at top left) or do not even have parallax information ($\sim$ 6%). For example, using a more relaxed $\varpi/\sigma(\varpi) \geq 5$ parallax constraint gives 2226 HAeBe candidates, but their distance and interstellar extinction uncertainties do not allow for separating them as nicely from the low-mass candidates. In contrast, using less precise parallaxes allow to retrieve candidates at farther distances and typically more massive. As the list of all PMS candidates is available in Table 3.3, future studies

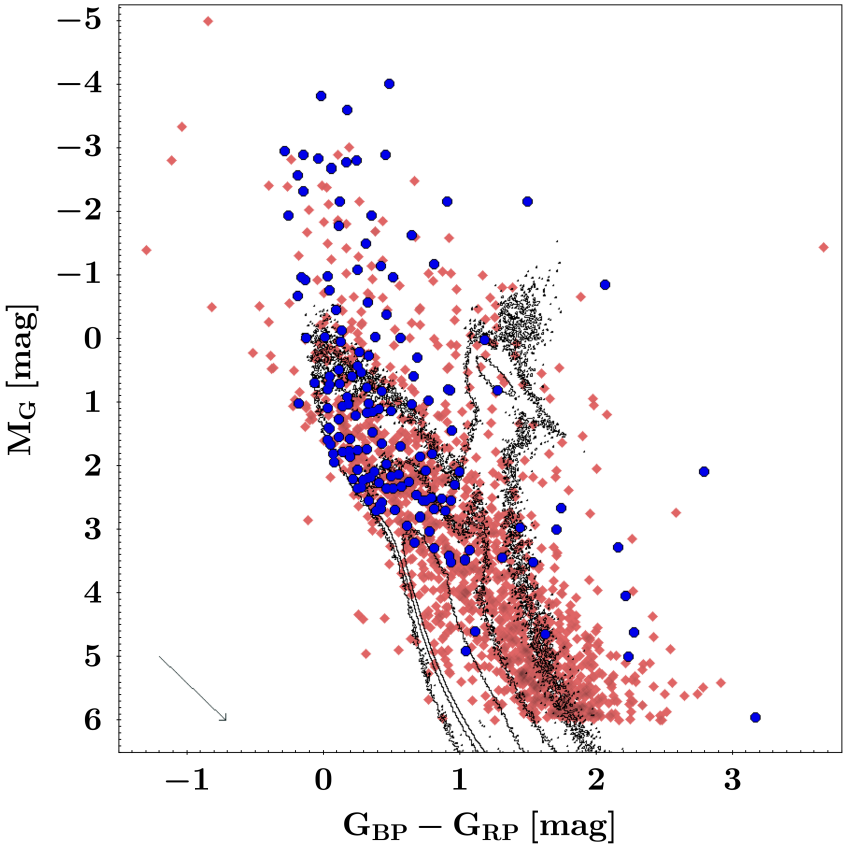


Fig. 3.13 Gaia colour vs. absolute magnitude HR diagram. Blue dots are previously known Herbig Ae/Be stars with good astrometric solution corrected from interstellar extinction. Red diamonds are new Herbig Ae/Be candidates corrected from interstellar extinction. An extinction vector corresponding to $A_G = 1$ is shown on the bottom left. Black contours trace Gaia sources within 500 pc with good astrometric solution.

may want to use less conservative thresholds to the astrometric quality and select their own set of Herbig Ae/Be candidates.

3.4.4 Variable candidates

As discussed in Chapters 1 and 2, UX Ori type objects (UXORs) are sources with irregular brightness variations from 2-3 magnitudes in the optical. Observed light gets bluer in the deep minima, and the fraction of polarised light increases. Many of them are catalogued as HAeBes and their extreme variability is explained by eclipsing dust clouds in nearly edge-on sources and scattering radiation in the circumstellar environment (Natta et al., 1997; Grinin, 2000; Natta and Whitney, 2000; Oudmaijer

et al., 2001). In Chapter 2 I provided strong support to the edge-on disc explanation using Hα line profiles of known HAeBes. In addition, using a similar variability indicator to G_{var} (Eq. 3.1), I found that all catalogued UXORs with V band detected variabilities above 0.5 mag are strongly variable (17 objects). This implies that these indicators effectively trace irregular photometric variability.

By using a variability threshold, in Chapter 2 I proposed 31 new UX Ori candidates among the previously known HAeBes. The equivalent G_{var} threshold is $G_{var} \geq 10$. PMS candidates with $G_{var} \geq 10$ are marked in Table 3.3 with the flag of 'Var' (3436 sources) and for the HAeBe candidates the UXOR phenomenon is the most likely explanation. As I am tracing variability by using the Gaia DR2 uncertainties, sources without intrinsically irregular photometry like binaries or extended objects can pop out as strongly variable.

This means that $\sim 41\%$ of the new PMS candidates are of this variability type. This proportion increases to $\sim 49\%$ when it comes to the HAeBe candidates. This number is consistent with the UXOR behaviour caused by an inclination effect (50% predicted by Natta and Whitney, 2000). Very probable PMS candidates are in general very variable (see Sects. 3.5.1 and 3.5.2), so most of the best PMS candidates appear as 'Var' in Table 3.3.

Another assessment of the variability proxies can be achieved by cross-matching the PMS candidates with variability surveys. A 5 arcsecond cross-match with ASAS-SN (Jayasinghe et al., 2019) gives 949 sources, of which 830 (87%) have $G_{var} \geq 10$. In addition, 557/949 (59%) appear as of Young-Stellar Object (YSO) variability type. A 5 arcsecond cross-match with the Zwicky Transient Facility (ZTF, Masci et al., 2019) gives 6438 sources, including 95% of those with $G_{var} \geq 10$ in the sky region covered by the survey. A 5 arcsecond cross-match with ATLAS-VAR (Heinze et al., 2018) gives 2216 objects. Of these, 1960 (88%) have variabilities which are hard to classify by the ATLAS-VAR machine classifier, suggesting that they are likely of an irregular type, similar to those of PMS sources. Finally, if I cross-match

my results with the catalogue of long-period variable candidates of Mowlavi et al. (2018), which also contains a small set of YSO candidates, I obtain 491 matches with my set of PMS candidates, of which 444 (90%) have $G_{var} \geq 10$. According to Mowlavi et al. (2018) classification, 297/491 (60%) are YSO candidates but 190/491 (39%) are long-period variable candidates (4/491 are undetermined because they lack parallaxes). These possible contaminants are addressed in Sect. 3.5.2.

3.4.5 Comparison with Marton et al. (2019) and other catalogues and surveys

Marton et al. (2019) did a similar study to the one presented here but looking for YSOs in general and only using Gaia DR2, 2MASS, WISE colours and passbands, and the optical depth from the *Planck* dust opacity map. Therefore, they did not use Hα or variability information and did use distance-dependent features. In addition, they restricted the search to high dust opacity regions. They found 1,768,628 potential new YSO candidates (with the recommended $p \geq 0.9$) using a random forest algorithm. In contrast with that study, in this chapter I have focused on identifying new high-mass pre-main sequence objects, and on differentiating them from classical Be stars.

Giving the differences between the two approaches in terms of considered sources, training data, and features; it should not be surprising if there are not many objects in common between the two studies and yet both are highly accurate. However, I find that 48% of my PMS candidates are within the Marton et al. (2019) catalogue. Moreover, this percentage slightly increases at higher probability thresholds of my catalogue (56% at $p \geq 0.95$). Regarding the Herbig Ae/Be candidates (see Sect. 3.4.3), 56% of them are present among the YSOs of Marton et al. (2019). In contrast, only 11% of my catalogue of classical Be stars appear as YSOs in Marton et al. (2019). When moving to $p \geq 0.85$ this number goes down to 0%.

This is a good assessment of the quality of the categorisation as an independent study, using a different algorithm and training data, has achieved relatively similar results regarding PMS sources (taking into account the differences between methodologies) but has not found almost any of the new CBe candidates presented in this chapter (and none of the best CBe candidates). This, in addition to support my selection, proves that the new HAeBe candidates are nicely separated from the population of CBe stars. The differences between the two studies probably lie in that I am using Hα and variability information and that Marton et al. (2019) searched only in dusty environments, being this way position-dependent. In addition, I demand detections up to W4 (22 µm), whereas these authors only demand detections up to W2 (4.6 µm). Further assessment is that, as in Marton et al. (2019), I find that 62 of my PMS candidates are within the Gaia Photometric Science Alerts published at the time of writing (a project that looks for transient events in the Gaia data, Delgado et al., 2019); 13 of them appear as YSO, 47 as unknown, and only two appear as non-PMS. Conversely, of the 87 YSOs in the Alerts, 18 are in the SoSt, which means that I only missed five that were classified as 'other source'.

Similarly, of the PMS candidates in SIMBAD (2607 within 1 arcsecond cross-match at the time of writing) 974 ($\sim$ 37%) appear catalogued as PMS or PMS candidate. There are 18 objects appearing as CBe, but these were mostly catalogued by Mathew et al. (2008) and Gkouvelis et al. (2016). These papers selected CBes using simple cuts in IPHAS+, 2MASS, or WISE observables and hence I understand that the analysis presented in this chapter supersedes theirs. In addition to this, 663 sources ($\sim$ 25%) appear as with emission lines, infrared bright, or variable. Only 356 sources ($\sim$ 14%) appear as clearly non-PMS. This includes 101 AGB candidates and 16 carbon star candidates that are addressed in Sect. 3.5.2. As explained in that section, I expect this number of 356 PMS candidates classified as non-PMS by other studies to be considerably lower, so this cross-match with SIMBAD is consistent with the estimated precision in Sect. 3.4 of $P \gtrsim 80\%$ for the catalogue of PMS

candidates. The other 596/2607 sources do not have a defined category in SIMBAD. VES 263, the new Herbig Ae/Be star discovered by Munari et al. (2019) is not within the Sample of Study.

Of the classical Be candidates in SIMBAD (280 within 1 arcsecond cross-match at the time of writing) 17 appear as CBe (again, most from Mathew et al., 2008 and Gkouvelis et al., 2016) and 197 as with emission lines. Only nine are clearly not CBes, of which four are of PMS nature and three appear as variable. This reinforces the idea that the algorithm is efficiently separating PMS sources from classical Be stars. The other 57/280 sources do not have a defined category in SIMBAD.

Finally, using a cross-match aperture of 20 arcsecond I find 26 matches between the set of PMS candidates and the Gaia-ESO Public Spectroscopic Survey (Gilmore et al., 2012). A fraction of 24/26 sources have hydrogen lines in emission: 14/26 show double-peaked emission (although two might be considered P-Cygni), 6/26 single-peaked emission, 3/26 are either single-peaked or double-peaked, and one shows a clear inverse P-Cygni profile. Only 2/26 spectra have Hα line in absorption. The line profile fractions agree with those studied in Chapter 2 for known HAeBes (31% single-peaked, 52% double-peaked, and 17% P-Cygni). This gives independent spectroscopic evidence for the PMS nature of the new PMS candidates.

3.4.6 Visualisation

It is possible to visualise the feature space and the selection using a dimensionality reduction algorithm. I used the UMAP algorithm (Uniform Manifold Approximation and Projection for Dimension Reduction, McInnes et al., 2018, a brief introduction to the algorithm is presented in Appendix B) to project the 12-dimensional feature space (Sect. 3.1.2) into two dimensions. This is done in Fig. 3.14. In the left panel I project the candidates using an euclidean metric (15% of the number of sources as number of neighbours and a minimum distance of 0.4). In the right panel I project

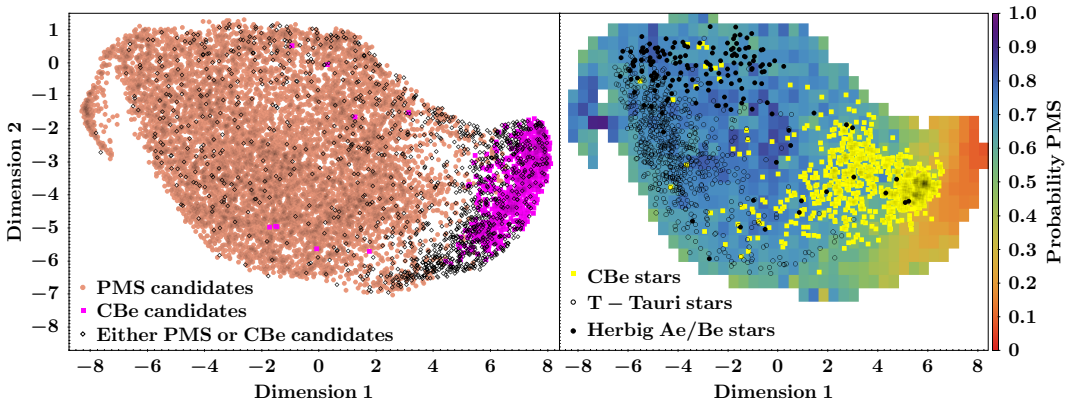


Fig. 3.14 UMAP dimensionality reduction from the 12D space of features to 2D. *Left:* dimensionality reduction of the PMS and classical Be candidates, together with those sources which belong to either category (i.e. $p(PMS)+p(CBe) \geq 50\%$ but $p(PMS) < 50\%$, $p(CBe) < 50\%$). *Right:* I project the known Herbig Ae/Be, T Tauri, and classical Be stars used for the training onto the same plane, which is colour-coded following the PMS probability distribution of the sources at left.

the known HAeBe, T Tauri, and CBe stars used for the training (Sect. 3.2) onto this same plane, which is colour-coded following the PMS probability distribution of the sources at left. This dimensionality reduction helps to understand the catalogue construction and to find trends within the data. However, information is lost when moving to 2D. The category of other objects was not included here because of size limitations.

First, it can be noticed that there is indeed a separation between known PMS sources and classical Be stars (Fig. 3.14 at right), which is used by the algorithm to learn how to separate these populations (Fig. 3.14 at left). It is remarkable that most of the retrieved CBe candidates are not found where known CBe stars are, but even farther away from the PMS region, which might imply that I am retrieving very extreme CBe candidates (see Sect. 3.5.2). In addition, most of the PMS candidates that are located very close to the CBe region are those with a high CBe probability and vice-versa (see Fig. 3.7). These PMS candidates have $r-H\alpha$ and $G_{BP}-G_{RP}$ values typical of known CBes and low IR-excesses for PMS objects. However, these

are still typically larger than the ones of known CBes. This, together with strong G_{var} variability explains their selection as PMS.

Second, in Fig. 3.14 an arm structure at the very top left of the space of PMS candidates seems special. A closer look at the 204 PMS candidates located in that arm shows that they are all placed in the red giant region of the HR diagram (see Fig. 3.9 at top left). They differentiate from the rest of the PMS candidates in that they have more variability in the indicators and are typically brighter. In addition, they show on average larger near-IR excesses but lower mid-IR excesses. None of them have reliable astrometry so they do not contaminate the sample of Herbig Ae/Be candidates. Potentially, they are evolved contaminants and are flagged in Table 3.3 as 'G_{UMAP}' (see Sect. 3.5.2 for further details of their nature). I cannot exclude sources only because of their HR diagram position, as I might remove HAeBes with high extinctions and some candidates do not have parallax information. Hence, candidates in similar red giant HR diagram positions but not in this UMAP region are not flagged in the final catalogues of Tables 3.3 and 3.4.

3.5 Quality assessment

Table 3.1 summarises the final number of sources in the resulting catalogues of PMS and CBe candidates. Table 3.1 also indicates the number of known sources considered in Sect. 3.2, of which those having all observables were used for the set of Labelled Sources. In this section I evaluate the classification from different perspectives and give insights on the relative importance of the different observables used for the selection. In addition, I discuss any detected bias or flaw in the final catalogues of PMS and classical Be candidates. In general, these mostly affect sources with a bad astrometric solution in Gaia DR2 so they do not implicate the catalogue of new Herbig Ae/Be candidates (Sect. 3.4.3).

Table 3.1 Summary of the number of known sources of each type considered together with those included in the set of Labelled Sources because of having all observables. The last column indicates the number of new candidates of each type classified by the algorithm.

	Considered sources	Labelled Sources set	Classified $p \geq 0.5$
Herbig Ae/Be	255	163	-
T Tauri	3171	685	-
PMS	3426	848	8470
Classical Be	1992	775	693

Note: To be considered, I demanded the sources to be present in Gaia DR2. For the algorithm there is just a single class of PMS objects, which is constructed by combining known Herbig Ae/Be and T Tauri stars. Herbig Ae/Be and T Tauri candidates can be selected from the set of PMS candidates using the HR diagram. In Sect. 3.4.3 1361 Herbig Ae/Be candidates were obtained this way.

3.5.1 Classification on the test sets

One way to analyse the classification is to study the evaluation on the test sets. As described in Sect. 3.3.4, I evaluated the performance of the ANN in 30 different test sets. As the selection of the test set is random in every iteration, almost all of the known PMS and classical Be sources were in the test set at some point. Averaging those 30 evaluations results in 793 PMS and 733 CBe known sources that have been independently assessed by the algorithm.

Regarding the classification of known PMS sources, the most noteworthy trend is that very variable PMS stars in either indicator (G_{var} and V_{htg}) are identified. Although those known PMS stars with $r - H\alpha > 1.3$ are identified, objects with $0 < r - H\alpha < 1.3$ are spread over the whole range of probabilities. Thus, $r - H\alpha$ does not seem to play an important role in detecting PMS sources (see Sect. 3.5.4). This also happens with $G_{BP} - G_{RP}$. However, in these two cases, known PMS sources with low $r - H\alpha$ or bluer are those who tend to be given high CBe probabilities. The known PMS sources that were not identified were mostly stars with low near-IR excess ($J - K_s$), which are also the ones that are given high CBe probabilities. This

is probably because these are more similar to CBe stars. Surprisingly, I miss many known PMS sources with high mid-IR excess ($W1-W4$) and those that had very low $W1-W4$ values were mostly not identified, which again are the ones with higher CBe probabilities. In general, very few known PMS sources are assigned to the CBe category, although many known PMS sources are not classified as such (algorithm recall on the PMS group is $R = 78.8 \pm 1.4\%$, Sect. 3.4.1).

Regarding the known classical Be sources, the algorithm also identifies the very variable ones correctly. CBe sources with high $r-H\alpha$ are normally given high PMS probabilities but in general they are not misclassified. There is no trend between $r-H\alpha$ values and CBe assigned probabilities. In contrast, there is a trend with $G_{BP}-G_{RP}$ and redder objects are less likely to be classified as CBe and are given higher PMS probabilities, but are rarely misclassified as PMS. In addition, CBe sources show no CBe probability trend with $J-K_s$ or $W1-W4$ although sources with $W1-W4 \gtrsim 3$ are normally not classified as CBe. Similarly, CBe sources with higher near- and mid-IR excesses are given higher PMS probabilities but are infrequently assigned to the PMS category.

Evaluation on the test sets indicates that the algorithm effectively identifies sources of different categories and uses the various observables to trace the main characteristics of PMS and classical Be stars.

3.5.2 Final catalogues assessment

In the following points I discuss a few biases and flaws detected in the final catalogues of PMS and CBe candidates:

1. I demand to have detections up to W4 (22 μm) and in the Hα passbands. Although I am training with sources that span the whole range of values in these observables, this induces some biases as I am excluding in the training many of the less extreme sources and hence biasing the posterior selection.

This is aggravated given that the Hα EW to $r-H\alpha$ colour conversion of Sect. 3.2 can only be applied to sources with observed emission above the continuum. This effect can be quantified if I compare the output catalogues with all the known sources gathered in Sect. 3.2 (see Table 3.1). This way, the mean value of $r-H\alpha$ for known PMS (classical Be) sources (using one standard deviation as error) is $r-H\alpha = 0.74 \pm 0.36$ mag (0.38 ± 0.18 mag) and for the candidates is $r-H\alpha = 0.87 \pm 0.46$ mag (0.63 ± 0.20 mag). The mean value of $W1-W4$ for known PMS (classical Be) sources is $W1-W4 = 4.0 \pm 2.2$ mag (1.7 ± 1.3 mag) and for the candidates is $W1-W4 = 5.2 \pm 1.4$ mag (2.24 ± 0.71 mag). In short, the retrieved candidates are the more extreme of their kind in terms of Hα emission and IR excess (specially mid-IR excess). This particularly affects the catalogue of CBe candidates, as these have typically less extreme values. In Fig. 3.15 I present the frequency density distribution of the final catalogues of PMS and CBe candidates for a subsection of key observables ($G_{BP}-G_{RP}$, $J-K_s$, $W1-W4$, $r-H\alpha$, G_{var}, and V_{htg}) together with the distribution of all known sources.

2. As mentioned in Sect. 3.1.4, WISE presents many spurious photometric detections in the Galactic plane. To investigate this, Koenig and Leisawitz (2014) used a set of AllWISE quality parameters and additional selection criteria to determine that only $\sim 28\%$ of the sources in their study have reliable W3 and W4 detections. Marton et al. (2019), using a different approach, concluded that only 10% of their set have reliable W3 and W4 photometry. These authors used very stringent criteria for the sake of purity and these percentages may be slightly pessimistic.

   I decided to use these passbands because of their expected importance in separating CBes from PMS sources (see Sect. 3.5.4). A more relaxed constraint, using the extended source flag of AllWISE distinct to 0 gives 44% and 27% of

badly behaved PMS and CBe candidates respectively (marked in Tables 3.3 and 3.4 with a 'W3W4' flag). I note that, in contrast to Marton et al. (2019), I am using many observables in addition to W3 and W4 so the algorithm can deal better with these being spurious.

3. As described in Sect. 3.1.4, because of the cross-match, I estimated that 1/42 (1/625) sources of the Sample of Study on average are fake, in the sense that their associated IR (Hα) photometry do not belong to them. However, only 17 (6) PMS and no CBe candidates appear with duplicated IR (Hα) information. The sources that have the AllWISE, or IPHAS or VPHAS+ name repeated in the SoSt are marked in Tables 3.3 and 3.4 with the 'ID AllW' or 'ID IPH/VPH' flag respectively.

4. There are 104 SIMBAD AGB stars in the catalogue of PMS candidates (only three appear as of confirmed AGB nature and the rest appear as candidates). This is because they were all classified in one single paper (Robitaille et al., 2008), where they attempted to separate YSO from AGB stars using a simple colour and magnitude selection criteria in the near- and mid-IR. I understand that these are contaminants in that work as my analysis supersedes theirs.

5. I have detected a high incidence of planetary nebulae (PN) detected as PMS candidates. Observational similarities between YSOs, B[e] stars, and PN have been reported before (e.g. Frew and Parker, 2010; Boissay et al., 2012; Miroshnichenko and Zharikov, 2015 and references therein). In fact, Akras et al. (2019a) found new PN objects by using machine learning techniques and a similar set of observables to the ones used in this chapter. This is mainly caused because PN show high $r - H\alpha$ colours. In addition, as they are extended they present large Gaia uncertainties, so they can appear as highly photometrically variable in the indicators (Eqs. 3.1 and 3.2). Of the PMS candidates in SIMBAD (Sect. 3.4.5), there are 57 ($\sim 3.5\%$) catalogued

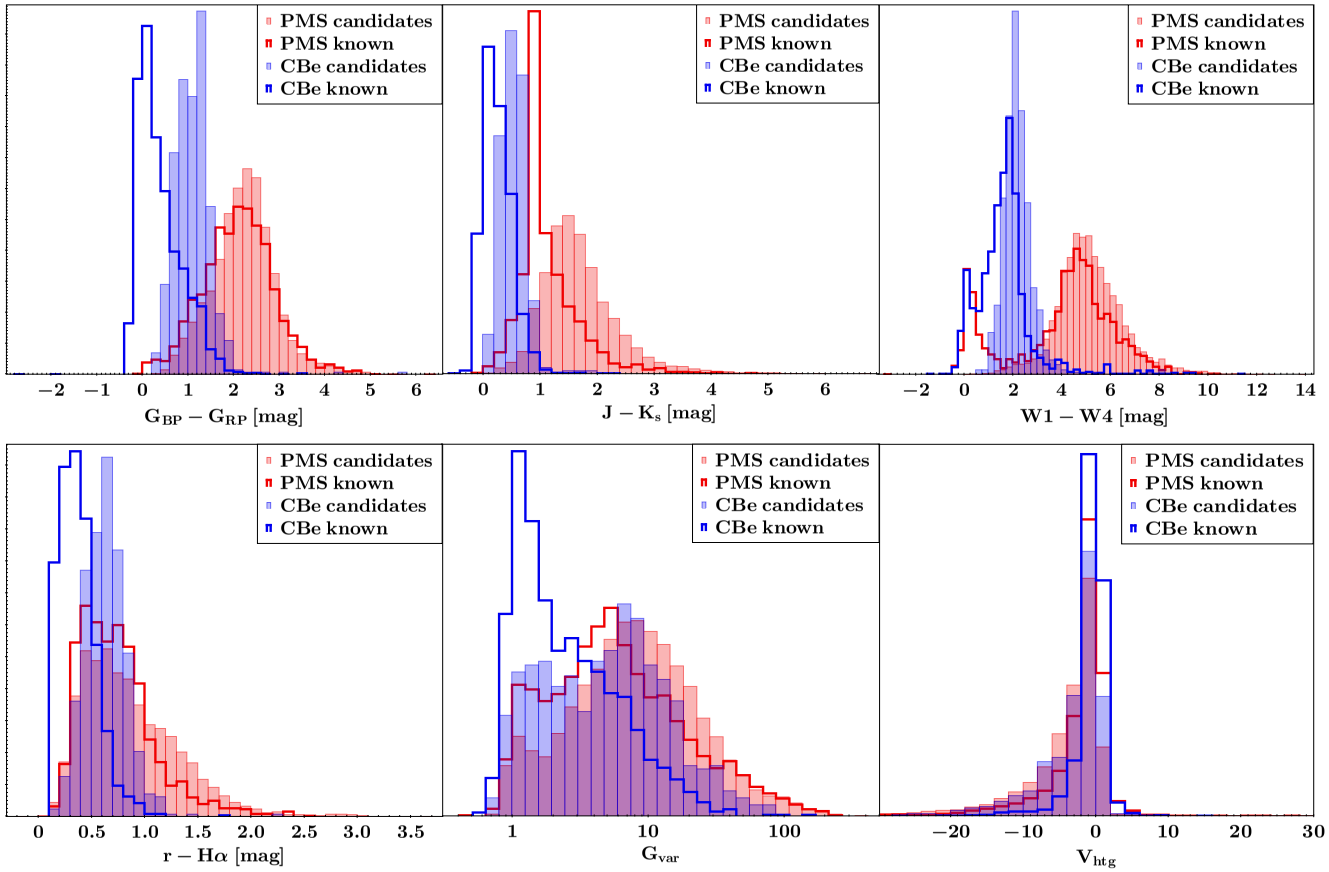


Fig. 3.15 Frequency density distribution of PMS candidates in shaded red and classical Be candidates in shaded blue for different selected observables. The red and blue lines respectively trace all considered known PMS and CBe objects, including those without all the observables. Area of each histogram has been normalised to one. For clarity, some individual extreme sources are out of bounds in the $r - H\alpha$, G_{var}, and V_{htg} plots.

as PN and 34 as possible PN. By studying their location in the observable space I concluded that any candidate with a $r - H\alpha \gtrsim 1.3$ should be treated with caution (16% of the sample of PMS candidates). Below that number I estimate the possible contamination by PN to be below 5%. Candidates with $r - H\alpha \geq 1.3$ are marked in Tables 3.3 and 3.4 with a 'PN' warning flag. Indeed, there are eight PN (within 5 arcsecond) from Sabin et al. (2014), 40 from Kerber et al. (2003), and three from Stanghellini et al. (2008) among the PMS candidates. A fraction of 46/51 (90%) have $r - H\alpha \geq 1.3$. I expect also some contamination in these works. PMS candidates with $r - H\alpha \geq 1.3$ have mostly absolute magnitudes $M_G > 6$, so they barely contaminate the sample of Herbig Ae/Be candidates (Sect. 3.4.3). In contrast, the few candidates with a 'PN' warning flag and $M_G < 6$ have a significant probability of being unclassified B[e] (FS CMa) stars.

6. Similarly, I detect a high number of carbon stars among my PMS candidates (71 confirmed and 16 candidates, according to SIMBAD). They stand out in variability and near-IR excess, but not in mid-IR excess, where they have a smaller excess than the rest of the candidates. Only two were classified as Herbig Ae/Be (Gaia DR2 1828276425855506304 and Gaia DR2 5336019093122634624 with PMS probability of 0.52 and 0.53 respectively) as the other 85 do not have reliable astrometry. Not surprisingly, 80/87 (92%) were identified as contaminants in Sect. 3.4.6 and marked in Table 3.3 with a 'G_{UMAP}' warning flag. Therefore, I do not expect them to have a high impact on the final catalogue of PMS candidates. In addition, 29 PMS candidates appear as variable stars of Mira Cet type in the cross-match with SIMBAD. A fraction of 17/29 (59%) are flagged as 'G_{UMAP}' in Table 3.3.

7. 51 PMS candidates are in the catalogue of OH/IR stars of Engels and Bunzel (2015, within 5 arcsecond). These 51 sources have different categories in

SIMBAD and may have been assessed in previous points and sections as of other nature (even YSO). From the catalogue of AGBs of Suh and Hong (2017), 26 (within 5 arcsecond) are among the PMS candidates (14, 54%, are flagged as 'G_{UMAP}' in Sect. 3.4.6). Similarly, a very small fraction of the CBe catalogue is potentially contaminated by sources with very strong mid-IR excess or that seem evolved. There are 0 post-AGB stars of Szczerba et al. (2007) in either catalogue.

Most of the contaminants discussed in the previous points can be avoided by constraining the position in the HR diagram and moving away from the giant and supergiant region. One way to do so is with the constraint applied in Fig. 3.9 at right. This also implies that the sample of Herbig Ae/Be candidates (see Sect. 3.4.3) is barely affected by these contaminants. Conversely, the HR diagram is a powerful tool to discard contaminants in other catalogues which were correctly classified in this chapter.

3.5.3 Probability of being either PMS or classical Be

There are 1309 sources whose probabilities of being PMS and CBe sum up to a probability $\geq 50\%$, but that are below 50% in either category (i.e. $p(PMS) + p(CBe) \geq 50\%$ but $p(PMS) < 50\%$, $p(CBe) < 50\%$, see Fig. 3.7). This means that the algorithm thinks that they belong to one of those two categories, but it is unable to say which. A closer look at these objects reveals that they behave very similarly to known CBe stars in terms of G_{var}, V_{htg}, and $r - H\alpha$ but their mean $G_{BP} - G_{RP}$ colour is redder and they have slightly larger near-IR excess ($J - K_s$). The $W1 - W4$ colour peaks where CBe stars do but it is quite homogeneously distributed. This is not surprising as in Fig. 3.14 at left they appear mostly mixed with the CBe candidates, in the region where the PMS candidates that are more similar to known CBes are placed.

These borderline objects are interesting in their own right and contain probably most of the less active PMS sources and in particular, most of the less active and probably more evolved Herbig Ae/Be stars. These sources are listed in a table equivalent to Tables 3.3 and 3.4 which is only available in electronic form (see Sect. 3.6). A more detailed analysis of these sources is presented in Chapter 4.

3.5.4 Important observables

In this section, I try to assess how important the different observables are for identifying PMS and classical Be stars. This is not trivial because of the intrinsic nature of the ANN-based algorithm, as the selection itself is not an obvious process.

What I did was to repeat the pipeline explained in Sect. 3.3 excluding some observables. I did not include the sources that were removed by demanding detections in those observables (see Sect. 3.2 and Table 5.1), as this would make it impossible to know whether the new results are caused by the new sources or the lack of those observables. Similarly, the ANN architecture was optimized for the whole set of observables (Sect. 3.3.2), so using fewer observables might alter the performance in an uncontrolled manner that can affect the conclusions. In order to minimise the impact of this I removed only a few observables at a time. Another problem is that by using fewer observables without changing the complexity of the algorithm I may start overfitting the selection. To assess this I checked that the retrieved precisions and recalls are within reasonable limits in each case.

In Table 3.2 I present the results (precision, recall, PMS and CBe candidates with $p \geq 0.5$) obtained when applying the same algorithm of Sect. 3.3 to the same sources of Sects. 3.1.4 and 3.2 but excluding certain observables (in the case of passbands this implies excluding the colours they appear on, see Sect. 3.1.1). In the last two columns I express the percentage of PMS and CBe candidates of those selections that were also retrieved when using all the observables and the percentage

of sources classified when using all the observables present in these new catalogues. These two values, in some sense, are equivalent to precisions and recalls if the catalogues obtained using all observables are taken as reference. As the algorithm was optimized to maximise the precision on PMS sources (see Sect. 3.3.3), this is maximum when using all observables. Similarly, when applying the algorithm to a smaller set of observables the results are going to be inevitably worse (as I do not include more sources). However, there is information in how much worse they get, although I can only talk in relative terms.

As outlined above, this table should be treated with caution but it gives information about the relative importance of the different observables in the selection of PMS and CBe candidates. I discuss the main outcomes of Table 3.2 in the following points:

1. Not using $r - H\alpha$ does not change the output tremendously. The number of PMS and CBe candidates retrieved is similar and there is only a small discrepancy (of $\sim 20\%$) with the case of using this colour, in the sense that mostly the same sources are identified and not many sources that were not identified when using $r - H\alpha$ are included. This is because cool stars have the same $r - H\alpha$ colour than hot stars with emission, see Drew et al. (2005), and hence this observable is not efficient in separating high-mass PMS and CBe sources from other objects.

2. G_{BP}, G, and G_{RP} are more relevant for the selection. If I do not use them the discrepancy with the original set of PMS candidates obtained using all the observables is higher than in the $r - H\alpha$ case. Many more sources are obtained (which can be considered a sort of contaminants) without losing many of the catalogued ones using all observables. In the case of the CBes, the effect is the opposite, not many contaminants are added but a few of the sources identified with these colours are lost. This might be caused by these colours carrying

Table 3.2 Evaluation of the impact of the different observables in the final selection.

	Precision PMS (%)	Recall PMS (%)	Precision CBe (%)	Recall CBe (%)	PMS ($p \geq 0.5$)	CBe ($p \geq 0.5$)	PMS P/R (%)	CBe P/R (%)
All observables	**40.7 ± 1.5**	**78.8 ± 1.4**	**88.6 ± 1.1**	**85.5 ± 1.2**	**8470**	**693**	**100/100**	**100/100**
No *r*, Hα	35.5 ± 1.4	73.1 ± 1.5	87.1 ± 1.1	85.52 ± 0.97	9108	718	77/82	80/83
No G_{BP}, G, G_{RP}	32.11 ± 0.99	74.2 ± 1.2	87.5 ± 1.0	82.2 ± 1.3	10772	650	63/80	80/75
No J, H, K_s	32.4 ± 1.2	81.9 ± 1.4	87.42 ± 0.84	86.59 ± 0.64	11123	758	69/91	81/88
No W1, W2, W3, W4	28.33 ± 0.68	80.4 ± 1.3	70.2 ± 1.5	76.6 ± 1.5	13303	1890	47/74	25/69
No W1, W2	37.6 ± 1.1	75.1 ± 1.3	86.0 ± 1.2	85.8 ± 1.1	8398	742	76/75	75/80
No W3, W4	31.7 ± 1.1	81.0 ± 1.2	83.9 ± 1.2	75.4 ± 1.9	11889	820	68/95	58/69
No G_{var}, V_{htg}	23.49 ± 0.66	74.0 ± 1.3	91.98 ± 0.83	80.7 ± 1.1	18055	358	40/84	100/52
No G_{var}	27.78 ± 0.84	74.7 ± 1.4	92.62 ± 0.80	80.2 ± 1.8	13667	355	53/85	99/51
No V_{htg}	40.0 ± 1.1	73.1 ± 1.4	91.89 ± 0.97	83.8 ± 1.6	7543	468	97/87	100/67

Notes: Metrics and number of sources obtained for the different classifications excluding the indicated observables. Precisions are lower limits (see Sect. 3.3.1). Last two columns indicate the 'precision' and 'recall' of these new catalogues with respect to the catalogues obtained when using all the observables. This is, the proportion of sources of the different catalogues that are in the catalogues obtained using all the observables and the proportion of the candidates obtained when using all the observables that are still retrieved when excluding the respective observables.

the photospheric information less affected by disc emission and hence more representative of temperature. Therefore, for the PMS case including them helps to remove candidates with unfeasible temperatures (like white dwarfs) and in the CBe case it helps a bit the selection as they are mostly blue with low extinctions (see Figs. 3.3 and 3.9).

3. J, H, and K_s: Not using these 2MASS passbands makes very few PMS and CBe candidates to be lost (less than in the previous cases) but I get many PMS contaminants, implying that the colours involving these passbands are relevant for differentiating PMS sources from other objects, although they are not critical for the selection. These observables do not seem to have a big impact for the classification of CBes.

4. As expected, W1, W2, W3, and W4 are very important. Not using these WISE passbands drastically increases the number of contaminants and significantly reduces the number of PMS and CBe candidates obtained when compared to the case of using this information. However, removing so many observables at a single time can cause the algorithm to start overfitting, so these results might be a bit exaggerated. If I choose to not use only W1 and W2 the selection is not much affected (only a bit more than in the $r - H\alpha$ case). Not surprisingly, it is if I choose not to use W3 or W4 when very poor results are obtained. I retrieve mostly the same PMS candidates but also many PMS contaminants. The larger impact is in the case of the CBes, as a lot of them are missed and almost half of the obtained catalogue is misclassified. This is expected, as at this wavelength range is where the discs of PMS stars and CBes start differing. Therefore, probably many CBes are moved to the PMS catalogue lowering its precision. This is the reason I opted to keep these passbands even though they suffer from a high incidence of spurious detections (see Sects. 3.1.4 and 3.5.2).

5. G_{var} and V_{htg}: The observable G_{var} proves to be of the utmost importance for identifying PMS sources. When excluding both variability indicators, I get twice as many PMS candidates than in the catalogue using all observables, with an almost full recovery of the later ones. Curiously, half of the CBes are lost, but not a single contaminant is added. If I exclude them independently I find that V_{htg} is almost irrelevant. It only helped to classify several CBes. In contrast, not using G_{var} doubles the number of PMS candidates retrieved (so half the sources can be considered contaminants) and halves the number of CBes obtained. The number of CBe contaminants is close to zero in every case. All these imply that this indicator is very useful for separating PMS sources from other objects and, in some cases, to differentiate them from CBes, but ineffective to identify CBes from the background sources. I showed in Chapter 2 that G_{var} mostly traces irregular photometric variations caused by edge-on dusty discs, so this result is just as expected.

It is clear that if I had optimized the algorithm for each situation using all the sources available in each case for the training I would have obtained more candidates and, from Table 3.2, it is safe to say that these would have been more contaminated.

3.6 Catalogue of new PMS and classical Be stars

Using the well-characterised sample of previously known Herbig Ae/Be stars of Chapter 2, in this chapter I have created a homogeneous and well-defined catalogue of new Herbig Ae/Be candidates by means of machine learning techniques. Standard techniques are not efficient for identifying new Herbig Ae/Be stars mainly because of their similarity with classical Be stars, with which they share many characteristics. By focusing on disentangling these two types of objects, my algorithm has also identified new classical Be stars. A catalogue of 8470 new pre-main sequence candidates and another catalogue of 693 new classical Be candidates were obtained.

From the catalogue of new pre-main sequence stars, at least 1361 sources are strong new Herbig Ae/Be candidates according to their position in the Hertzsprung-Russell diagram. Note that many more Herbig Ae/Be candidates can be obtained by relaxing the constraints on the parallax quality (see Sect. 3.4.3). This catalogue of new Herbig Ae/Be stars increases the number of known objects of the class by an order of magnitude.

The catalogue of new pre-main sequence stars (Table 3.3), new classical Be stars (Table 3.4), Sample of Study with probabilities (described in Sect. 3.4), and table of sources that belong to either category (described in Sect. 3.5.3) are only available in electronic form in their entirety[4].

Below I present a portion of Tables 3.3 and 3.4 for guidance regarding their form and content. The electronic version of the tables contains the uncertainties of the magnitudes, quality flags and rest of Gaia parameters together with angular distances from AllWISE and IPHAS or VPHAS+ sources to Gaia sources. In the points that follow I describe the possible warning flags of Tables 3.3 and 3.4 by alphabetical order. See the referred text for further discussion.

- G_{UMAP}: Possible evolved star contaminant. Identified through UMAP visualisation. Discussed in Sect. 3.4.6.
- ID AllW: Source with an AllWISE name repeated in the Sample of Study. Discussed in Sect. 3.5.2, point 3.
- ID IPH/VPH: Source with an IPHAS or VPHAS+ name repeated in the Sample of Study. Discussed in Sect. 3.5.2, point 3.
- PN: Possible planetary nebula or 'unclassified B[e]' contaminant. Defined as those candidates with $r - H\alpha \geq 1.3$. Discussed in Sect. 3.5.2, point 5.

[4] https://vizier.u-strasbg.fr

- Var: Photometrically variable PMS candidate. Defined as those PMS candidates with $G_{var} \geq 10$. Discussed in Sect. 3.4.4.

- W3W4: Source which extended source flag of AllWISE catalogue is different of 0. Discussed in Sect. 3.5.2, point 2.

- X-mtch: Likely false candidate because of incorrect cross-match with IPHAS or VPHAS+. Discussed in Sect. 3.4.

Table 3.3 PMS candidates ($p \geq 50\%$, 8470 sources) ordered by probability. In boldface sources that are strong Herbig Ae/Be candidates according to their position on the HR diagram ($M_G < 6$, 1361, see Fig. 3.13). These Herbig Ae/Be candidates are also indicated in the column 'HAeBe'.

Gaia source id	RA	DEC	Probability	Probability	Probability	V_{htg}	G_{var}
	h:m:s	deg:m:s	PMS	CBe	Other		
189190423171051136	**05:39:18.1**	**+36:17:16**	**1.00±0.00**	$\mathbf{(7.8 \pm 2.0) \cdot 10^{-6}}$	$\mathbf{(4.2 \pm 1.6) \cdot 10^{-6}}$	**-52.13**	**158.74**
513068993519575808	01:39:32.5	+64:53:02	1.00±0.00	$(4.18 \pm 0.91) \cdot 10^{-5}$	$(7.7 \pm 2.9) \cdot 10^{-6}$	-66.16	165.17
5546522453993928960	08:12:40.5	-34:14:12	1.00±0.00	$(4.4 \pm 2.0) \cdot 10^{-5}$	$(1.38 \pm 0.58) \cdot 10^{-5}$	-117.08	225.56
181175743617347328	05:23:04.3	+33:28:46	1.00±0.00	$(1.67 \pm 0.44) \cdot 10^{-4}$	$(4.7 \pm 2.0) \cdot 10^{-6}$	-47.67	131.31
3430718965791698176	**05:52:51.7**	**+26:47:30**	**1.00±0.00**	$\mathbf{(3.04 \pm 0.96) \cdot 10^{-4}}$	$\mathbf{(3.6 \pm 1.0) \cdot 10^{-5}}$	**-67.90**	**152.91**
3046391307734381184	07:09:22.3	-10:30:57	1.00±0.00	$(2.06 \pm 0.55) \cdot 10^{-4}$	$(1.52 \pm 0.57) \cdot 10^{-4}$	-44.33	77.08
3449189833426211840	05:32:13.9	+34:06:01	1.00±0.00	$(8.0 \pm 2.8) \cdot 10^{-5}$	$(2.8 \pm 1.8) \cdot 10^{-4}$	-107.39	183.40
189577661722195840	05:37:53.2	+37:24:56	1.00±0.00	$(2.61 \pm 0.60) \cdot 10^{-4}$	$(1.52 \pm 0.51) \cdot 10^{-4}$	-66.15	117.40
2006046771487587328	22:24:13.8	+56:11:33	1.00±0.00	$(2.3 \pm 1.1) \cdot 10^{-4}$	$(2.38 \pm 0.99) \cdot 10^{-4}$	-117.72	162.47
4279295067717293696	**18:47:32.6**	**+02:12:06**	**1.00±0.00**	$\mathbf{(1.31 \pm 0.21) \cdot 10^{-4}}$	$\mathbf{(3.82 \pm 0.94) \cdot 10^{-4}}$	**-33.04**	**96.61**
3114712967419900544	07:09:42.8	+01:53:11	1.00±0.00	$(5.5 \pm 2.4) \cdot 10^{-5}$	$(5.4 \pm 4.0) \cdot 10^{-4}$	-46.52	143.75
3102576519414606464	06:58:34.4	-03:56:46	1.00±0.00	$(1.75 \pm 0.58) \cdot 10^{-4}$	$(4.3 \pm 1.2) \cdot 10^{-4}$	-26.55	58.08
...	...	...	...	...	...	...	...

G_{BP} (mag)	G (mag)	G_{RP} (mag)	IPHAS or VPHAS+ name	r (mag)	i (mag)	Hα (mag)	AllWISE name	J (mag)	H (mag)	K_s (mag)
15.97	**15.04**	**14.04**	**J053918.09+361716.2**	**13.78**	**13.06**	**13.21**	**J053918.08+361716.1**	**12.98**	**11.23**	**9.79**
14.34	13.80	12.98	J013932.55+645302.3	12.91	12.43	12.35	J013932.53+645302.2	11.43	10.21	9.11
14.97	14.18	13.20	J081240.5-341411.7	12.86	12.19	11.99	J081240.50-341411.7	10.86	10.08	9.59
14.89	14.40	13.66	J052304.26+332846.5	13.80	13.24	12.90	J052304.26+332846.4	12.01	11.12	10.25
14.40	**13.94**	**13.05**	**J055251.75+264730.2**	**12.62**	**12.22**	**12.28**	**J055251.74+264730.0**	**12.16**	**11.04**	**10.06**
16.75	15.56	14.31	J070922.3-103057.0	16.04	14.94	14.78	J070922.28-103057.0	12.47	10.98	9.90
16.17	14.76	13.63	J053213.92+340601.5	13.70	12.39	13.25	J053213.92+340601.4	10.80	9.83	8.83
14.73	13.68	12.63	J053753.15+372456.1	14.30	13.07	13.80	J053753.15+372456.0	10.54	9.83	9.31
14.00	13.31	12.40	J222413.78+561133.2	13.63	12.92	13.29	J222413.75+561133.2	9.56	8.69	8.01
15.15	**14.11**	**13.05**	**J184732.62+021205.9**	**15.01**	**13.84**	**14.23**	**J184732.61+021205.7**	**11.34**	**9.97**	**8.79**
13.62	11.79	10.19	J070942.85+015311.0	13.86	11.84	13.15	J070942.75+015311.1	8.20	6.62	5.25
18.24	17.57	16.41	J065834.4-035645.6	21.41	20.84	20.13	J065834.39-035645.6	13.73	12.40	11.38
...	...	...	...	...	...	...	...	...	...	...

W1 (mag)	W2 (mag)	W3 (mag)	W4 (mag)	RUWE	Parallax (ϖ) (mas)	Distance (pc)	A'_G (mag)	M_G (mag)	$G_{BP}-G_{RP}$ (mag)	HAeBe	Flag
7.90	**7.05**	**4.84**	**3.23**	**1.36**	$\mathbf{0.636 \pm 0.048}$	$\mathbf{1510^{+120}_{-100}}$	**1.06**	**3.09**	**1.41**	**Yes**	**Var**
7.95	7.26	5.41	3.90	1.86	1.116 ± 0.029	874^{+23}_{-22}	-	-	-	-	Var
8.63	7.81	5.37	3.32	1.33	2.877 ± 0.025	$344.2^{+3.0}_{-2.9}$	0.06	6.43	1.75	-	Var, W3W4
9.12	8.39	6.35	4.21	1.37	0.234 ± 0.040	3710^{+650}_{-490}	1.04	0.51	0.72	-	Var, W3W4
9.26	**8.76**	**7.23**	**5.56**	**1.09**	$\mathbf{1.950 \pm 0.032}$	$\mathbf{505.6^{+8.3}_{-8.0}}$	**0.27**	**5.15**	**1.22**	**Yes**	**Var**
7.19	5.90	3.19	1.06	3.28	1.00 ± 0.20	1020^{+320}_{-200}	-	-	-	-	Var, W3W4
6.92	5.79	3.95	2.94	1.01	-0.068 ± 0.080	7200^{+2500}_{-1800}	-	-	-	-	G_{UMAP}, Var
8.34	7.37	5.01	3.66	1.32	0.025 ± 0.041	8200^{+2300}_{-1600}	-	-	-	-	Var, W3W4
7.06	5.99	3.79	2.43	1.72	0.326 ± 0.027	2810^{+230}_{-200}	-	-	-	-	Var, W3W4
7.61	**6.97**	**4.84**	**3.10**	**1.30**	$\mathbf{1.614 \pm 0.046}$	$\mathbf{609^{+18}_{-17}}$	**1.99**	**3.19**	**1.13**	**Yes**	**Var**
3.99	1.79	1.93	1.44	1.01	0.103 ± 0.089	4500^{+1600}_{-1100}	-	-	-	-	G_{UMAP}, Var, W3W4
10.58	9.25	6.40	4.02	1.15	0.36 ± 0.11	2580^{+1120}_{-640}	-	-	-	-	Var, W3W4
...	...	...	...	...	...	...	...	...	...	...	...

Notes: This table is available in electronic form in its entirety with uncertainties of the magnitudes, quality flags, and rest of Gaia parameters together with angular distances from AllWISE and IPHAS or VPHAS+ sources to Gaia sources. A portion is shown here for guidance regarding its form and content. The probabilities are expressed to the precision of their uncertainties. Distances from Bailer-Jones et al. (2018). A'_G only traces the interstellar extinction. It is used to correct M_G and $G_{BP}-G_{RP}$. A'_G, M_G, and $G_{BP}-G_{RP}$ are only presented for sources with RUWE<1.4 and $\varpi/\sigma(\varpi) \geq 5$. By construction Herbig Ae/Be candidates are astrometrically well behaved (RUWE<1.4 and $\varpi/\sigma(\varpi) \geq 10$). These constraints can be relaxed to obtain more Herbig Ae/Be candidates. The probabilities of the three categories sum up to 1. The different catalogue warning flags are discussed throughout the chapter and summarised in Sect. 3.6.

Table 3.4 A representative sample of the full table of classical Be candidates ($p \geq 50\%$, 693 sources) ordered by probability.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Probability CBe	Probability Other	V_{htg}	G_{var}
...	...	...	...	...	...	...	...
2012831922144678272	23:43:35.7	+61:27:44	0.119±0.010	0.831±0.015	0.0500±0.0061	-1.87	7.75
3106114885277650944	06:49:27.0	-02:33:30	0.105±0.014	0.831±0.020	0.064±0.010	-1.36	2.40
509370275062324608	01:34:15.6	+59:30:59	0.146±0.012	0.830±0.014	0.0239±0.0032	-7.87	13.95
1889519708863266 56	05:11:53.5	+40:13:11	0.1127±0.0099	0.830±0.015	0.0572±0.0081	0.14	1.62
2173867430250308608	21:50:31.8	+54:04:53	0.122±0.016	0.830±0.019	0.0483±0.0058	-10.14	15.16
4268987932187374848	19:07:10.6	+03:29:22	0.107±0.010	0.828±0.017	0.0644±0.0090	-0.42	6.07
518388499496394752	01:58:16.4	+65:49:52	0.107±0.010	0.828±0.015	0.0647±0.0071	-7.34	11.67
2201241387128842240	22:18:59.0	+59:47:10	0.137±0.018	0.827±0.022	0.0356±0.0083	0.00	0.67
2060197062012903296	20:00:28.0	+37:06:52	0.103±0.012	0.827±0.017	0.0700±0.0080	0.08	1.18
3325867711602501504	06:21:13.0	+09:07:21	0.101±0.011	0.827±0.019	0.072±0.016	-0.23	0.99
3106856785040836224	06:37:36.4	-01:54:03	0.107±0.010	0.824±0.016	0.0694±0.0079	-6.52	12.02
3356939124228643968	06:35:02.5	+15:17:52	0.125±0.014	0.823±0.017	0.0518±0.0061	-0.27	0.87
...	...	...	...	...	...	...	...

G_{BP} (mag)	G (mag)	G_{RP} (mag)	IPHAS or VPHAS+ name	r (mag)	i (mag)	Hα (mag)	AllWISE name	J (mag)	H (mag)	K_s (mag)
...	...	...	...	...	...	...	...	...	...	...
13.14	12.80	12.26	J234335.71+612744.0	12.92	12.39	12.28	J234335.70+612743.9	11.59	11.36	11.15
12.93	12.67	12.18	J064926.96-023330.0	12.56	12.14	11.88	J064926.98-023330.5	11.41	11.24	10.87
12.27	11.99	11.51	J013415.63+593058.7	12.00	11.60	11.40	J013415.64+593058.6	10.98	10.69	10.48
12.66	12.43	11.99	J051153.45+401310.9	12.50	11.92	11.64	J051153.44+401310.9	11.55	11.38	11.21
12.87	12.32	11.58	J215031.85+540452.9	12.73	11.91	11.94	J215031.83+540452.7	10.71	10.41	10.24
12.30	12.04	11.59	J190710.55+032922.3	12.15	11.60	11.37	J190710.55+032922.4	10.95	10.74	10.54
12.81	12.32	11.67	J015816.40+654951.9	12.16	11.56	11.61	J015816.37+654951.9	10.89	10.65	10.46
12.73	12.33	11.75	J221859.05+594710.0	12.60	11.84	12.18	J221859.03+594709.9	11.02	10.90	10.77
12.43	12.25	11.94	J200028.00+370651.9	12.33	11.93	11.99	J200028.00+370653.0	11.54	11.45	11.37
11.99	11.90	11.66	J062112.96+090721.4	12.48	11.71	11.44	J062112.95+090721.4	11.37	11.28	11.13
13.13	12.68	12.06	J063736.41-015403.0	12.58	12.04	12.21	J063736.41-015402.9	11.26	11.02	10.84
12.56	12.37	12.01	J063502.46+151752.5	12.50	12.05	12.11	J063502.46+151752.5	11.57	11.52	11.37
...	...	...	...	...	...	...	...	...	...	...

W1 (mag)	W2 (mag)	W3 (mag)	W4 (mag)	RUWE	Parallax (mas)	Distance (pc)	A'_G (mag)	M_G (mag)	$G_{BP} - G_{RP}$ (mag)	Flag
...	...	...	...	...	...	...	...	...	...	...
10.82	10.56	9.76	9.15	0.96	0.321 ± 0.023	2850^{+200}_{-180}	2.31	-1.78	-0.25	-
10.23	9.99	9.35	9.00	0.89	0.356 ± 0.035	2600^{+270}_{-230}	1.37	-0.78	0.08	W3W4
10.14	9.83	8.92	7.70	1.11	0.286 ± 0.033	3130^{+360}_{-300}	1.36	-1.85	0.09	-
10.63	10.33	9.40	8.21	1.30	0.108 ± 0.044	6000^{+1700}_{-1200}	-	-	-	-
9.17	8.91	8.33	7.50	1.03	0.256 ± 0.028	3470^{+390}_{-320}	2.11	-2.50	0.25	W3W4
10.23	10.03	9.32	8.71	0.82	0.388 ± 0.078	2400^{+590}_{-400}	-	-	-	W3W4
10.25	10.00	9.31	8.19	0.99	0.347 ± 0.034	2660^{+270}_{-230}	2.47	-2.27	-0.07	-
10.68	9.75	10.42	8.46	0.95	1.137 ± 0.032	858^{+24}_{-23}	1.52	1.14	0.24	W3W4
10.15	9.53	9.55	8.86	1.06	0.864 ± 0.027	1121^{+35}_{-33}	0.61	1.39	0.18	W3W4
11.06	10.84	9.91	8.63	0.93	0.340 ± 0.046	2710^{+410}_{-320}	0.81	-1.08	-0.06	X-mtch
10.47	10.22	9.40	8.70	0.90	0.217 ± 0.036	3990^{+670}_{-510}	1.48	-1.80	0.35	W3W4
11.20	10.78	10.53	8.66	0.97	0.446 ± 0.056	2110^{+290}_{-230}	0.91	-0.16	0.11	-
...	...	...	...	...	...	...	...	...	...	...

Notes: This table is available in electronic form in its entirety with uncertainties of the magnitudes, quality flags, and rest of Gaia parameters together with angular distances from AllWISE and IPHAS or VPHAS+ sources to Gaia sources. A portion is shown here for guidance regarding its form and content. The probabilities are expressed to the precision of their uncertainties. Distances from Bailer-Jones et al. (2018). A'_G only traces the interstellar extinction. It is used to correct M_G and $G_{BP} - G_{RP}$. A'_G, M_G, and $G_{BP} - G_{RP}$ are only presented for sources with RUWE<1.4 and $\varpi/\sigma(\varpi) \geq 5$. The probabilities of the three categories sum up to 1. The different catalogue warning flags are discussed throughout the chapter and summarised in Sect. 3.6. I note that some of the sources presented here have a 'X-mtch' flag (57 out of 693), and hence they are likely false candidates because of an incorrect cross-match with IPHAS or VPHAS+.

Chapter 4

Analysis of candidates and independent observations

Non ignoravi me mortalem genuisse

In this chapter I perform an analysis from different perspectives of the catalogues of pre-main sequence (PMS) and classical Be (CBe) candidates constructed in Chapter 3. To this I add an analysis of the catalogue of sources that have a probability of larger than 50% of belonging to one of the two categories. Although little importance was given to it in the previous chapter, this latter catalogue is in some respects the most interesting one as it may contain most of the massive PMS sources.

The analysis is complemented with independent spectroscopic observations that I conducted for a subsample of the catalogues. These observed objects can be characterised and studied in better detail than the larger catalogues they were obtained from. Therefore, the set of observed objects is used as a control sample. I start the chapter with a description of the observations (Sect. 4.1). After this I describe the properties of the observed objects (Sects. 4.2, 4.3, and 4.4), analyse the possible contaminants (Sect. 4.5), and discuss the results of the observations (Sect. 4.6). Finally, I move on to analyse the general catalogues in Sect. 4.7.

4.1 Observations

I carried out three independent observing runs to sample the catalogues presented in Chapter 3 with optical spectroscopy, with emphasis on the PMS candidate catalogue. The goal was to evaluate the accuracy and quality of the catalogues, and also to characterise spectral types and hence obtain total extinction values for a subset of interesting objects, allowing them to be placed in the HR diagram. As explained in Chapter 3, knowledge of the total extinction is essential to derive accurate stellar parameters.

In total 145 PMS candidates and 14 classical Be candidates were observed. In order to obtain precise stellar parameters I tried to select candidates with good astrometric quality when possible (all but two have $\varpi/\sigma(\varpi) \geq 5$ and RUWE < 1.4). This is because the parallax dominates the uncertainty of the stellar parameters when these are obtained from the location of the objects in the Hertzsprung-Russell diagram (see Chapter 2). I intentionally observed objects that were flagged as possible contaminants in Chapter 3. In particular, nine observed candidates are tagged as 'PN', 81 as 'W3W4', and one as both 'ID AllW' and 'ID IPH/VPH'. The 'PN' flag traces possible planetary nebula or 'unclassified B[e]' contaminants; it applies to those objects with $r - H\alpha \geq 1.3$. The 'W3W4' flag traces sources which 'extended source flag' of AllWISE catalogue is different from 0, and hence might have spurious WISE photometry. The 'ID AllW' and 'ID IPH/VPH' flags respectively indicate sources with a possible incorrect cross-match with the AllWISE or IPHAS or VPHAS+ catalogues. These observed tagged contaminants are useful to test the accuracy of the catalogue warning flags.

The PMS target list was chosen to confirm the pre-main sequence nature of some of the objects which are expected to be more massive. Thus, I selected the observed PMS candidates by their expected mass as traced by the absolute magnitude corrected from interstellar extinction (presented in Chapter 3, see discussion in Sect. 3.2). As

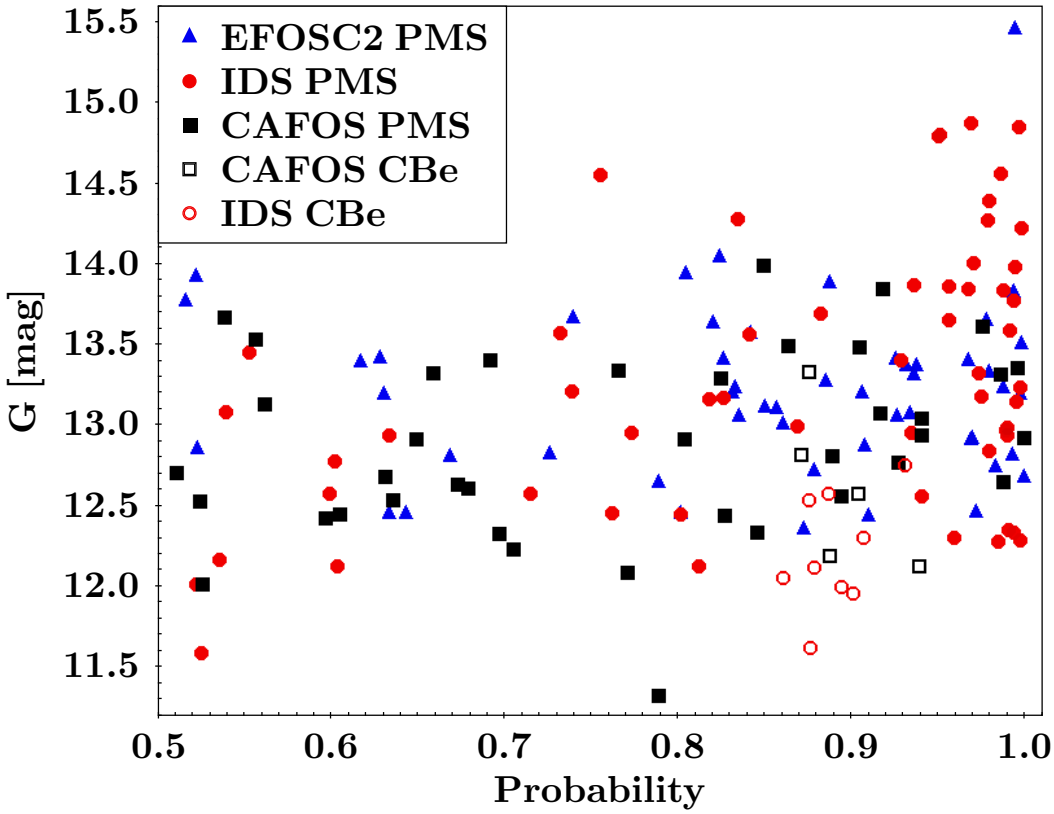


Fig. 4.1 Probability of each catalogue vs. Gaia white G band magnitude for the subsample of candidates independently observed with optical spectroscopy. The different shapes indicate the instrument used to observe each source: triangles for EFOSC2, circles for IDS, and squares for CAFOS. Filled symbols trace PMS candidates and unfilled symbols trace CBe candidates. The number of objects of each category observed at each telescope appears in Table 4.1.

the interstellar extinction is a lower limit, the lower limits to their masses should ensure they are massive objects indeed. In addition, I included PMS candidates that span the whole probability range to trace the sensitivity of the probability values (presented in Chapter 3). This was also done because the most massive objects tend to have lower probabilities as the algorithm struggles more to differentiate them from CBe sources.

In addition, a small number of 14 CBe candidates was observed, these were selected only based on their high CBe probability values.

The PMS and CBe catalogue probabilities of the observed candidates (each type of candidate with its corresponding catalogue probability) can be seen in Fig.

4.1. The number of sources observed in each observing run are tabulated in Table 4.1, together with the reciprocal dispersion and the spectral resolution of each configuration used. The individual sources are detailed in Appendix C. The three observing runs can be summarised as follows:

- 4th – 11th August 2019. Eight nights with the Intermediate Dispersion Spectrograph (IDS) instrument which is at the Cassegrain focus of the 2.54-metre Isaac Newton Telescope (INT). The INT is located at the Roque de los Muchachos Observatory in the island of La Palma, Spain. In two blocks of four nights each, two different configurations were used. One block used the EEV10 CCD detector together with the medium resolution R900V diffraction grating which covers the $\sim 3600-5000$ Å spectral range. This setting has a reciprocal dispersion of 0.63 Å/pixel and a spectral resolution of ~ 1.3 Å (or resolving power $R \sim 3400$). The other block used the higher resolution R1200R grating with the Red+2 CCD detector ($\sim 5700-6700$ Å). This setting has a reciprocal dispersion of 0.52 Å/pixel and a spectral resolution of ~ 1.0 Å (or resolving power $R \sim 6000$). Bias, flat and arc frames were taken each night for the reduction of the data.

- 3rd – 5th September 2019. Three nights with the Calar Alto Faint Object Spectrograph (CAFOS) and SITe-1d 2Kx2K CCD chip at the Calar Alto 2.2-metre telescope (CAHA2.2m) in Calar Alto Observatory, Spain. I employed the B-100 grism ($\sim 3200-5800$ Å) which produces a reciprocal dispersion of 2.0 Å/pixel and a spectral resolution of ~ 4.0 Å (or resolving power $R \sim 1100$). Bias, flat and arc frames were taken each night for the reduction of the data.

- 11th – 13th and 17th – 20th March 2020. Seven nights with the ESO Faint Object Spectrograph and Camera (v.2) or EFOSC2 in two settings. EFOSC2 is installed at the Nasmyth B focus of the 3.58-metre New Technology Telescope (NTT) at La Silla Observatory, Chile. The first block used a medium resolution

Table 4.1 Number of candidates of each category observed in each observing run together with the dispersion and resolution of each configuration used.

Telescope	Instrument	Dispersion blue/red range (Å/pixel)	Resolution blue/red range (Å)	PMS candidates	CBe candidates	Total
INT	IDS	0.63/0.52	1.3/1.0	56 (6)	9	65
CAHA2.2m	CAFOS	2.0/-	4.0/-	39 (1)	5	44
NTT	EFOSC2	0.96/0.55	7.4/3.5	51 (2)	0	51
Total				145*	14	159

Notes: Blue range corresponds to $\sim 3300-5400$ Å and red range corresponds to $\sim 5800-7000$ Å. Between brackets the number of candidates that have a 'PN' warning flag in the catalogues of Chapter 3. * Gaia 4094703381988286592 PMS candidate is repeated between NTT and INT observations.

G7 diffraction grism and the second block used the higher resolution G20 grism. The G7 grism ($\sim 3300-5300$ Å) provides a reciprocal dispersion of 0.96 Å/pixel and a spectral resolution of ~ 7.4 Å (or resolving power $R \sim 580$). The higher resolution G20 grism ($\sim 6000-7200$ Å) provides a reciprocal dispersion of 0.55 Å/pixel and a spectral resolution of ~ 3.5 Å (or resolving power $R \sim 1900$). Bias, flat and arc frames were taken each night for the reduction of the data.

The blue spectral region considered in the three observing runs ($\sim 3300-5400$ Å) covers the main wavelength range to determine spectral types. This region is especially useful for the earlier spectral type stars such as A and B stars, as their spectrum in the wavelength range beyond 5000 Å is fairly line free. The chosen grisms allow for efficiently obtaining spectral types which in turn can be converted to effective temperatures (T_{eff}), which are needed to deredden the sources by means of intrinsic colours. The blue region also covers the Balmer jump and Hβ, which allows the study of the circumstellar activity. Examples of the normalised spectra obtained in this spectral range are shown in Fig. 4.2 in the top panels.

At the INT and NTT telescopes additional observations covering the red spectral range ($\sim 5800-7000$ Å) were performed for each source that was observed

with the blue setting. The exceptions are Gaia DR2 3348898739291597568 and 4253225574005033088, which were not observed in the red spectral range even though they were observed in the blue range at the NTT. The red range covers the important diagnostic Hα line, which enables the determination of accretion rates. In addition, next to Hα, the red spectra cover an additional number of lines that are useful for diagnostics of (circumstellar) activity, such as the He 5876 line and the Na D doublet at 5890 Å. It also contains the strong Diffuse Interstellar Bands (DIBs) at 5797 and 5870 Å that trace interstellar extinction. The higher resolution allows one to better resolve the line-profiles, which are important for further assessing the nature of the candidates. Examples of the Hα lines observed within the red spectral range are shown in Fig. 4.2 in the bottom panels.

I wrote the observing proposals and was awarded time for these observations at telescopes from three different organisations. The observations were also prepared and carried out by myself but the spectra were reduced by Mr. C. Wichittanakom (for the INT data) and Prof. R. D. Oudmaijer (for the CAHA2.2m and NTT data). The signal-to-noise ratio of the spectra were typically of order 100. More detailed information regarding the observations will be provided in a future publication. The resulting effective temperatures or spectral types and the Hα equivalent widths and line profile information were forwarded to me as a private communication, which I include and discuss in the following sections.

4.1.1 Results from the blue spectral range

The observed 145 PMS candidates and 14 CBe candidates and their derived spectral types and temperatures from the blue spectral range ($\sim 3300 - 5400$ Å) are listed in Tables C.1 and C.4, respectively.

For several objects the temperature was estimated directly from model fitting the spectra, and thus no spectral type was derived for those sources. For the rest

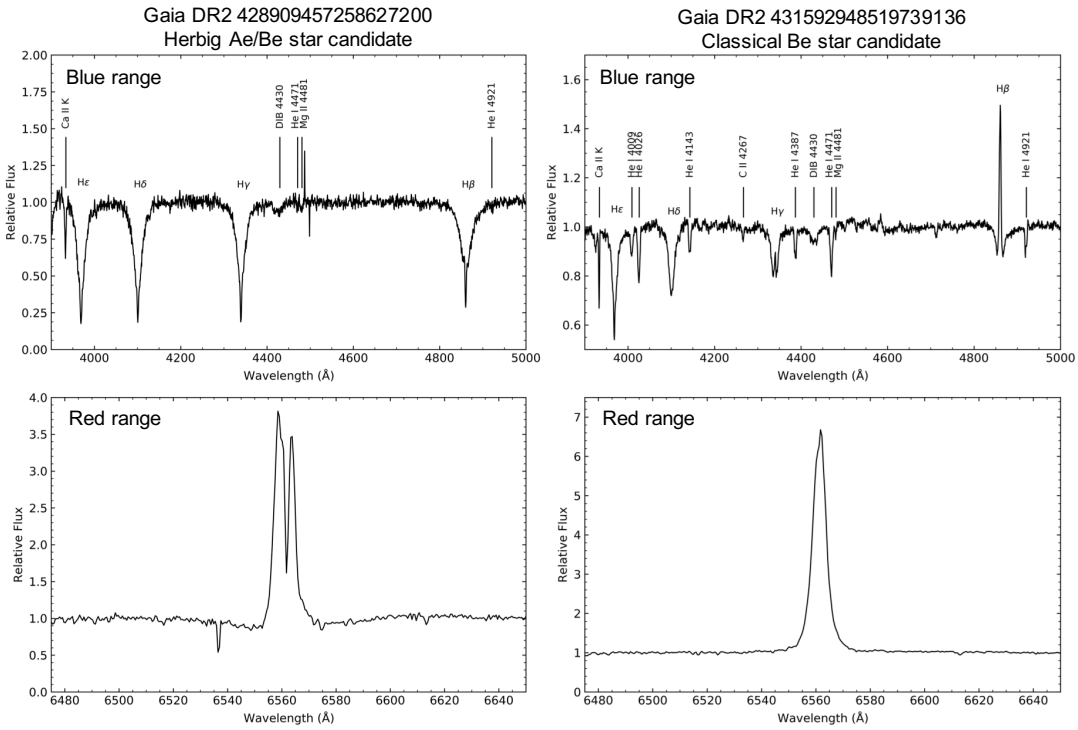


Fig. 4.2 Examples of the normalised spectra obtained with IDS of a PMS candidate and a classical Be candidate (see Table 4.1). The top panels contain the blue spectral range of each source and the bottom panels contain the corresponding Hα line within the red spectral range. Spectra are courtesy of Mr. C. Wichittanakom.

spectral types were obtained by comparison with model spectra and published spectral standards, and the T_{eff}s appearing in Tables C.1 and C.4 are the values that correspond to those spectral types according to Pecaut and Mamajek (2013). In the latter case the T_{eff} uncertainties are larger than with the previous method.

The brightness of the observed sources is shown in Fig. 4.1. Because of sensitivity limitations, only candidates at the bright end of the catalogues were observed. One could therefore argue that the observed sample is biased, however as the selection was distance independent, the conclusions that arise from these observations can be extrapolated to the fainter objects of the catalogues, given that they were all selected homogeneously by the machine learning algorithm. I note that it is at the fainter end of the catalogues where the vast majority of the sources are totally novel and unexplored. Thus, several of the observed objects have been already characterised

and studied by other surveys and authors. The independent classification of sources in my catalogues was thoroughly discussed in Chapter 3, where it was shown that those previous results often agree with my categorisation.

In this regard, I performed a 5 arsecond cross-match of the observed candidates with SIMBAD. The resulting 'SIMBAD object types' (the category the SIMBAD database assigns to each object; e.g. YSO, emission-line star) and spectral types from previous literature, if available, are tabulated in Tables C.1 and C.4 together with our determinations. There are 26 candidates with some previous spectral type information in SIMBAD. Only two of those sources show discrepancies greater than a few spectral subtypes with our spectral types or $T_{\rm eff}$ values. These are Gaia DR2 4272195138879459200 and 2013474586693428096. Although Gaia DR2 4272195138879459200 appears as a F6 star in SIMBAD, it has been classified by most previous studies, including ours, as of B type (e.g. The et al., 1994). Gaia DR2 2013474586693428096 was classified as B2 from its photometric $B-V$ colour by Uemura et al. (2004). However, the spectra of this object evidences that it is a much cooler object. Our determination is $T_{eff} = 6500 \pm 500$ K. Therefore, the independent SIMBAD spectral types evidence the accuracy of our derivations. I proceed with our determinations as they were obtained homogeneously.

In addition, there are 116 observed candidates with SIMBAD object types. These are consistent with the precisions derived in Chapter 3 of $\sim 81\%$ and $\sim 89\%$ for the PMS and CBe catalogues respectively. Individual analysis of the object types leads to similar conclusions to what was discussed in Chapter 3 (see Sects. 3.4.5 and 3.5.2). I note that the 'Candidate AGB' sources were catalogued by Robitaille et al. (2008) and in Sect. 3.5.2 it was concluded that my analysis supersedes theirs. See Sect. 4.5 for further discussion of possible contaminants.

4.1.2 Results from the red spectral range

Several spectral features in the red spectral range ($\sim 5800-7000$ Å) that can trace circumstellar activity and accretion in PMS objects were observed. In this section I mainly focus on describing what was observed in the Hα emission line.

In Sect. 4.5, I discuss that 12 clear contaminants among the observed PMS candidates can be directly identified from the observations. Excluding those, of the 95 PMS candidates with observations in the red spectral range, 85 ($\sim 89\%$) show Hα emission. In general, of the 133 PMS candidates remaining ($145-12$), 97 show emission in Hα or Hβ ($\sim 73\%$). I note that Hβ emission may not be present even if Hα emission is. All nine CBe candidates observed in the red spectral range show Hα emission and 13/14 show emission in Hα or Hβ ($\sim 93\%$). This high percentage of sources with hydrogen emission supports the PMS or CBe classification.

The Hα line profile was classified into single-peaked, double-peaked or P-Cygni profile, both regular or inverse, following the classification of Chapter 2. Of the aforementioned 95 PMS candidates with Hα line observations, 37 show single-peaked emission, 40 double-peaked emission and eight P-Cygni emission (ten sources have Hα in absorption). Therefore, of the PMS candidates with Hα emission, 44% are single-peaked, 47% are double-peaked and 9% are P-Cygni. These percentages are somewhat different to those observed for known HAeBes (31%, 52%, and 17% respectively; see Chapter 2). I suspect that this difference is caused by the low-resolution of my spectroscopical observations, which moved many P-Cygni and double-peak profiles to the 'single-peaked' group; and because of the presence of contaminants (see Sect. 4.5). Indeed, the main references for the line profiles of the known HAeBes in Chapter 2 were Vieira et al. (2003) and a private communication of the spectra used for Fairlamb et al. (2017). Both studies have a spectral resolution of $R \sim 10000$, which is two to five times larger than the resolution of my observations.

Of the 9 CBe candidates with Hα line observations, 7 (78%) have single-peaked emission and 2 (22%) have double-peaked emission. No P-Cygni profile was observed in them. I also expect these percentages to be affected by the low-resolution of my observations.

The measured Hα equivalent widths (observed above the continuum) and the Hα line profiles are tabulated in Tables C.2 and C.5 for all observed PMS and CBe candidates respectively. In addition, in these tables I present information of whether the Hβ line from the blue range ($\sim 3300-5400$ Å, Sect. 4.1.1) appears in emission.

4.1.3 SIMBAD set

In addition to the observed objects, I checked the literature for sources among the catalogues of Chapter 3 with spectral type determinations. For doing this I cross-matched the catalogues of PMS and CBe candidates, and the list of sources with a probability of larger than 50% of belonging to either category (which I call the 'Either' group hereafter), with the SIMBAD database using a 5 arcsecond aperture. This resulted in 187 stars (103 PMS, 56 CBe, and 28 Either) that were not observed by myself but have a spectral type in the literature ('SIMBAD set' hereafter). I found that the cross-match is consistent by comparing optical and infrared colours and no incorrect cross-match was detected.

These sources are tabulated in Appendix D. I note that the heterogeneity of the SIMBAD database introduces some uncontrolled scatter and spurious determinations, which makes this sample less useful than my own observations.

4.2 Intrinsic Gaia $G_{BP} - G_{RP}$ colours

In the next sections I use the observed spectral types or effective temperatures to determine the total extinction. However, for many objects, only Gaia photometry is

available, and the intrinsic $G_{BP} - G_{RP}$ colour of the different spectral types are not yet published, so it was necessary to perform my own derivations. This is not trivial. A set of stars with well-characterised spectral types is needed, with the additional complication that these objects need to have a negligible or well controlled extinction and photometric data that are not problematic (not variable and systematic free). As discussed below, there is no ideal set of sources with observationally obtained spectral types and Gaia data to derive empirical $G_{BP} - G_{RP}$ intrinsic colours.

However, at present I need to determine Gaia intrinsic colours empirically, because the very broad Gaia passbands do not yet have a properly characterised sensitivity curve (see later discussion in this section). A further complication is that Gaia photometry suffers from systematic effects and biases in the published values. This was first discussed in Evans et al. (2018) for the case of Gaia DR2. Later, Weiler (2018) and Maíz Apellániz and Weiler (2018) proposed new sensitivity curves and, building on previous results, detailed and summarised the biases and systematic effects of the Gaia photometry.[1] In particular, two magnitude ranges with different G_{BP} sensitivity curves are necessary, with the split at $G = 10.87$. Because of these issues, the Gaia intrinsic colours should not be obtained from theoretical models, as these are yet highly uncertain, and should be derived only from empirical observations.

All large surveys with temperature determinations and Gaia DR2 colours (e.g. Gaia DR2: Andrae et al., 2018; RAVE DR6: Steinmetz et al., 2020; or GALAH DR2: Buder et al., 2018) are largely incomplete at the spectral type range of interest here (O, B, A). Similarly, the LAMOST survey (Luo et al., 2019) is also inefficient because

[1] It turns out there is a small systematic, magnitude dependent, effect in the white G band. Its correction, which is applied from here onward, can be summarised by:

$$\begin{aligned} 2 < G \leq 6 &\longrightarrow G_{corr} = -0.047344 + 1.16405G - 0.046799G^2 + 0.0035015G^3, \\ 6 < G \leq 16 &\longrightarrow G_{corr} = G - 0.0032(G-6), \\ G > 16 &\longrightarrow G_{corr} = G - 0.032. \end{aligned} \tag{4.1}$$

I note that this is a small systematic effect that would not affect any of my conclusions.

of its sparse coverage of B type stars. In addition, those works that derive both $T_{\rm eff}$ and A_V simultaneously (e.g. Anders et al., 2019) suffer from the degeneracy between those parameters, especially for objects that do not lie on the main sequence and suffer circumstellar extinction. The Tycho-2 Spectral Type Catalogue (Wright et al., 2003), which includes some of the better characterised spectral types catalogues, was also discarded as each of its spectral sub-types exhibit a large spread in the $G_{BP} - G_{RP}$ colour. This indicates that the catalogued objects would require major extinction corrections, or have a large uncertainty in their spectral types.

After considering this, I decided that the Bright Star Catalogue (BSC, Hoffleit and Jaschek, 1991) was the best set of stars to use. The BSC lists the spectral types of 9110 of the brightest stars in the sky. This catalogue is not ideal as Gaia suffers significantly from saturation for sources brighter than $G < 6$, whereas the optimal range is for objects with $G > 13$ (Riello et al., 2018). However, because they are bright, they are typically not far away and hence are not expected to suffer much from extinction. This is further evidenced by the small scatter of the BSC spectral sub-types in the $G_{BP} - G_{RP}$ colour (see Fig 4.3).

Some dwarf BSC objects appear in the sub-giant or red-giant branch. Therefore, the luminosity class classification of the BSC is somewhat affected by reddening and erroneous determinations. Thus, I use the theoretical main-sequence location to select the set of BSC objects to use for deriving the Gaia $G_{BP} - G_{RP}$ intrinsic colours. The PARSEC 1.2S isochrones of Bressan et al. (2012) and Marigo et al. (2017) have a label that describes the evolutionary stage of each point, and those corresponding to a main sequence stage were used to define a theoretical main sequence in the Gaia $G_{BP} - G_{RP}$ vs. absolute G magnitude diagram (see Fig 4.3). I used 4000 isochrones of solar metallicity from log(Age) = 1.00 to log(Age) = 10.48 yr in steps of 0.0025 dex. This set of isochrones is used throughout the whole chapter. The passband profiles of Maíz Apellániz and Weiler (2018) were used (in this case with the blue sensitivity curve for bright objects, see Fig. 4.4).

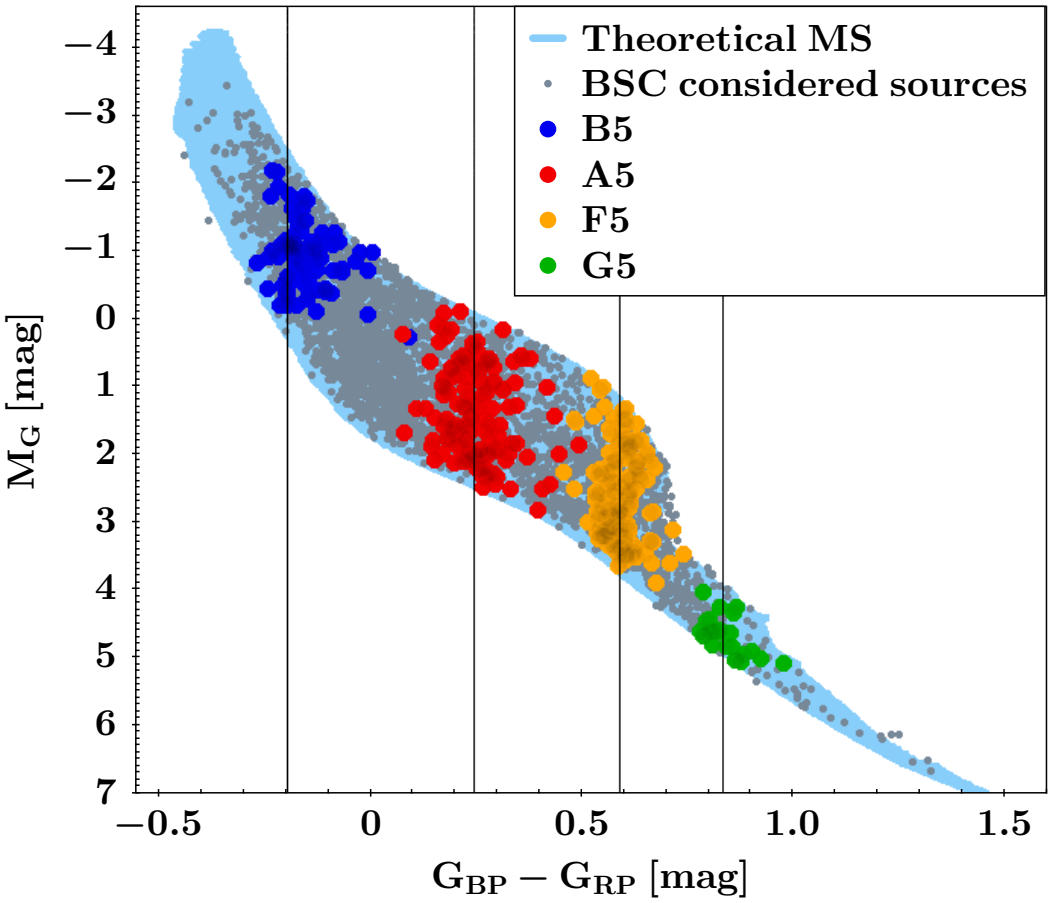


Fig. 4.3 Gaia colour vs. absolute magnitude diagram of the Bright Star Catalogue sources considered for deriving Gaia $G_{BP} - G_{RP}$ intrinsic colours (in grey). Only four spectral sub-types are highlighted for clarity. Black lines trace the derived intrinsic colours for those sub-types (B5: −0.197, A5: 0.245, F5: 0.591, G5: 0.834 mag). Note that for B5 a posterior reddening correction was applied (see Sect. 4.2). The theoretical main sequence from PARSEC 1.2S isochrones appears in light blue (Marigo et al., 2017).

BSC sources with incorrect Gaia photometric determinations according to the criterion introduced in Chapter 2 (Eq. 2.5) from Gaia Collaboration et al. (2018a) were also removed. In total, 3765 BSC sources have a good photometry according to this constraint, have Gaia distances from Bailer-Jones et al. (2018), and are on the theoretical main sequence according to the PARSEC 1.2S isochrones. This is the set of sources I use to derive the Gaia $G_{BP} - G_{RP}$ intrinsic colours and it is shown in the Gaia colour vs. absolute magnitude diagram of Fig. 4.3. No constraints to the astrometric quality were applied as only one of the BSC sources used has

$\varpi/\sigma(\varpi) < 5$. I note that 16% of the used BSC sources have RUWE ≥ 1.4, and hence their astrometry might be unreliable according to this Gaia astrometric indicator. However, the parallax determination does not affect the colour (i.e. the astrometric solution is independent of the photometry) and for this exercise I chose to keep these objects for their colour information. This is because the exact absolute magnitudes of the used BSC sources are not relevant (see Fig. 4.3). As the selection of the theoretical main sequence is already an arbitrary cut in absolute magnitude, I expect the potential impact of these sources with a possible unreliable astrometry to be very minor.

Artificially biasing the empirical observations to those objects in theoretical main sequence positions seems to contradict the previous discussion. However, after inspection I concluded that this approach is less affected by uncontrolled uncertainties and biases than to use the luminosity class classification of the BSC.

I took the median $G_{BP} - G_{RP}$ and $B-V$ value of each BSC spectral subtype. The median colours obtained this way are presented in Table 4.2 (see also Fig. 4.3). The obtained $B-V$ median colours are compared with the intrinsic colours of Pecaut and Mamajek (2013). This allows me to assess the quality of the selected set of BSC sources for spectral typing. The result is that the difference between the two is below 0.03 mag for almost all spectral subtypes (44/49, see 'residual' column of Table 4.2). This implies that the selected BSC sources constitute a photometrically well-behaved sample which is good for deriving intrinsic colours. A 0.03 mag error is good enough for the purposes of this chapter; for example, the mass as determined from the isochrones and Gaia colour throughout the chapter is typically affected by $\lesssim 2\%$ under $G_{BP} - G_{RP} = 0.03$ mag deviations.

Finally, the observed median $G_{BP} - G_{RP}$ colour was taken directly as the intrinsic colour (see Fig. 4.3). However, I do note a systematic deviation in the aforementioned residual values for O and B objects (see Table 4.2). As these sources are typically further away this is likely caused by reddening. Therefore, O and B Gaia $G_{BP} - G_{RP}$

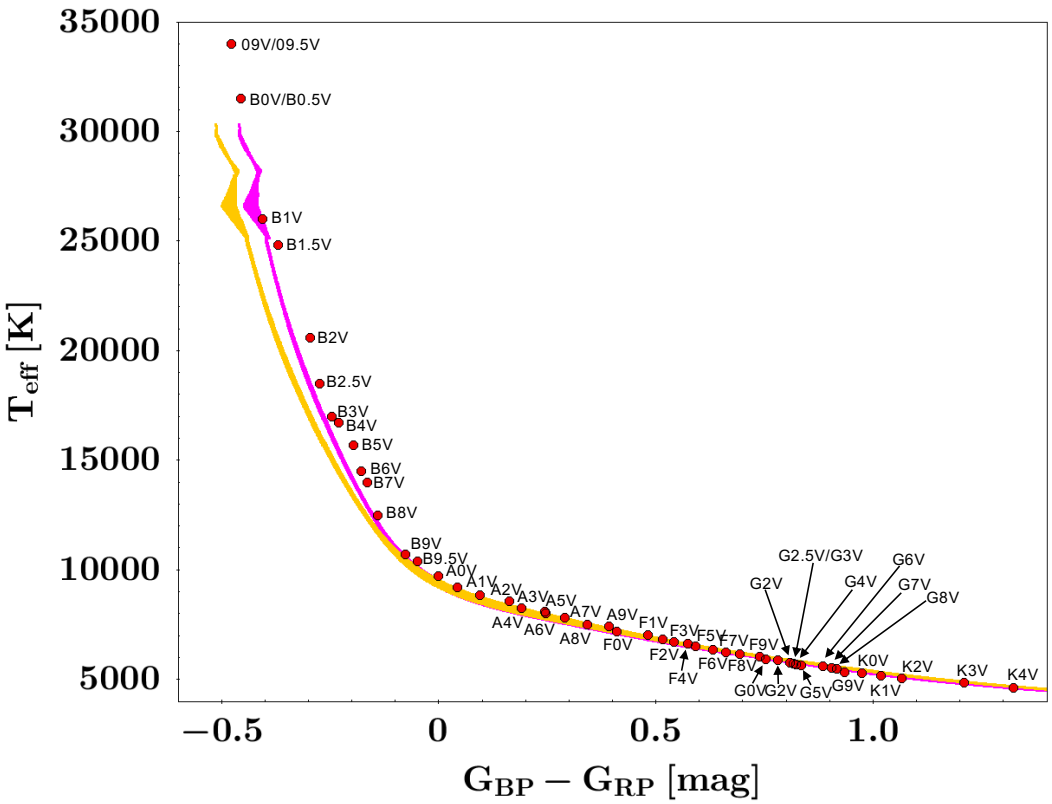


Fig. 4.4 Gaia DR2 $G_{BP} - G_{RP}$ intrinsic colours for dwarf stars of different spectral types. Magenta and yellow lines trace the theoretical main sequence colours using the bright and faint G_{BP} sensitivity passband profiles of Maíz Apellániz and Weiler (2018) respectively. Note the difference between empirical and theoretical colours in the B1.5V to B7V range.

median colours were corrected for reddening using the residual values. For this I used the E$(B-V)$ to E$(G_{BP} - G_{RP})$ conversions of Casagrande and VandenBerg (2018). In this latter case, these colours that were corrected for reddening were chosen to be the intrinsic colours. The resulting intrinsic colours are tabulated in Table 4.2 and shown in Fig. 4.4, in which the theoretical main sequence intrinsic colours arising from the two Maíz Apellániz and Weiler (2018) G_{BP} sensitivity curves are also plotted. The sensitivity curve for bright objects was used to constrain the BSC set of sources used to derive the Gaia $G_{BP} - G_{RP}$ intrinsic colours and the sensitivity curve for faint objects ($G > 10.87$) is used hereafter in the chapter. The difference between the two passband profiles is significant for the early type regime

Table 4.2 Gaia $G_{BP} - G_{RP}$ median and intrinsic colour for each spectral sub-type from Bright Star Catalogue selected objects. The $B-V$ median colour from the Bright Star Catalogue is also derived and compared with the intrinsic $B-V$ colours of Pecaut and Mamajek (2013, PM13). The difference between these latter two appear in the fifth column. Temperatures from PM13 are listed in the seventh column.

SpType	Gaia DR2$_{med}$ $G_{BP} - G_{RP}$	BSC$_{med}$ $B-V$	PM13$_{int}$ $B-V$	Residual (mag)	Gaia DR2$_{int}$ $G_{BP} - G_{RP}$	PM13 T_{eff} [K]
O9/O9.5V	-0.374*	-0.24*	-0.318	0.078	-0.478*	34000
B0/B0.5V	-0.366	-0.24	-0.307	0.067	-0.455	31500
B1V	-0.315	-0.21	-0.278	0.068	-0.406	26000
B1.5V	-0.281	-0.185	-0.252	0.067	-0.371	24800
B2V	-0.255	-0.18	-0.21	0.03	-0.295	20600
B2.5V	-0.237	-0.17	-0.198	0.028	-0.274	18500
B3V	-0.209	-0.15	-0.178	0.028	-0.246	17000
B4V	-0.196	-0.14	-0.165	0.025	-0.230	16700
B5V	-0.162	-0.13	-0.156	0.026	-0.197	15700
B6V	-0.146	-0.115	-0.14	0.025	-0.179	14500
B7V	-0.126	-0.10	-0.128	0.028	-0.163	14000
B8V	-0.102	-0.08	-0.109	0.029	-0.141	12500
B9V	-0.064	-0.06	-0.07	0.01	-0.077	10700
B9.5V	-0.035	-0.04	-0.05	0.01	-0.048	10400
A0V	0.000	-0.01	0.00	-0.01	0.000	9700
A1V	0.044	0.02	0.043	-0.023	0.044	9200
A2V	0.096	0.06	0.074	-0.014	0.096	8840
A3V	0.162	0.11	0.09	0.02	0.162	8550
A4V	0.191	0.15	0.14	0.01	0.191	8270
A5V	0.245	0.17	0.16	0.01	0.245	8080
A6V	0.246	0.19	0.17	0.02	0.246	8000
A7V	0.290	0.20	0.21	-0.01	0.290	7800
A8V	0.343	0.24	0.25	-0.01	0.343	7500
A9V	0.392	0.28	0.255	0.025	0.392	7440
F0V	0.410	0.28	0.294	-0.014	0.410	7200
F1V	0.481	0.34	0.334	0.006	0.481	7030
F2V	0.515	0.36	0.374	-0.014	0.515	6810
F3V	0.541	0.39	0.389	0.001	0.541	6720
F4V	0.574	0.41	0.412	-0.002	0.574	6640
F5V	0.591	0.43	0.438	-0.008	0.591	6510
F6V	0.632	0.47	0.484	-0.014	0.632	6340
F7V	0.662	0.49	0.51	-0.02	0.662	6240
F8V	0.694	0.53	0.53	0.00	0.694	6150
F9V	0.740	0.55	0.552	-0.002	0.740	6040
G0V	0.753	0.58	0.588	-0.008	0.753	5920
G1V	0.781	0.62	0.604	0.016	0.781	5880

Table 4.2 continued.

SpType	Gaia DR2 $G_{BP} - G_{RP\mathrm{med}}$	BSC $B - V_{\mathrm{med}}$	PM13 $B - V_{\mathrm{int}}$	Residual (mag)	Gaia DR2 $G_{BP} - G_{RP\mathrm{int}}$	PM13 T_{eff} [K]
G2V	0.808	0.63	0.65	-0.02	0.808	5770
G2.5/G3V	0.817	0.64	0.661	-0.021	0.817	5720
G4V	0.822	0.65	0.674	-0.024	0.822	5680
G5V	0.834	0.67	0.68	-0.01	0.834	5660
G6V	0.885*	0.72*	0.704	0.016	0.885*	5590
G7V	0.905*	0.73*	0.713	0.017	0.905*	5530
G8V	0.917	0.73	0.737	-0.007	0.917	5490
G9V	0.935*	0.73*	0.777	-0.047	0.935*	5340
K0V	0.974	0.805	0.816	-0.011	0.974	5280
K1V	1.019	0.84	0.847	-0.007	1.019	5170
K2V	1.067	0.87	0.893	-0.023	1.067	5040
K3V	1.210	0.995*	0.99	0.005	1.210	4840
K4V	1.322*	1.105*	1.1	0.005	1.322*	4620

Note: *Values obtained with fewer than 5 objects. 44 spectral types have residuals below 0.03 mag. Those who do not are O9/O9.5, B0/B0.5, B1, B1.5 and G9. O and B type $G_{BP} - G_{RP}$ median colours were corrected from extinction using the residuals to derive the corresponding intrinsic colours; for the other spectral types the median $G_{BP} - G_{RP}$ colour equals the intrinsic $G_{BP} - G_{RP}$ colour.

and it exemplifies a problem with the Gaia photometric system and the requirement of using only empirically obtained intrinsic colours.

In Fig. 4.4 the empirically derived intrinsic colours match the theoretical values for most of the spectral types, which is another positive assessment of the quality of the derived intrinsic colours (although there is an induced bias by demanding the BSC sources to be on theoretical MS locations, see Fig. 4.3). However, there is a small offset of a few tens of milli-magnitudes in the B1.5V to B7V range. This offset could be caused by a flaw in the theoretical isochrones but also by an incomplete reddening correction or even by photometrically problematic BSC sources. As no intrinsic colour is bluer than theoretically expected, I have reasons to believe that an incomplete reddening correction is the more likely explanation.

I wish to stress that the intrinsic colours derived in this section are not optimal, but they are the first empirical determinations. Probably in the future superior derivations

will be achieved from a better behaved set of objects in the adequate magnitude range for the Gaia data.

Finally, it is worth remarking that I use intrinsic colours for dwarf sources throughout the chapter. Although this is correct for most CBe sources, Pecaut and Mamajek (2013) already noted that this may be inaccurate for PMS stars. Nonetheless, the error introduced is often insignificant when compared to other uncertainties, and most of the observed objects are indeed of a dwarf type.

4.3 HR diagram and stellar parameters

Using the intrinsic colours obtained in Sect. 4.2 and the temperatures and spectral types of Tables C.1 and C.4 I derived the intrinsic $G_{BP} - G_{RP}$ colours of the observed sources described in Sect. 4.1. For the sources which spectral types were determined I took one spectral sub-type as uncertainty to derive the error on the intrinsic colour. For those sources which effective temperatures were directly derived, the T_{eff} uncertainty was used, with a minimum value of one spectral sub-type (using the Pecaut and Mamajek, 2013 values; see Table 4.2). Then, extinctions were derived using the Gaia observed colours ($A_G = 2.046\,\mathrm{E}(G_{BP} - G_{RP})$; Casagrande and VandenBerg, 2018). Finally, I used those extinctions and the Gaia G magnitude and parallax to obtain absolute magnitudes (using the parallax to distance conversion of Bailer-Jones et al., 2018).

This could be done for all sources but for five PMS candidates. Gaia DR2 4318785810234714752, 506799479443438080 and 5333545642950621696 could not be spectral typed, mainly because of their strong emission line spectrum. In addition, stars cooler than K4 could not be assigned an intrinsic colour because there are no sources in the BSC with those spectral types in the demanded MS positions (see Sect. 4.2 and Fig. 4.3). Thus, the observed PMS candidates Gaia DR2 4319343855414540 80 and 2071705173505640448, which are a late-K/M and

M1 star respectively, do not have intrinsic colours tabulated in Table 4.2 and no extinction or absolute magnitude were derived for them. I note that these five objects are very likely contaminants: Gaia DR2 4318785810234714752 is a known planetary nebula (and was tagged with a 'PN' flag in my catalogue) and 506799479443438080 appears as dwarf nova in SIMBAD (from e.g. Sahman et al., 2015 or Dubus et al., 2018). Gaia DR2 5333545642950621696 has a PN warning flag in my catalogue and might be a B[e] (FS CMa) star. This is supported by the extremely strong emission it shows in all H lines observed (see Sect. 4.5). Gaia DR2 431934385541454080 and 2071705173505640448 are probably evolved stars given their low temperature and low absolute magnitude (high luminosity) in the HR diagrams of Chapter 3.

Gaia 4094703381988286592 was observed with both IDS and EFOSC2 settings and was assigned a B1.5 and a B3 spectral type respectively. Given the better resolution of the IDS configuration and its smaller spectral type uncertainty I used the B1.5 spectral type for this object in the derivations of this section.

In the end, 140 PMS candidates and 14 CBe candidates can be placed accurately on the Gaia Hertzsprung-Russell diagram. This is done in Fig. 4.5. I note that all of them have a $\varpi/\sigma(\varpi) \geq 5$ and only two have a RUWE > 1.4. These are Gaia DR2 5884829984147822976 and 5544618168572613504, which are highlighted in Fig. 4.5. Therefore, the indicated absolute magnitudes for those two objects might be spurious.

In Fig. 4.5 pre-main sequence tracks of 1, 2, 4, and 7 $M_{\odot}$ are shown together with the theoretical main sequence described in Sect. 4.2. In Fig. 4.5 it can be seen that most observed PMS candidates are massive hot objects. In addition, many occupy clear early-stage PMS locations, whereas some of them are closer to the ZAMS. No PMS candidate is located outside PMS positions in this HR diagram, which resembles very much that of Fig. 2.1 in Chapter 2 for the known Herbig Ae/Be stars. I note that from the PMS tracks of Fig. 4.5 most of the observed PMS candidates are very strong intermediate mass Herbig Ae/Be candidate stars, with

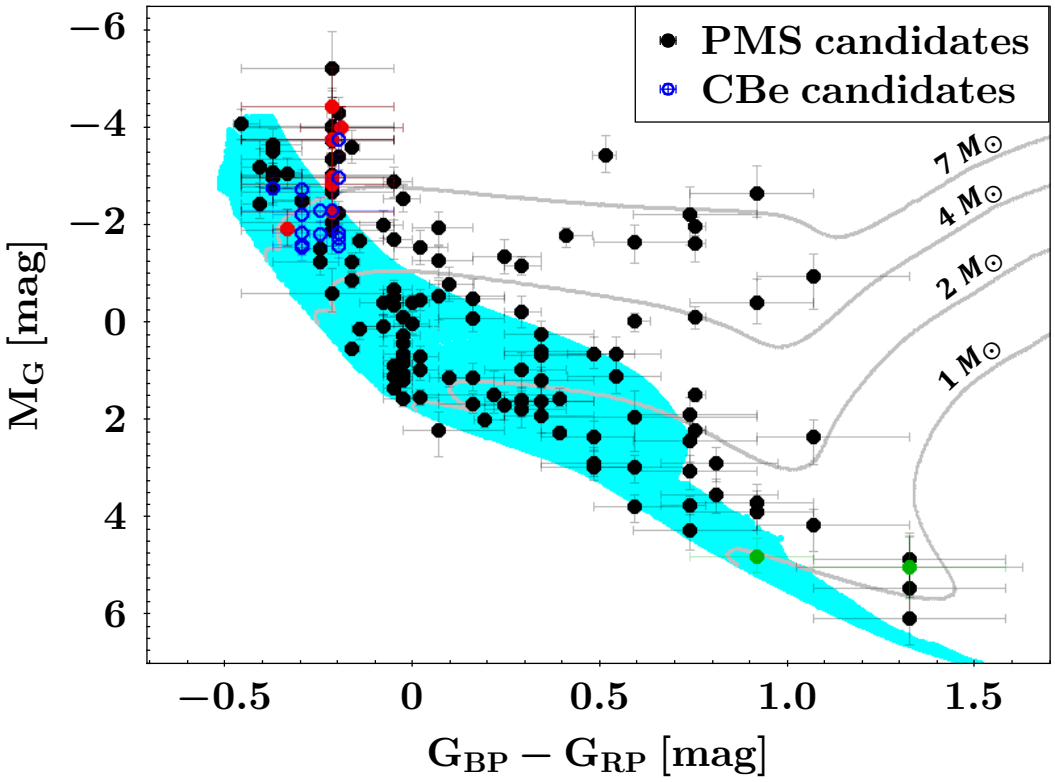


Fig. 4.5 Gaia colour vs. absolute magnitude diagram of the observed candidates. PMS candidates (140) appear as full circles and CBe candidates (14) appear as blue open circles. In green the two PMS candidates with an unreliable astrometric solution. In red those PMS candidates with a 'PN' warning flag in the catalogue of Chapter 3. The theoretical main sequence and the pre-main sequence tracks appear in blue and grey respectively (from Marigo et al., 2017). PMS tracks correspond to 1, 2, 4, and 7 $M_{\odot}$.

the exception of a handful that are placed around the 1 $M_{\odot}$ track. As in the HR diagrams of Chapter 2 for the known HAeBes, most massive sources appear close to the ZAMS whereas the scatter along PMS tracks is larger for low mass objects. However, around 10% of the observed candidates appear located at high-mass tracks (above 4 $M_{\odot}$) and at early stages of evolution (i.e, far away from the ZAMS, see Fig. 4.5). These are studied in more detail in Sect. 4.6. In addition, in Fig. 4.5 I highlight the seven PMS candidates with a 'PN' warning flag. These are discussed further in Sect. 4.5.

Most classical Be candidates observed appear on B main sequence positions in the HR diagram of Fig. 4.5. Only two appear in sub-giant locations, which

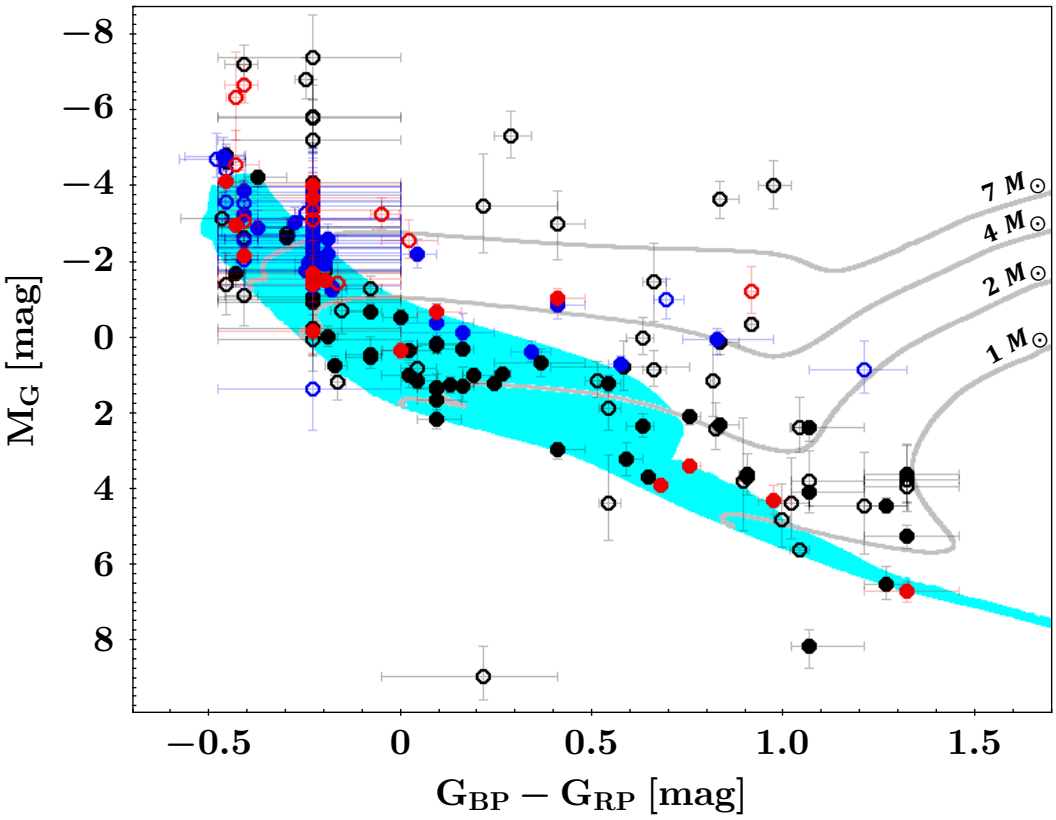


Fig. 4.6 Gaia colour vs. absolute magnitude diagram of candidates with spectral types from SIMBAD. PMS candidates (103) appear in black, CBe candidates (56) in blue and Either candidates (28) in red. Open circles are those sources which do not satisfy the $\varpi/\sigma(\varpi) \geq 5$ and RUWE < 1.4 astrometric condition. The theoretical main sequence and the pre-main sequence tracks appear in blue and grey respectively (from Marigo et al., 2017). PMS tracks correspond to 1, 2, 4, and 7 $M_\odot$.

have also been reported for several known classical Be stars (Rivinius et al., 2013, Bodensteiner et al., 2020a; see also Chapter 3).

In addition, I derived extinctions and absolute magnitudes for the sources in the SIMBAD set (Sect. 4.1.3) as was done for the sample of observed objects. The corresponding HR diagram for these SIMBAD objects is shown in Fig. 4.6. In this case many sources have poor astrometry, and only 47/103 PMS, 37/56 CBe, and 17/28 Either candidates have $\varpi/\sigma(\varpi) \geq 5$ and RUWE < 1.4. As the sources with poor astrometry typically have too large uncertainties in the HR diagram, hereafter I only consider the sources of the SIMBAD set that satisfy those astrometric constraints. These sources are highlighted in Fig. 4.6.

In Sect. 4.1.3 the heterogeneity and problematic of the SIMBAD database was discussed. An example of this is the nine CBe candidates of the SIMBAD set that appear outside of CBe locations in the HR diagram when using the SIMBAD spectral type information (see Fig. 4.6). These nine sources do appear in CBe positions in the HR diagrams of Chapter 3, where only the observed Gaia passbands and distances were used (see Fig. 3.9). In addition, seven of them have CBe probabilities above 70% in my catalogue. Therefore, these nine CBe candidates probably have an inaccurate or erroneous spectral type classification in SIMBAD. Obtaining a new spectrum of those would help in settling the matter.

Nevertheless, the candidates of the SIMBAD set in Fig. 4.6 follow the general trends observed for the sources with our own spectral determinations (Fig. 4.5): most PMS candidates are hot, massive and appear in PMS locations on the HR diagram. This has an added value as for this set I did not bias the selection towards massive objects. In addition, most CBe candidates are of B type and have luminosities compatible with those of known classical Be stars.

I derived stellar parameters (mass, luminosity, and age) for all sources placed in the HR diagrams of Figs. 4.5 and 4.6. For doing this I used the PARSEC 1.2S isochrones introduced in Sect. 4.2. I set a maximum of 50% relative error for the derived uncertainties. This is to avoid unrealistically large uncertainties for the sources with large spectral type uncertainties. Only isochrones with a pre-main sequence label were used for the PMS candidates, and only isochrones with a main sequence, sub-giant branch or RGB label were used for the CBe candidates. For the sources of the Either catalogue I applied both sets of isochrones independently and hence obtained two sets of stellar parameters. The difference between these sets of parameters can be significant. For example the PMS isochrones result in a mass typically $28 \pm 13\%$ larger than that predicted by the non-PMS isochrones.

The derived extinctions, absolute magnitudes, masses, luminosities, and ages are presented in Tables C.3 and C.6 for all the observed PMS and CBe candidates,

respectively. I also include in those tables the Gaia parallaxes and distances. I note that the ages are very model dependent, are based on an arbitrary decision of the age 'zero' and, as discussed in Chapter 2, are very susceptible to the HR diagram uncertainties. The parameters used in the following sections from the PMS, CBe, and Either candidates of the SIMBAD set are tabulated in Tables D.1, D.2, and D.3, respectively.

4.4 Hα line and variability trends

In this section I discuss how the Hα line properties discussed in Sect. 4.1.2 correlate with the stellar mass parameter derived in Sect. 4.3 and the G_{var} observable.

I correct the measured Hα EWs of Tables C.2 and C.5 for the underlying line absorption for the 95 PMS candidates and 14 CBe candidates with Hα line observations discussed in Sect. 4.1.2. For this I use the typical absorption EW values of each spectral sub-type (from Joner and Hintz, 2015). These corrected Hα equivalent widths for both sets of candidates are shown in Fig. 4.7 as a function of mass and are tabulated in Tables C.2 and C.5. In this figure I add the Hα EWs of the known Herbig Ae/Be stars from the high-quality sample of Chapter 2, which also have been corrected for the line absorption. For the PMS candidates a similar trend to what was found in Chapter 2 for the known HAeBes is also present here, with low-mass stars typically showing low levels of emission and high-mass stars displaying both high and low Hα EWs. At each mass range the median EW values are similar for both the PMS candidates and the known Herbig Ae/Be stars. The CBe candidates show levels of emission comparable to those of the HAeBes of the same mass.

In Fig. 4.8 I plot the Hα line shape information of the observed candidates (excluding the 12 clear contaminants discussed in Sect. 4.5) as a function of mass and G_{var}. The left panel of this figure can be compared with its equivalent for the known Herbig Ae/Be stars in Fig. 2.9 in Chapter 2. As in Chapter 2, low-

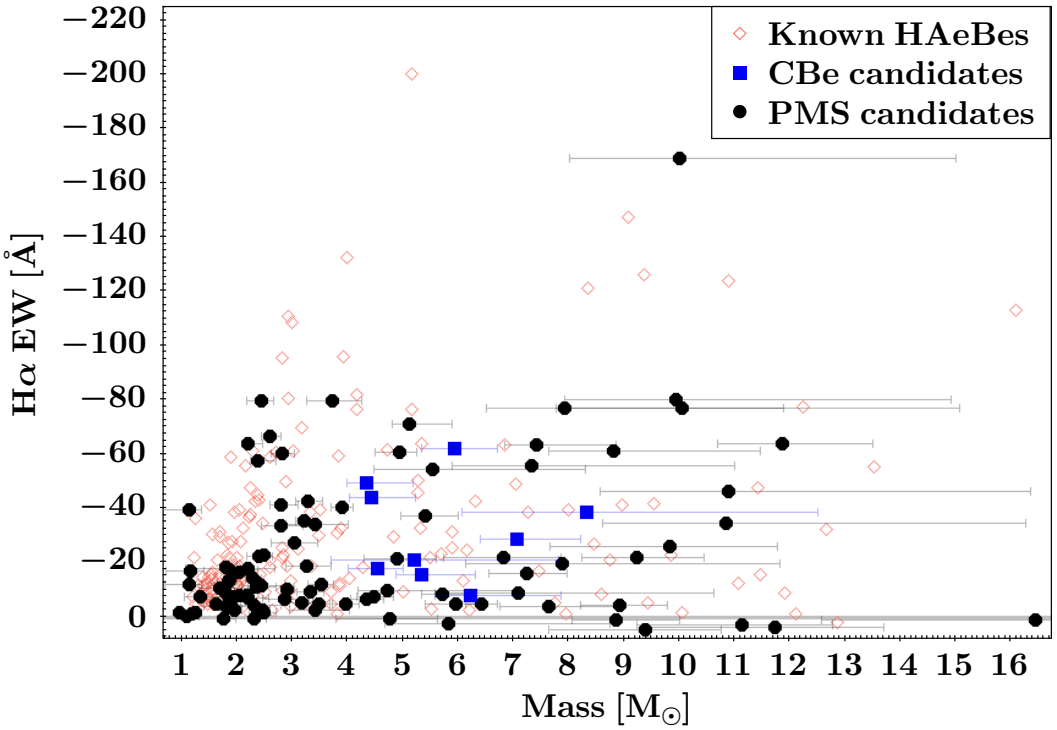


Fig. 4.7 Mass vs. Hα equivalent width of the observed candidates (excluding the 12 clear contaminants discussed in Sect. 4.5). Red diamonds trace the known Herbig Ae/Be stars of the high-quality sample of Chapter 2 within the same mass range. The equivalent widths have been corrected for the line absorption and have a typical uncertainty of 10% (not including the line variability). The grey line traces EW = 0 Å, and thus separates the sources with Hα emission from those without it.

mass PMS stars show both low and high levels of variability; this variability range reduces gradually from ~ 4 $M_\odot$, with a break appearing at ~ 7 $M_\odot$ from where more massive PMS stars only show low levels of variability. I note that there are four PMS candidates which appear to be very variable and massive. The fact that they are clearly deviating from the general trend and that they all have the Hα line in absorption or with very little emission ($EW_{obs} = -1.28$ Å for the one with a double-peaked Hα line profile) indicates that they are likely contaminants (see Sect. 4.5 for further discussion on the nature of these objects). The right panel of Fig. 4.8 shows that the CBe candidates have a different behaviour; the more massive in general appear more variable.

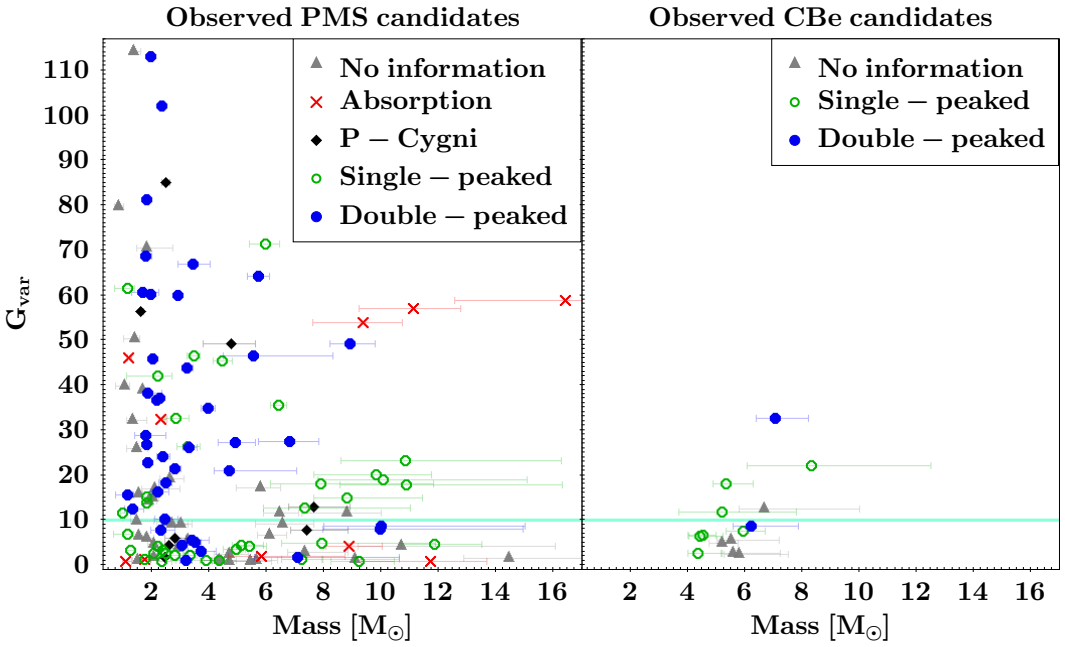


Fig. 4.8 Mass vs. photometric variability as traced by the variability indicator G_{var}. Different symbols indicate different Hα line profiles. The horizontal line marks the $G_{var} = 10$ threshold from which G_{var} is sensitive to disc orientation (see discussion in Chapters 2 and 3). *Left*: 133 observed PMS candidates (I have excluded the 12 clear contaminants discussed in Sect. 4.5). *Right:* 14 observed CBe candidates.

In Chapter 2 it was shown that the photometric variability as traced by the Gaia proxies I have developed along this thesis is correlated with the Hα line profiles. In particular, it was discussed that above a certain variability threshold (equivalent to $G_{var} = 10$), Herbig Ae/Be stars tend to have double-peaked Hα line profiles. This indicates that the traced variability is mainly caused by edge-on discs. As observed for the known HAeBes of Chapter 2, it can be seen in the left panel of Fig. 4.8 that the amount of optical variability for the observed PMS candidates is correlated with the Hα profiles. Above $G_{var} = 10$, the proportions of the observed profiles change to 34% single-peaked, 58% double-peaked, and 8% P-Cygni (from 44%, 47%, and 9% respectively, see Sect. 4.1.2). However, the increase of double-peaked profiles at high variabilities is not as large as for the previously known Herbig Ae/Be stars (see Chapter 2). As discussed in Sect. 4.1.2, this is mainly because of the low resolution of my observations and the presence of contaminants.

4.5 Contaminants

The analysis of the previous sections allow identifying several potential contaminants among the observed objects. From the study of the line profiles in both the blue ($\sim 3300-5400$ Å) and red ($\sim 5800-7000$ Å) spectral ranges eight sources were catalogued as new 'unclassified B[e]' (FS CMa) discoveries. These are listed in Tables C.1 and C.2.

To this list I add the planetary nebula and the dwarf nova discussed in Sect. 4.3 for a total of ten contaminants that were identified because of their strong emission line spectra. Seven of these contaminants have a 'PN' warning flag in the catalogues of Section 3 (see Table C.1). This flag traces possible contamination by planetary nebulae or B[e] stars. Therefore, the PN flag is highly efficient at tracing strong emission line contaminants. Of the seven PMS candidates in Fig. 4.5 with a PN warning flag only two, Gaia DR2 2200017424528999936 and 205118464010485632, could not be discarded from the PMS group.

The set of detected contaminants is completed with the two evolved stars identified from their HR diagram locations that were discussed in Sect. 4.3. In total, from the observations 12 PMS candidates appear as clear contaminants. These 12 sources are marked in Tables C.1 and C.2.

In principle, Herbig Ae/Be stars are defined as with the presence of emission lines. Thus, the ten PMS candidates with the Hα line in absorption (see Sect. 4.1.2) could be considered potential contaminants. This is further evidenced by the three very massive and variable sources without Hα emission in the left panel of Fig. 4.8. These ten objects are marked in Tables C.1 and C.2 with an 'abs' comment.

To conclude the analysis of contaminants, I examine in detail the observed candidates with a SIMBAD object type that were not flagged as possible contaminant in the previous paragraphs. Only three appear to be probable contaminants. The double-peaked source in the left panel of Fig. 4.8 which is also placed at

atypically high mass and variability values is Gaia DR2 2035402872974695936, a well known Cepheid variable (e.g. Schmidt et al., 2004). In addition, Gaia DR2 1828276425855506304 is very likely a carbon star (e.g. Groenewegen et al., 2002), and Gaia DR2 4253225574005033088 is a long-time considered eclipsing binary (e.g. Jiang et al., 2012). These three sources are also marked in Tables C.1 and C.2.

None of the 14 classical Be candidates observed can be discarded as contaminant, not even from their position in the HR diagram (see Fig. 4.5).

In total I have identified 25 possible contaminants among the observed PMS candidates (12+10+3), of which 12 appear as clear contaminants in my observations because of strong emission line spectra or HR diagram locations. This number of possible contaminants (25/145 or 17%, see Table 4.1) is consistent with the estimated precision of the catalogue of PMS candidates ($P \sim 81\%$). These contaminants span the whole range of observed PMS probabilities (Fig. 4.1). This implies that those probabilities mainly indicate the confidence of the algorithm at separating CBes from HAeBes, which was the main goal of Chapter 3, but they are not sensitive to other types of contaminants.

4.6 Discussion of the observations

In the previous sections I discuss the results of the observations of a sample of 145 PMS and 14 CBe candidates. The objectives of the observations were to derive accurate stellar parameters and to assess the accuracy of the catalogues constructed in Chapter 3. In this section, I summarise the main results and conclusions of those observations.

4.6.1 Observed PMS candidates

In Sect. 4.5 25 possible contaminants were found among the 145 PMS candidates observed. The HR diagram with the remaining 120 PMS objects is shown in Fig. 4.9. As in Fig. 4.5, most of the observed PMS candidates are hot and massive. In Fig. 4.9 the 218 known Herbig Ae/Be stars of the high-quality sample of Chapter 2 are also plotted. Most of the observed PMS candidates are distributed in the HR diagram similarly to the known HAeBes. However, the are some massive candidates at early stages of evolution (i.e. far away from the ZAMS) that are younger than most previously known Herbig Ae/Be stars of the same mass. I note that these less evolved Herbig Ae/Be candidates are significantly more variable than the candidates of the same mass that are closer to the ZAMS. This could be caused by these younger sources hosting larger discs than their more evolved counterparts, which already have suffered from the fast dust dispersal mechanisms discussed in Chapter 2. No clear trend in the IR colours is detected.

In the bottom panel of Fig. 4.9 I present the mass distribution of the 120 PMS candidates. In this panel I also plot the mass distribution of the 218 known Herbig Ae/Be stars. The first thing to note is that the observed PMS candidates have masses that correspond to the Herbig Ae/Be category. In particular, 42 of them are above the 4 $M_{\odot}$ threshold of Wichittanakom et al. (2020) for the break in accretion properties (from magnetospheric accretion to the still unknown accretion mechanism for high-mass stars, see Chapter 1), and 21 are above the 7 $M_{\odot}$ threshold of Chapter 2 (from which different or differently acting disc-dispersal mechanisms appear). Therefore, the PMS candidates observed and characterised in this chapter are indeed new strong Herbig Ae/Be candidates of the greatest interest for understanding the mechanisms of high-mass star formation. It is noteworthy that in Fig. 4.9 no very high-mass PMS candidates ($M > 15$ $M_{\odot}$) were observed, whereas there are several known HAeBes at this mass range. This is because the PMS candidates with those masses, in addition to

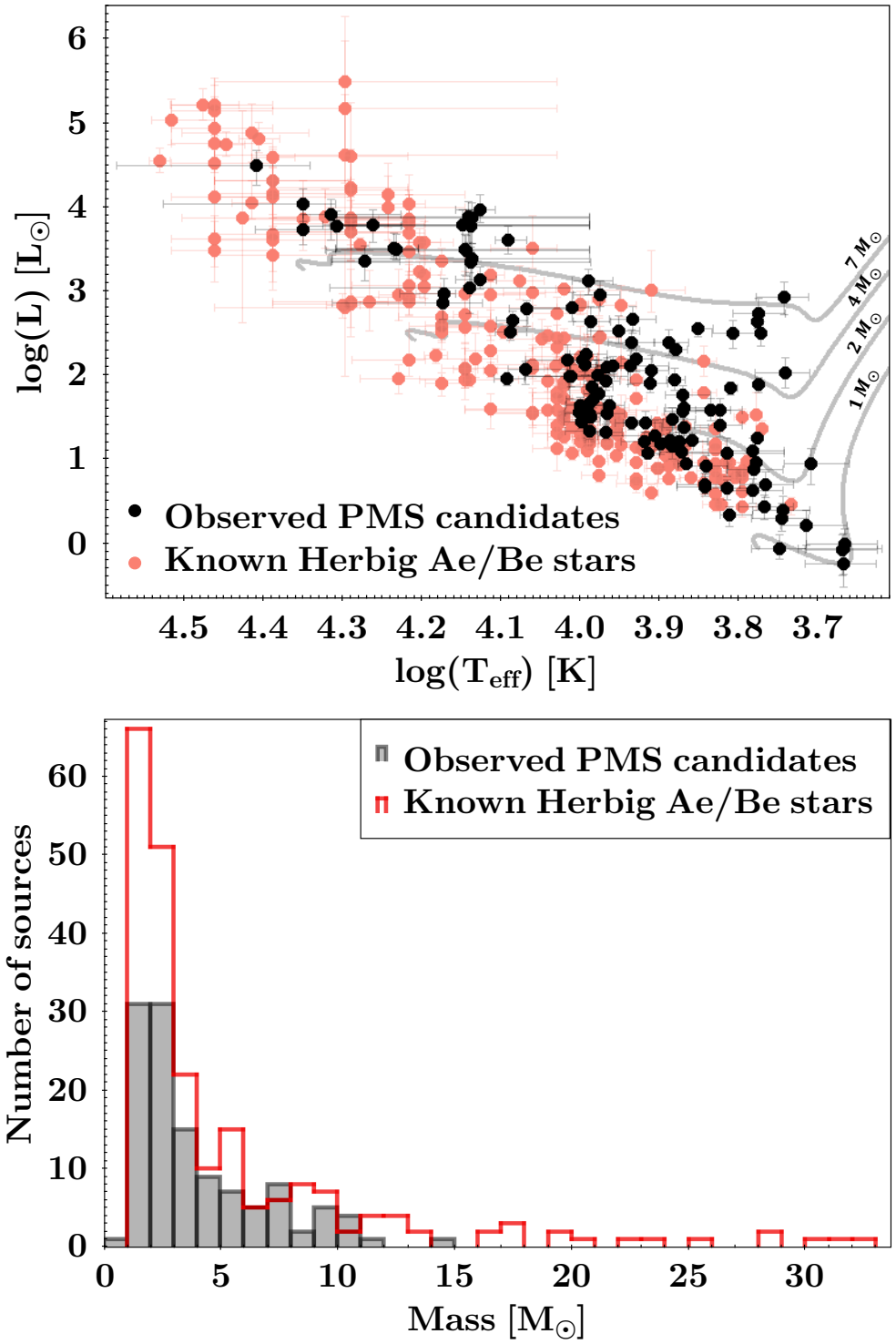


Fig. 4.9 *Top*: Hertzsprung-Russell diagram of the 120 observed PMS candidates that could not be discarded as possible contaminants. The 218 known Herbig Ae/Be stars of the high-quality sample of Chapter 2 are also shown. PMS tracks correspond to 1, 2, 4, and 7 $M_{\odot}$ (from Bressan et al., 2012 and Marigo et al., 2017). *Bottom*: Histogram of the number of objects in the sets of the top panel per 1 $M_{\odot}$ mass bin.

being rare and extreme, are typically at large distances and thus normally have poor parallaxes. Therefore, they were excluded from the target list of the observations (Sect. 4.1).

It is beyond the capabilities of the data presented in this chapter to assert with absolute certainty whether the observed Herbig Ae/Be candidates are indeed new Herbig Ae/Be discoveries. In fact, it even proved difficult and controversial for much more intensely studied objects (e.g. HD 45677, Oudmaijer and Miroshnichenko, 2017). Nevertheless, in this and the previous sections I have provided ample and independent evidence to conclude that, at least, a large fraction of the proposed objects have a Herbig Ae/Be nature. This can be summarised by their PMS position in the HR diagram (Figs. 4.5 and 4.9) and the presence of emission lines in their spectra (Hα in particular, Fig. 4.7). In addition, similar trends to those found in Chapter 2 for the known HAeBes were found for the observed Herbig Ae/Be candidates. In particular, I have found the same correlation of the stellar mass with the Hα EW and the photometric variability (Figs. 4.7 and 4.8, respectively), and also the same correlation between the Hα line profiles and the photometric variability (Fig. 4.8).

4.6.2 Observed classical Be candidates

Regarding the 14 classical Be candidates observed, they were all classified as of B type and are placed at the typical HR diagram location of classical Be stars. None could be identified as a possible contaminant. In addition, the nine that were observed in the red spectral range show Hα in emission, and of the five that were only observed in the blue spectral range, four show Hβ in emission.

Variability in CBes has been studied by a number of authors (e.g. Kurtz et al., 2015, Baade et al., 2018, Borre et al., 2020 or Neiner et al., 2020), and its underlying cause is complex. However, the variability is believed to be mainly related to mass

changes in the disc and stellar pulsations. Therefore, it should not be strongly dependent on the disc orientation. In Chapter 2 I showed that variability proxies like G_{var} trace edge-on discs in PMS sources, as it is indicated by the higher incidence of double-peaked Hα line profiles for variable sources. In Fig. 4.8 the observed CBe candidates lack of this correlation with the Hα line profile. This is another argument in favour of their correct classification. In addition, the CBe candidates of Fig. 4.8 display an increasing photometric variability with mass, the opposite trend to what is expected for PMS objects (see Chapter 2).

Most CBe stars are known to be binaries (Klement et al., 2019); and their gaseous, dust-free, and diffuse discs and the Be phenomena are commonly explained by the interaction with the companion (see e.g. Bodensteiner et al., 2020b or El-Badry and Quataert, 2020). Unresolved binaries might appear as variable in the variability indicators of this thesis. This is because the moving photocenter of an unresolved binary system might result in large Gaia flux uncertainties (see definition of G_{var}, Eq. 3.1, in Chapter 3). Therefore, a part of the variability I am tracing for classical Be stars might be due to binarity.

Further evidence for CBe candidates being correctly classified was given to me by Mr. C. Wichittanakom from the analysis of the INT data. Of the 65 sources observed in that observing run, 19 show absorption line broadening consistent with a fast rotational velocity. All nine CBe candidates observed in that run but Gaia DR2 5224710591951790 08 are in this group of fast rotators. Fast rotational velocity is one of the defining characteristics of the CBe class (see e.g. Rivinius et al., 2013). Six other fast rotators are among the 120 Herbig Ae/Be candidates previously discussed in this section. The last five were tagged as possible contaminants in Sect. 4.5 because of having the Hα line in absorption (I note that only ten objects show Hα in absorption). These latter five sources are all of B type, and it is tempting to consider that the observed line broadening might have been caused by unresolved binarity. A more detailed analysis of these sources is planned for the future.

As for the observed PMS candidates, it is not possible to state the category of each observed CBe candidate with an absolute certainty, but the previous discussion is consistent with the lower limit to the precision of $P \sim 89\%$ derived for the catalogue of CBe candidates (see Chapter 3). Therefore, the independent observations described in this chapter lead me to conclude that the algorithm is very efficient finding classical Be stars, and separating classical Be stars from Herbig Ae/Be stars.

4.6.3 General remarks of the observations

In Sect. 4.1 I mentioned that objects with different warning flags were intentionally observed to study the accuracy of those flags. I conclude that the 'PN' flag is very efficient at tagging extreme emission line objects which are not PMS, in particular planetary nebulae and B[e] (FS CMa) stars. I note that at the mass range of Herbig Ae/Be stars, this flag mainly traces contamination by B[e] stars. The source with both 'ID AllW' and 'ID IPH/VPH' flags, Gaia DR2 2208194286334441344, is not an incorrect cross-match in AllWISE or IPHAS. Finally, 72% of the sources selected as possible contaminants in Sect. 4.5 have the 'W3W4' flag, whereas this proportion is of 46% for the candidates that could not be marked as possible contaminants. The 'W3W4' flag is too general and frequent to be traced appropriately, but it would seem from these proportions that sources with this flag are more likely to be contaminants.

Finally, in Sect. 4.1.3 I presented the SIMBAD set, a set of sources that were not observed by myself but which have spectral types in the literature and could be placed in the HR diagram of Fig. 4.6. Therefore, this is a random subsample of the catalogues of Chapter 3, which in general has a greater degree of uncertainties and biases than the subsample of sources selected for my own observations. However, in Sect. 4.3 I discussed that most of the PMS and CBe candidates of this SIMBAD set seem to be of a Herbig Ae/Be and CBe nature respectively, according to their positions in the HR diagram. Moreover, the HR diagram of the SIMBAD set is

similar to the HR diagram of the observed candidates (Figs. 4.5 and 4.6 respectively). This is important because it evidences that the conclusions of this section are not biased by the decisions made when constructing the target list (which, for example, selected some of the best candidates; see Sect. 4.1).

In conclusion, the independent observations described in this section and the previous ones provide further support to the accuracy and high-quality of the catalogues of new PMS and CBe sources presented in Chapter 3. The degree of contamination by other sources in the observations described in this chapter is similar to what it was proposed in Chapter 3 for the general catalogues. In addition, I have accurately characterised the stellar parameters of a large set of new probable Herbig Ae/Be and classical Be stars.

4.7 Analysis of the catalogues

In this section I discuss the general catalogues obtained in Chapter 3. The sample of PMS candidates has 8470 sources, the classical Be candidate star sample contains 693 sources while the list of sources with a probability of larger than 50% of belonging to one of the two categories (Either group) has 1309 sources.

Following the discussion of Chapter 3, for this section I excluded objects which have a 'X-mtch' warning flag as they are likely wrong cross-matches. In addition, I removed those tagged as 'G-UMAP' as they are likely evolved contaminants (the G-UMAP flag indicates the sources that the unsupervised UMAP algorithm found as likely not PMS, see Sect. 3.4.6). I did not exclude those marked as 'PN', 'W3W4', 'ID AllW' or 'ID IPH/VPH' because that would remove many correctly classified sources. However, I note that in the previous section it was concluded that the PN flag is very efficient at tracing contaminants. This exercise results in a sample of 8248 PMS candidates, 636 CBe candidates and 1264 sources belonging to the Either group. These different sets are summarised in Table 4.3. Not all the

Table 4.3 Number of sources per catalogue for different thresholds on the astrometric quality.

Catalogue	Total	With parallax (ϖ) RUWE < 1.4	$\varpi/\sigma(\varpi) \geq 5$ RUWE < 1.4	$\varpi/\sigma(\varpi) \geq 10$ RUWE < 1.4
PMS	8248	6381	3369	2038
CBe	636	578	462	230
Either	1264	1094	707	397
E_D/D	-	0.35 ± 0.40	0.101 ± 0.070	0.058 ± 0.028

Note: The bottom row shows E_D/D, the mean relative error in the distance for the different astrometric constraints (error is one standard deviation of the mean).

objects in these catalogues have parallax detections, and those with parallax have different astrometric qualities. In Table 4.3 the mean relative error in the distance is indicated for different astrometric constraints. By constraining the parallax quality with $\varpi/\sigma(\varpi) \geq 10$ we ensure that the uncertainties in the distance are below 10% (which is why this constraint was used in Chapter 3). However, as discussed in Sect. 4.6.1, applying a strong constraint on the parallax quality comes at the price of excluding many sources, especially massive ones that are typically further away. In contrast, using a more relaxed constraint introduces a larger uncertainty on the parameters that are derived from HR diagram locations. The number of sources within each set that satisfy the different astrometry quality constraints discussed in this thesis are summarised in Table 4.3.

4.7.1 General considerations

In Chapter 3 it was discussed that the main observables that can be used to separate PMS from CBe objects are the mid-IR colours because of the different nature of their circumstellar discs. In Fig. 4.10 the mid-IR $W1-W4$ colour is plotted against the near-IR ($J-K_s$), photometric variability (G_{var}), and $r-H\alpha$ observables for the three catalogues (PMS, CBe, and Either; see Table 4.3). The sample of the spectroscopically observed sources described in the previous sections is also shown.

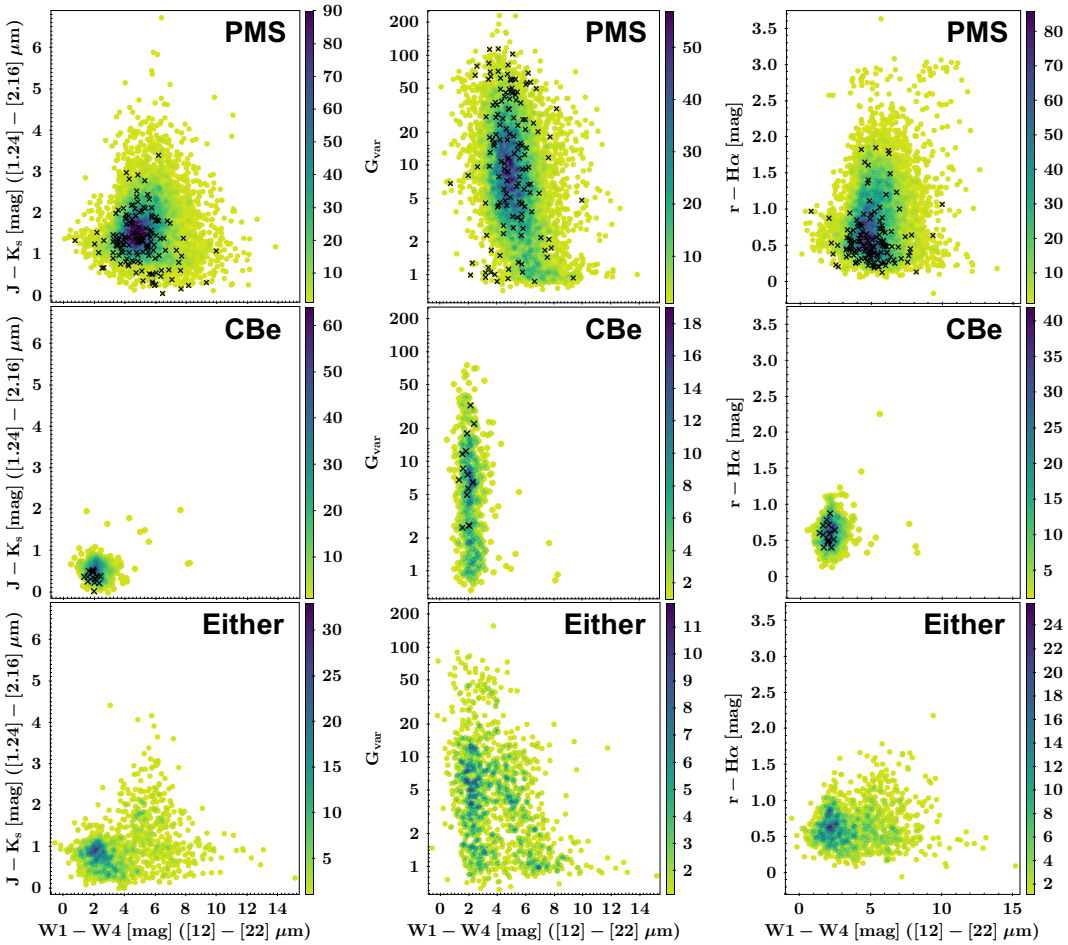


Fig. 4.10 $W1-W4$ vs. $J-K_s$, G_{var}, and $r-H\alpha$ for the sample of PMS candidates (top), CBe candidates (middle) and Either group (bottom). The crosses mark the observed candidates (not including the SIMBAD set). In the case of $r-H\alpha$ the PMS candidate Gaia DR2 2069751818049277824, with $r-H\alpha = 7.39$, has been removed from the plot for visualisation purposes. It has a 'PN' warning flag and it is likely a contaminant.

It is important to assess whether the spectroscopically observed sources are a representative subsample of the general catalogues. In Fig. 4.10 they roughly match the distribution of all sources in the space of observables. The observed PMS candidates also cover the whole probability range (see Fig. 4.1). Therefore, it could be said that the observed PMS candidates described in Sect. 4.6.1 are a representative selection of the entire catalogue of PMS objects. In contrast, only CBe candidates with high ($> 85\%$) probability were observed. Nonetheless, from Fig. 4.10 it is fair to say that the CBe candidates observed are not extreme in any observable, and thus

they can also be considered as a representative subsample of the catalogue of CBe objects. I thus conclude that it is possible to work with the catalogues of PMS and CBe candidates constructed in Chapter 3 with a high degree of confidence. This conclusion is based on the independent assessment of their high accuracy provided by a representative subsample of them (see Sect. 4.6).

In addition, in order to have a complete view of the high-mass pre-main sequence candidates it is necessary to take into account the Either catalogue. This is because the PMS candidates obtained tend to have large $W1-W4$ colours (see Fig. 4.10), whereas most Either sources appear at lower $W1-W4$ values. In Chapter 2 I demonstrated that massive Herbig Be stars dissipate their discs much faster than their lower mass counterparts, and hence typically show lower $W1-W4$ values. In this respect, in the Gaia colour vs. absolute magnitude diagram of Fig. 4.11 we can see that most of the high-mass sources are in the Either category (and thus show low IR excess, see Fig. 4.10). Therefore, the Either catalogue probably contains most of the more evolved new Herbig Ae/Be stars with dissipated discs. In contrast, many of the newly classified high-mass PMS stars from the PMS catalogue must either be younger or have suffered different, less efficient, dispersion mechanisms than most of the previously known Herbig Ae/Be stars. I note that in Fig. 4.11 only interstellar extinction corrections have been applied to the full samples (using the dust map of Lallement et al., 2019), so all sources could move significantly to brighter and bluer positions if corrected from the total extinction.

4.7.2 Stellar parameters

In Sect. 4.3 I used total extinction corrections to determine HR diagram positions and derive stellar parameters. In addition, I used the uncertainty on the total extinction to derive error bars for those stellar parameters. However, for the general catalogues (PMS, CBe, and Either) I only have the interstellar extinction A'_G described in Chapter

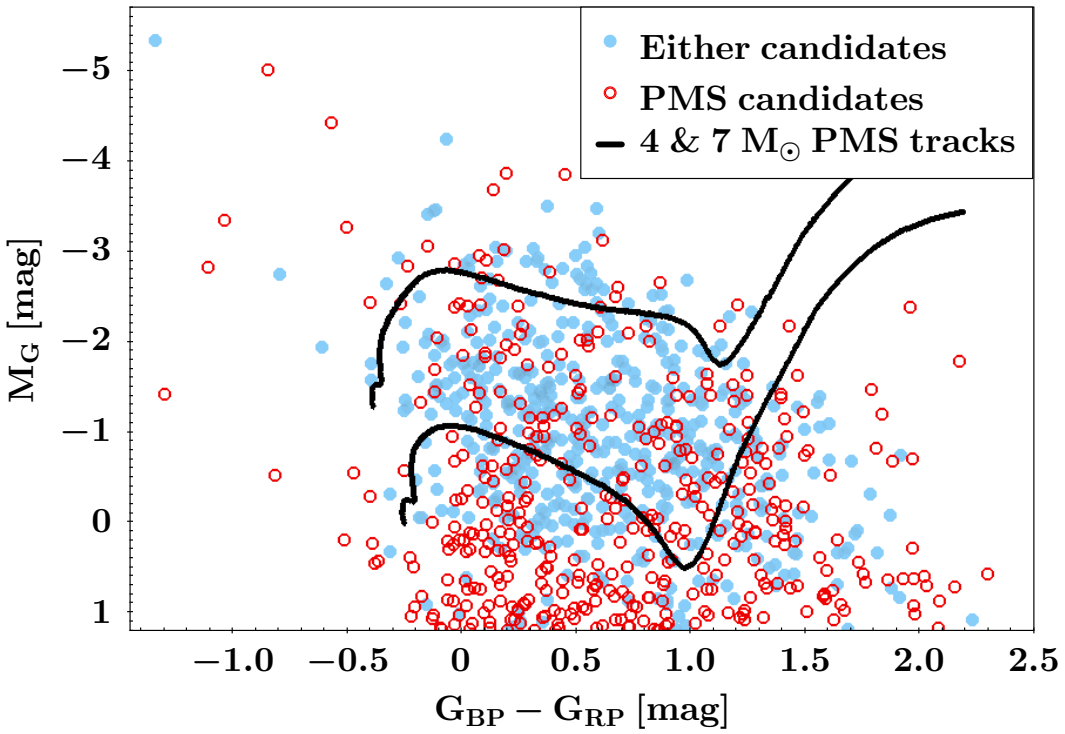


Fig. 4.11 Gaia colour vs. absolute magnitude diagram of high-mass PMS and Either sources with RUWE < 1.4 and $\varpi/\sigma(\varpi) \geq 5$. This plot contains $\sim 71\%$ of the Either sources with these astrometry constraints. Only corrections to the interstellar extinction were applied. Two PMS tracks from Marigo et al. (2017) appear in black. One track indicates the threshold at ~ 4 $M_\odot$ where Wichittanakom et al. (2020) found a break in accretion properties. The other track indicates the threshold found at ~ 7 $M_\odot$ in Chapter 2, from which more efficient disc-dispersion mechanisms were detected.

3 from the dust map of Lallement et al. (2019). Therefore, by placing the PMS, CBe, and Either catalogues in the HR diagram (see e.g. Fig. 4.11 or Fig. 3.9 in Chapter 3) I can only derive lower limits to the stellar mass, effective temperature, and luminosity. This is because to correct from circumstellar extinction would only make the sources bluer and brighter. I use the same PARSEC 1.2S isochrones of Sect. 4.2. The main parameters controlling the uncertainty on a star's position in the HR diagram are the parallax and the extinction. In this case, the extinction obtained from the dust map of Lallement et al. (2019) depends on the distance and therefore its uncertainty is also heavily influenced by the parallax. Therefore, it is necessary to impose some constraints to the parallax quality. Contrary to what was

done in Chapter 3, I derive stellar parameters for sources with RUWE < 1.4 and $\varpi/\sigma(\varpi) \geq 5$ (instead of $\varpi/\sigma(\varpi) \geq 10$, see Table 4.3) to be able to study massive objects.

I compare the (lower limit) stellar parameters obtained this way with those obtained in Sect. 4.3 for the objects I observed and those of the SIMBAD set (see Sect. 4.1). This is done for the mass parameter in Fig. 4.12. The main result is that most masses obtained from HR diagrams that were only corrected from interstellar extinction are indeed lower limits to the masses obtained from HR diagrams that were corrected from total extinction. The stars for which this is not the case have incorrectly large interstellar extinctions. This is because the spatial resolution of the Lallement et al. (2019) dust map is quite large and might lead to erroneous extinctions. Similar trends are found with the other stellar parameters and, as expected, the $T_{\rm eff}$ and L values are in general lower limit derivations.

It is useful to quantify the typical difference between the lower limit approximations of this section and the values obtained in Sect. 4.3 for the observed objects. In Fig. 4.12 the identity line is indicated, together with lines that trace a 20% deviation from it. 122 out of 248 sources ($\sim 49\%$) fall within these 20% deviation lines, which means that for them the mass lower limit is a reasonably good approximation. This percentage goes up to $\sim 74\%$ for sources below 4 $M_\odot$. Some sources, mainly belonging to the SIMBAD set, are very massive and appear in Fig. 4.12 with very large error bars. These are mostly candidates with a generic 'B' spectral type, that were assigned a central $T_{\rm eff}$ value with generous error bars.

Therefore, it can be concluded that the masses, luminosities, and temperatures derived in this section for the general catalogues are, in general, robust lower limits. Furthermore, a large fraction of these stellar parameter determinations are reasonably close to the ones obtained when the total extinction is known. In general, more massive stars tend to behave worse in this respect (see Fig. 4.12). This is because they are often further away and thus have poorer parallaxes and perhaps incomplete

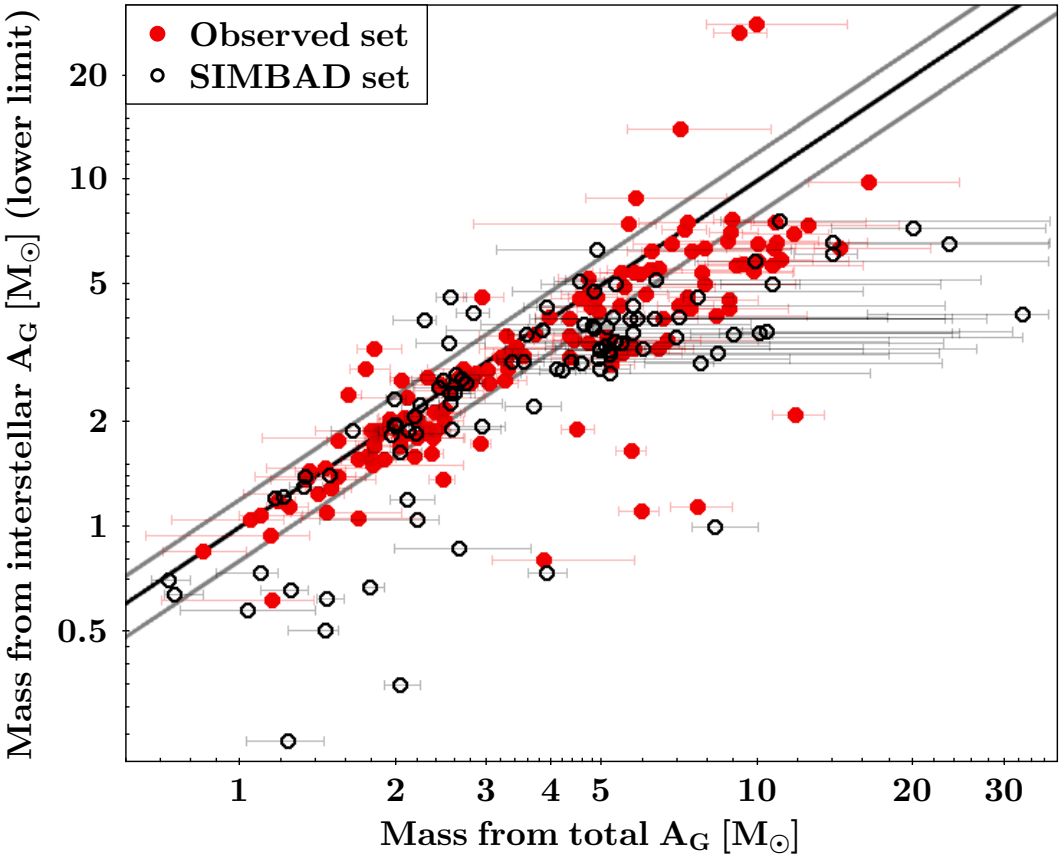


Fig. 4.12 Mass (using extinctions determined from spectra) vs. mass lower limit (using interstellar extinctions determined from the dust map of Lallement et al., 2019). The identity line is shown in black. Grey lines trace 20% deviations from the identity line.

interstellar extinctions (the dust map of Lallement et al., 2019 typically does not extend beyond 3 kpc).

The distribution of the mass lower limits obtained for the PMS, CBe, and Either catalogues is shown in Fig. 4.13. The derived mass lower limits imply that there are at least 129 PMS candidates with $M > 4$ $M_\odot$ (and an additional 317 in the Either group). These sources are interesting for the study of the accretion properties of the high-mass regime, as these objects are expected to not undergo magnetospheric accretion (Wichittanakom et al., 2020). The steep decline of new discoveries at around 7 $M_\odot$ (see Fig. 4.11 and Fig. 4.13) can be explained by the lack of previously known Herbig Ae/Be stars with non-dispersed discs that are far from ZAMS positions.

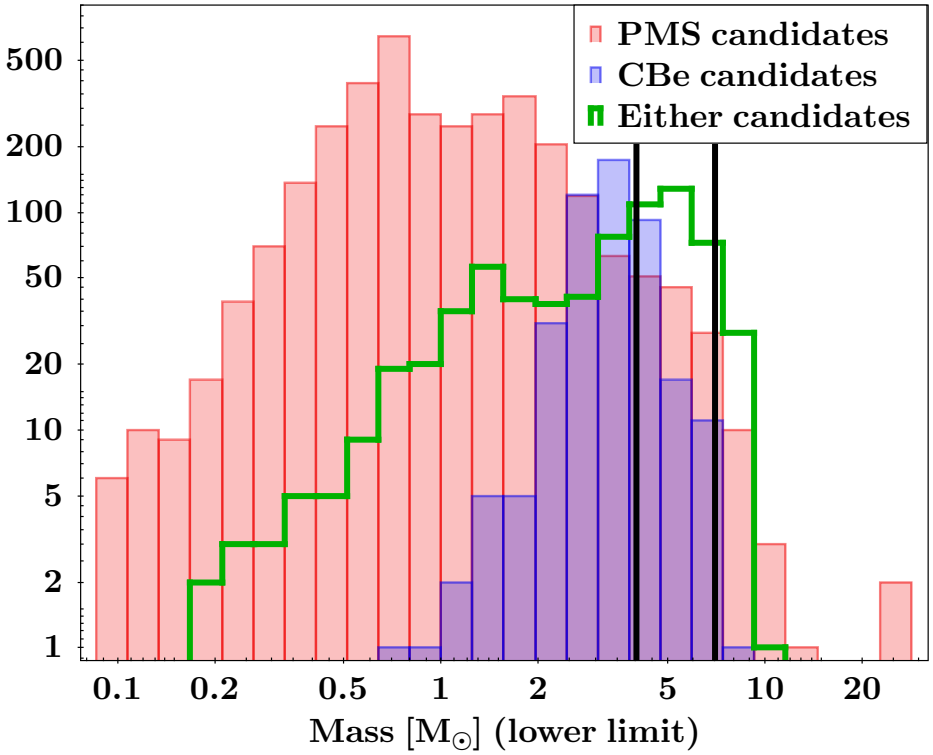


Fig. 4.13 Histogram of the number of sources from the PMS, CBe, and Either catalogues at different masses (lower limit). Note that both axes are logarithmic. The masses of the Either catalogue were obtained from PMS isochrones. The black vertical lines are the two mass thresholds at 4 and 7 $M_\odot$ where, respectively, the break in accretion and disc-dispersal properties are expected to occur.

This meant that very few objects in the training set with $M > 7$ $M_\odot$ were emphasised for the algorithm by the observables used in Chapter 3. The constraint on the parallax quality removes many massive objects that as a result are not considered in Fig. 4.13 or for the statistical purposes of this chapter (see Table 4.3). However, these massive sources might be of great interest for future surveys.

The mass distribution of the CBe catalogue in Fig. 4.13 shows that most CBe candidates appear in a narrow range between $\sim 2.5 - 5$ $M_\odot$, which is precisely the mass range that corresponds to main sequence B stars (according to Marigo et al., 2017 isochrones and Table 4.2). This, in addition to further confirming the accuracy of the CBe classification, implies that the circumstellar extinction of these objects is very minor, as is typical for CBe stars (e.g. Chapter 3). Fig. 4.13 also evidences that

the Either group peaks at a high-mass where the PMS and CBe catalogues contain much fewer sources. This supports the discussion of Sect. 4.7.1 which stated that most of the massive sources are in the Either group. These massive PMS objects in the Either group are those more similar to CBe stars and to most of the previously known Herbig Be stars of Chapter 2.

Some sources appear outside the theoretical tracks in the HR diagram and it proved not possible to obtain an estimate of their stellar parameters. Sources with too large extinction corrections might appear in unphysical positions to the left of the MS. In addition, some sources are too red for the tracks, implying they might be evolved contaminants. However, some of these latter sources might also be very young objects with inadequate extinction corrections. The number of objects without derived stellar parameters are 115, 2, and 62 for the PMS, CBe, and Either catalogues respectively.

Mass completeness and the IMF

In Chapter 2 it was mentioned that one of the main biases of the previously known set of Herbig Ae/Be stars is the existence of a gap between T Tauri and Herbig Ae stars because of historical selection effects. Using the derivations of this section I assess whether the new catalogue of PMS candidates covers this artificial gap and smoothly transits between high- to low-mass PMS objects, including the elusive Intermediate-Mass T Tauri stars. In Fig. 4.14 I show the mass distribution of the previously known Herbig Ae/Be stars of the high-quality sample of Chapter 2 (see discussion in Sect. 2.4.2), together with the mass distribution (lower limit) of the new PMS candidates. Two lines that trace the same initial mass function (IMF) from Kroupa (2001) are plotted. They are normalised to the last increasing bin of each set. As discussed in Chapter 2, the previously known HAes are consistent with the IMF above ~ 2 M$_\odot$, but the range between $1-2$ M$_\odot$ is rather incomplete. In contrast,

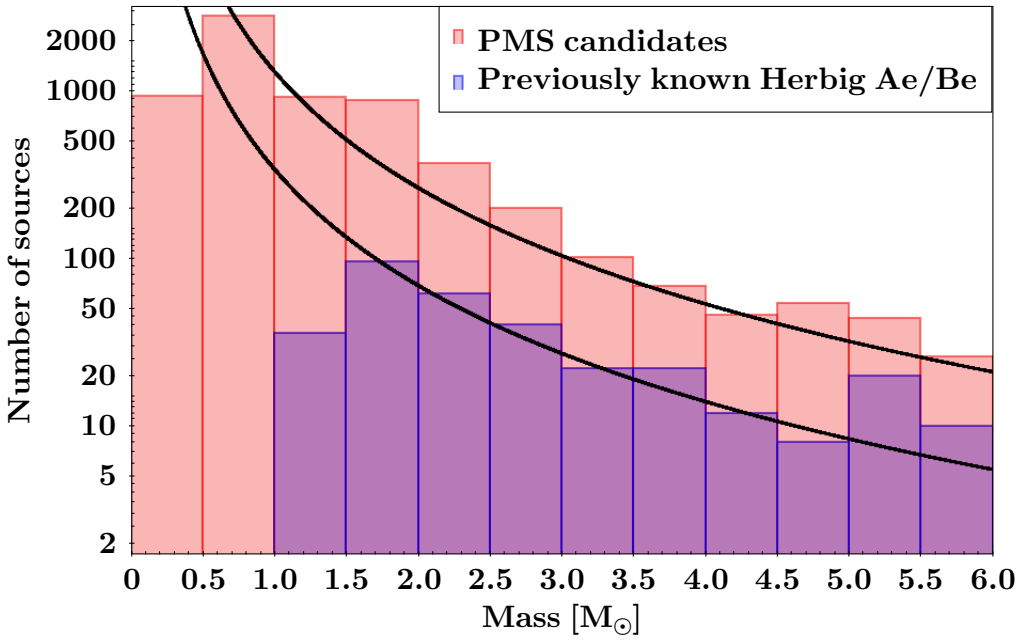


Fig. 4.14 Mass distribution up to 6 $M_\odot$. In blue the known Herbig Ae/Be stars of the high-quality sample of Chapter 2, and in red the new PMS candidates with lower limits to the mass. Black lines trace the initial mass function of Kroupa (2001), normalised to the last increasing bin of each set.

the new catalogue of PMS candidates is consistent with the IMF up to the $0.5-1$ $M_\odot$ bin. Therefore, the new catalogue links the HAeBe stars with the T Tauri stars continuously, without artificial breaks. It is remarkable that this set adjusts nicely to the IMF even though its masses are only lower limits. This further supports the previous discussion about the accuracy of the stellar parameter determinations of this section, especially for objects with $M \lesssim 4$ $M_\odot$ (see Fig. 4.12).

Therefore, the new catalogue of PMS sources does not suffer from the historical bias the previously known Herbig Ae/Be stars have at the low-mass end. This was achieved because the algorithm of Chapter 3, and thus the classification, are independent of the mass, spectral type or temperature.

4.7.3 Hα line emission

In order to study the Hα line emission properties of the PMS, CBe and Either catalogues I converted the IPHAS or VPHAS+ $r-H\alpha$ and $r-i$ colours to Hα equivalent widths (EW_{obs}, observed above the continuum), using the models of Drew et al. (2005). It is not necessary to deredden the $r-H\alpha$ and $r-i$ colours as the reddening line moves parallel to that of equal EW_{obs} (Drew et al., 2005). This conversion can only be done effectively for objects approximately hotter than 5900 K (see Chapter 3). Hence, the EW_{obs} derived for cooler stars can be considered as an upper limit (see Fig. 6 of Drew et al., 2005). Nevertheless, the error introduced is expected to be as much as ~ 20 Å. I provide EW_{obs} for all the PMS, CBe and Either sources for which the $r-H\alpha$ and $r-i$ colours are available from the cross-matches of Chapter 3 (8179, 623, and 1239 sources respectively). Only lower limits to the EW_{obs} could be derived for 68 of the PMS candidates because of their large $r-H\alpha$ colour (these have a 'PN' warning flag in the catalogue and are probably contaminants).

To assess this methodology I compare the EW_{obs} obtained in this section with those measured in Sect. 4.1.2 from the independent spectroscopic observations. This comparison is shown in Fig. 4.15, where it can be seen that the agreement between the two derivations is remarkably good for most objects (within ± 10 Å, which is the typical Hα variability of Herbig Ae/Be stars; Mendigutía et al., 2011b). In this comparison the ten strong emission line contaminants described in Sect. 4.5 were excluded.

I also add to Fig. 4.15 the Hα EW_{obs} measured in Wichittanakom et al. (2020) for some of the known HAeBes of Chapter 2. In this case, the Hα EW_{obs} obtained from $r-H\alpha$ and $r-i$ colours were derived after a 1 arcsecond cross-match with IPHAS and VPHAS+. For this set the agreement between methodologies is also good, although there is a larger discrepancy for high-emission objects ($EW_{obs} < -100$ Å), a regime that my observations barely covered. The scarcity of points in this region

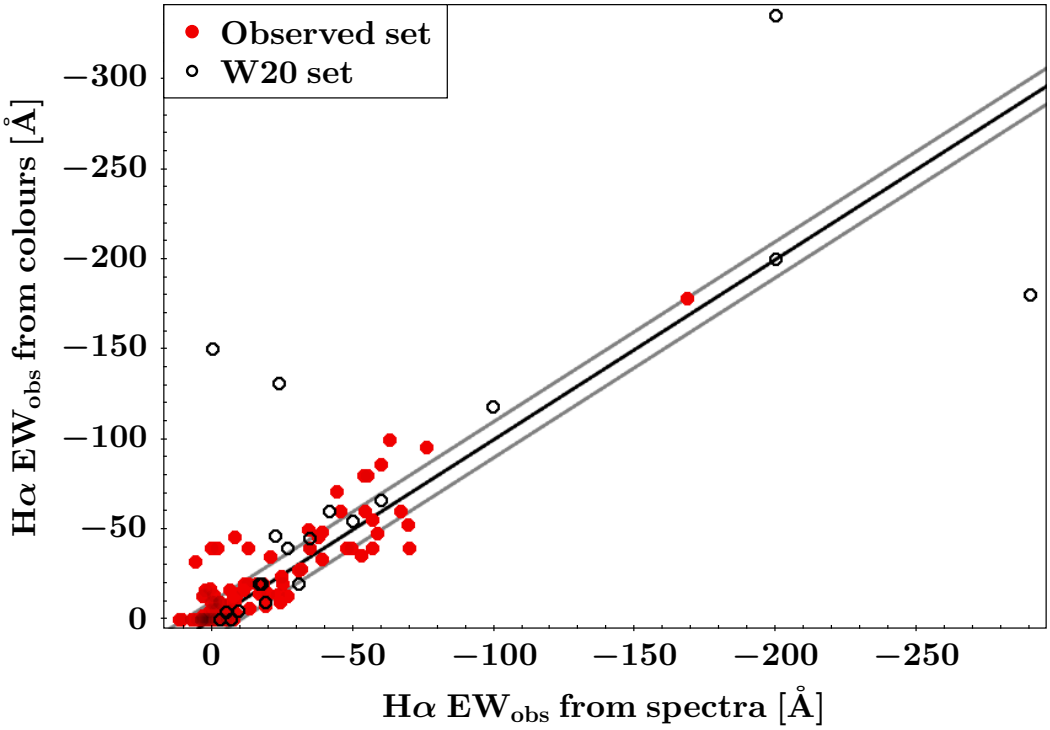


Fig. 4.15 Comparison of the Hα EWs measured from independent spectroscopic observations (excluding the ten strong emission line contaminants of Sect. 4.5) with the Hα EWs determined from the $r-H\alpha$ and $r-i$ colours. The identity line is shown in black. Grey lines trace ±10 Å deviations from the identity line. The W20 set is formed by those known HAeBes with Hα EWs from Wichittanakom et al. (2020).

make it hard to investigate what the reason might be behind this discrepancy, but the extreme emission of these objects suggests that they might have strong Hα variability.

The number of points which are deviant beyond ±10 Å is 39 out of 114 (34%), probably caused by Hα variability. Therefore, I conclude that the Hα EW$_{obs}$ that can be obtained for the general catalogues from the $r-H\alpha$ and $r-i$ colours are in general a good approximation to the Hα EW$_{obs}$ that can be measured from spectra.

4.7.4 Observed trends in the catalogues

In Sects. 4.7.2 and 4.7.3 I have shown that the derived stellar parameters and Hα EW$_{obs}$ for the PMS, CBe, and Either catalogues are adequate estimations of the stellar parameters and Hα EW$_{obs}$ that could be derived from spectroscopic data. This

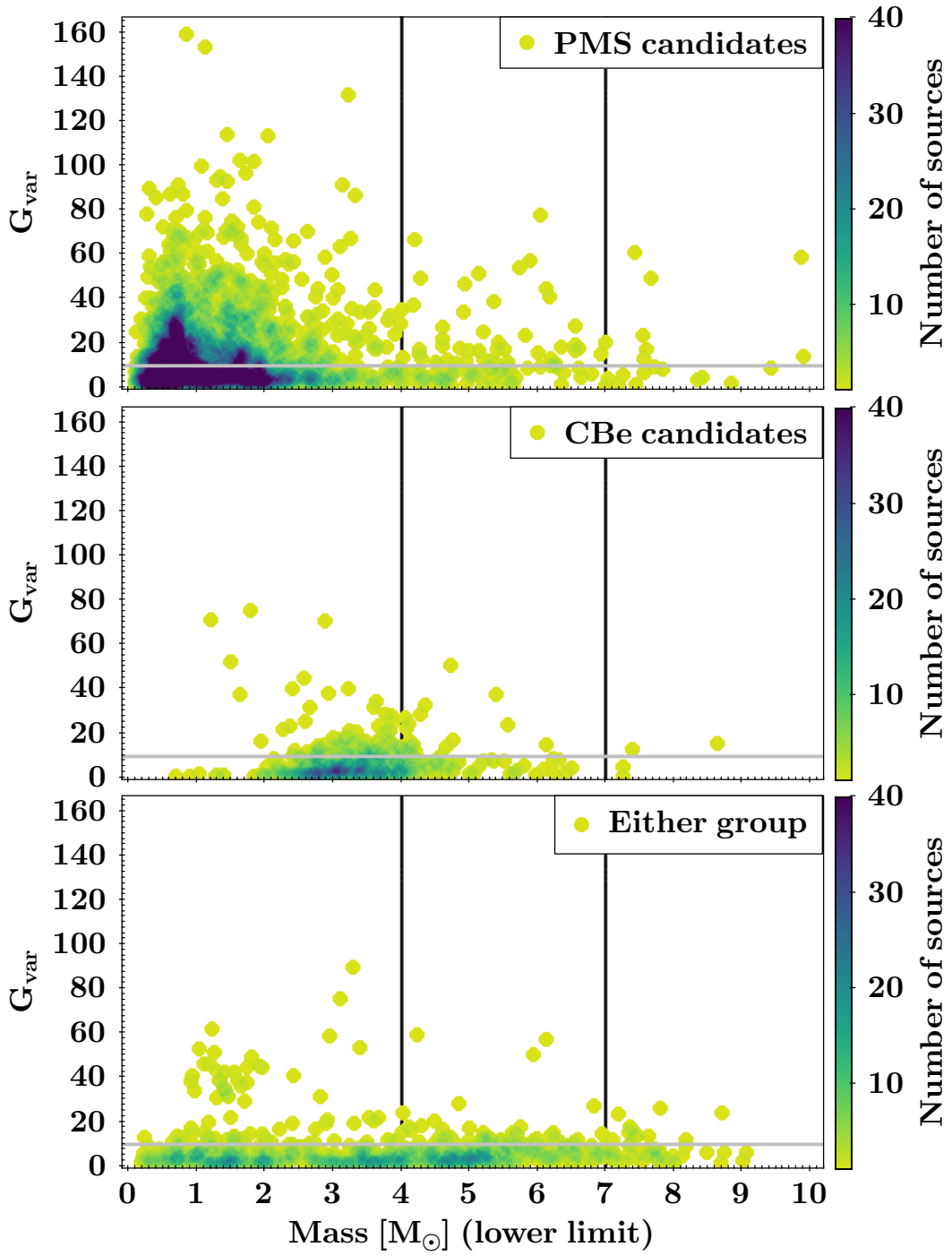


Fig. 4.16 Lower limit to the mass vs. photometric variability as traced by G_{var}. Horizontal grey lines indicate the $G_{var} = 10$ threshold from which G_{var} is sensitive to disc orientation (see discussion in Chapters 2 and 3). Vertical lines trace the two mass thresholds at 4 $M_\odot$ and 7 $M_\odot$ discussed in this chapter. Five extreme sources were left out of bounds for visualisation purposes. *Top:* PMS candidates. *Middle:* CBe candidates. *Bottom:* Either group.

was evidenced by using the observed sources of Sect. 4.1, which are a representative selection of the general catalogues (Sect. 4.7.1). In this section I study the catalogues as a function of the stellar mass lower limit. This parameter was chosen instead of the effective temperature or the luminosity because PMS objects evolve at almost constant mass values (see Chapter 1). The masses of the Either catalogue presented in this section are those obtained from PMS isochrones (see Sect. 4.3).

In Fig. 4.16 I study the irregular photometric variability as traced by G_{var} as a function of mass for the three catalogues. The first thing to note is that the sources of the PMS catalogue follow the same trend observed for the known Herbig Ae/Be stars in Chapter 2. Low-mass PMS objects show both low and high levels of variability, whereas sources with masses above 4 $M_{\odot}$, and especially 7 $M_{\odot}$, are typically weakly or not variable. This trend was also detected for the spectroscopically observed sample in Sect. 4.4.

In Chapter 2 it was concluded that G_{var} is sensitive to variability caused by disc orientation for values higher than $G_{var} \sim 10$ (see Chapter 3 for the comparison of G_{var} and V_i variability proxies). Therefore, high-mass objects only show low levels of variability because they have suffered from faster or more efficient disc dispersal mechanisms. However, in Chapter 2 it was discussed that at least a small fraction of HBe stars should still have discs large enough to be traced by this variability proxy. In that chapter I concluded that the scarcity of HBe stars at high-variability levels was caused by a bias in the set of known Herbig Ae/Be stars; most previously known HBe stars are very close to the ZAMS, and thus may have lost their discs during their evolution towards it. In Fig. 4.16 several high-mass PMS candidates appear with large G_{var} values. In Sect. 4.6.1 some of these objects could be placed accurately in the HR diagram of Fig. 4.9, and they are typically those further away from the main sequence. This is an important supporting argument to the conclusion of Chapter 2, and it also shows that I have detected new high-mass PMS stars that are less-evolved than most of the Herbig Ae/Be stars that were previously known.

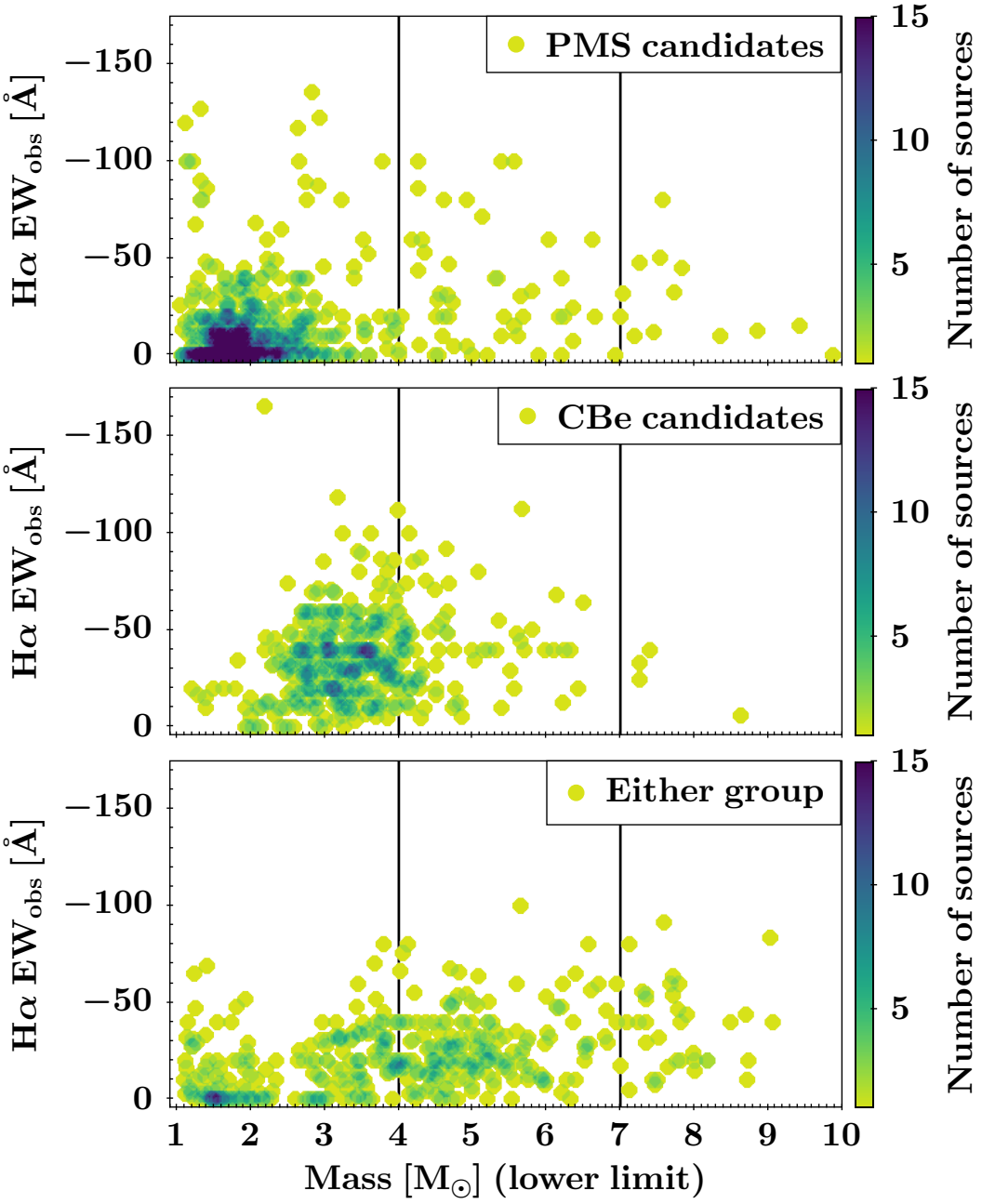


Fig. 4.17 Lower limit to the mass vs. Hα EW$_{obs}$. Vertical lines trace the two mass thresholds at 4 M$_\odot$ and 7 M$_\odot$ discussed in this chapter. Only sources with $T_{eff} > 5900$ K are shown. The sources with 'PN' warning flags (17) were removed. Three extreme sources were left out of bounds for visualisation purposes. *Top:* PMS candidates. *Middle:* CBe candidates. *Bottom:* Either group.

In Fig. 4.16 the variability of the CBe candidates seems to increase with mass. This was also found for the spectroscopically observed CBe candidates (Fig. 4.8, see discussion in Sect. 4.6.2). As discussed in Sect. 4.6.2, the variability of classical Be stars is probably not dominated by the disc orientation, and hence for the CBe candidates a different behaviour in this plot is expected with respect to PMS objects.

For the Either catalogue, Fig. 4.16 shows that sources in this catalogue mostly have low levels of variability for any mass value ($G_{var} < 10$). This is because this catalogue contains, together with some CBe stars, the PMS sources that are more similar to CBe stars (see Chapter 3). In other words, all the PMS objects with large variability were readily classified by the algorithm into the PMS group. Following the discussion of Sect. 4.7.1, Fig. 4.16 also shows that the high-mass range of the Either group contains the new high-mass PMS stars that are more similar to the previously known Herbig Be stars with dissipated discs (which only show low levels of variability) of Chapter 2.

In Fig. 4.17 I study the Hα EW$_{\rm obs}$ as a function of mass for the three catalogues. I only considered sources with $T_{eff} > 5900$ K as only for those the Hα EW$_{\rm obs}$ are accurate (see Sect. 4.7.3). In addition, I removed those candidates with 'PN' warning flags (17 PMS candidates in total).

In Fig. 4.17 it can be observed that, similar to what was reported for the known HAeBes in Chapter 2, low-mass PMS candidates show a very low median value of Hα EW$_{\rm obs}$ whereas higher mass sources have a more evenly distributed Hα EW$_{\rm obs}$. This is not exactly true beyond 7 M$_{\odot}$, probably because of the scarcity of PMS candidates in this mass range (see Sect. 4.7.2 and Fig. 4.13). This trend was also detected for the spectroscopically observed sample in Sect. 4.4. In Fig. 2.7 of Chapter 2, no low-mass Herbig Ae/Be star ($\sim 1-2$ M$_{\odot}$) shows levels of Hα emission $EW < -60$ Å. However, in the top panel of Fig. 4.17 a few PMS candidates appear at these values. This could be caused by the artificial incompleteness of the set of previously known HAeBes at low-masses (see discussion in Sect. 4.7.2), or these

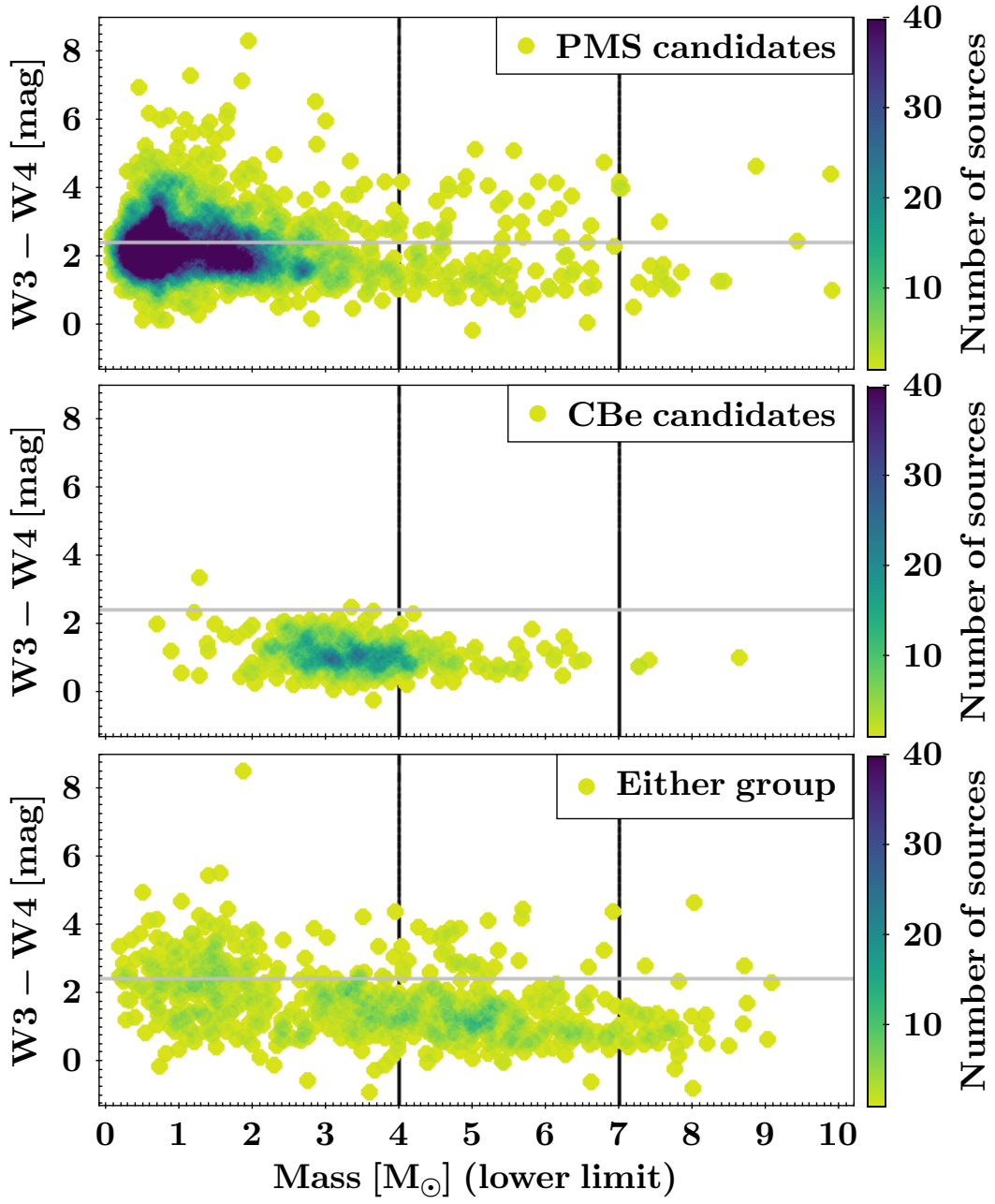


Fig. 4.18 Lower limit to the mass vs. mid-IR excess as traced by $W3-W4$ ([12]-[22] µm). Horizontal grey lines indicate the typical maximum $W3-W4$ value for CBe candidates. Vertical lines trace the two mass thresholds at 4 $M_\odot$ and 7 $M_\odot$ discussed in this chapter. Four extreme sources were left out of bounds for visualisation purposes. *Top:* PMS candidates. *Middle:* CBe candidates. *Bottom:* Either group.

sources could just be contaminants. The distribution of the Hα EW_{obs} as a function of mass for the CBe and Either catalogues shows no correlation.

To conclude the analysis of the catalogues, in Fig. 4.18 I study the mid-IR excess as traced by $W3-W4$ ([12]-[22] μm) as a function of mass for the three catalogues. In principle, the $W3-W4$ colour traces the IR excess arising from warm dust and will have little contribution by the re-radiated emission from hot gas. Therefore, the $W3-W4$ colour is expected to be typically smaller for classical Be stars, that have a dust-free gaseous disc, than for most PMS objects. Indeed, in Chapter 3 I showed that the $W3$ and $W4$ passbands were effectively used by the algorithm to separate PMS objects from classical Be stars. However, many high-mass HBe stars with dispersed discs show $W3-W4$ colours compatible with those of CBe stars. This is observed for the known objects that were considered for the training in Chapter 3. In Fig. 4.18 a grey horizontal line marks the typical maximum $W3-W4$ value for the CBe candidates. Many PMS candidates in the top panel of that figure are above this line, although many others have levels of mid-IR excess compatible with the values observed for CBe objects.

Fig. 4.18 evidences that most PMS and Either sources with $M \gtrsim 7$ $M_{\odot}$ have low $W3-W4$ values. This supports the conclusion of Chapter 2 which stated that, from ~ 7 $M_{\odot}$, Herbig Ae/Be stars suffer from faster or more efficient (dust-) disc dispersal mechanisms.

The discussion of this section about the general properties of the catalogue of new pre-main sequence candidates validates the results and conclusions of Chapter 2 for the set of previously known Herbig Ae/Be stars. In addition, the catalogue of new classical Be candidates also resembles in the studied properties the set of known classical Be stars considered in Chapter 3. This evidences the quality of the classification and reassures the analysis performed for the previously known sources.

In addition to analysing the catalogues, in this chapter I have discussed the results of independent spectroscopic observations conducted for a selection of 145 new

Herbig Ae/Be candidates and 14 new classical Be candidates. These observations further confirm the high quality and precision of the catalogues, and help to better characterise the properties of these newly discovered sources.

Chapter 5

Conclusions

Siente pequeños susurros. En seguida oye el percutir de su corazón en palpitaciones desiguales. Al través de sus párpados cerrados entrevé la llama de la luz.
Pedro Páramo, 1955 – Juan Rulfo

This chapter summarises the main conclusions and results of this doctoral thesis. First, I briefly recapitulate the context and motivation for the work presented in the previous chapters (Sect. 5.1). Secondly, I detail the main results of the thesis and explain how they fulfill the original motivation and expectations (Sect. 5.2). In addition, I introduce ideas and concepts for future work which were left unfinished in this thesis (Sect. 5.3). Finally, I conclude with some closing remarks and thoughts in Sect. 5.4.

5.1 Context and motivation

Herbig Ae/Be stars (HAeBes) are pre-main sequence (PMS) sources of intermediate-mass (historically defined as having masses 2 $M_\odot \lesssim M \lesssim 10\ M_\odot$; spectral type B, A, and F) that cover the gap between the lower-mass T Tauri stars and the deeply embedded infrared-bright Massive Young Stellar Objects (MYSOs). HAeBes are essential to understanding the properties of high-mass star formation. This is because

HAeBes present significant differences with respect to lower mass objects regarding their formation mechanisms and protoplanetary disc structures. In addition, HAeBes are optically bright so they are much more easy to observe and study than their more massive MYSO counterparts.

However, a large caveat in all of the studies dedicated to Herbig Ae/Be stars is that *only* 272 of them were known and proposed at the start of this thesis work. Only about half of those are widely accepted HAeBes, while the others have been proposed without much extra tests or verification (108/272 are in the master list of The et al., 1994, see Chapter 2). In addition, this is a very heterogeneous and biased set. In particular, few objects were known at the high-mass end (Herbig Be stars), with many of them having a doubtful nature as they are easily confused with classical Be stars (CBes, rapidly rotating main sequence B stars with Keplerian gas discs, Rivinius et al., 2013). This situation contrasts with the thousands of T Tauri stars known in the literature.

As a consequence, many open problems involving high-mass star formation suffer from these biases and the lack of completeness. For example, it is commonly accepted that T Tauri stars accrete through magnetically-driven flows arising from the protoplanetary disc, which is truncated at a distance of a few stellar-radii (see Bouvier et al., 2007 and Hartmann et al., 2016). However, higher-mass PMS objects have radiative envelopes and hence normally present negligible magnetic fields (Alecian et al., 2013; Villebrun et al., 2019). Therefore, the magnetospheric accretion paradigm may not apply to them. Several studies have indicated that the change in accretion mechanism occurs within the mass range of Herbig Ae/Be stars (see Chapter 1). The recent work of Wichittanakom et al. (2020) placed the change in accretion properties at $3.98^{+1.37}_{-0.94}$ $M_{\odot}$. Nonetheless, most Herbig Be stars considered in these studies are very close to the main sequence. Therefore, these studies barely include younger high-mass pre-main sequence sources, which are obviously of paramount importance for understanding high-mass accretion.

In addition, there is observational evidence that points towards differences between the discs of low- and high-mass PMS sources. This can be seen in the amount of infrared and millimeter excess, which are much lower for high-mass sources (Alonso-Albi et al., 2009; Ribas et al., 2015; Arun et al., 2019) or in morphology; for instance, spirals have only been found in the discs around early spectral type stars (Garufi et al., 2018). Similarly, there is a clear observational bias in these results, as so far mostly long-lived, massive discs around low-mass stars have been observed. Furthermore, more planets should have been detected around Herbig Ae/Be stars given their discs mass (albeit it is hard to determine the exact masses of their discs; Pascucci et al., 2016, Panić and Min, 2017). It has been proposed that the scarcity of detected planets is because the HAeBes observed are too old to have active planet formation (e.g. Kennedy and Kenyon, 2008 or Manara et al., 2018).

Independently, it is known that high-mass stars tend to form in clusters (Hillenbrand et al., 1995; Testi et al., 1999). Studies of massive field runaway stars have shown that at least a small fraction ($\sim 4\%$, de Wit et al., 2005) of O-type stars are formed without a cluster environment. Nonetheless, the existence of isolated high-mass star formation is still questioned (e.g. Stephens et al., 2017 or Lucas et al., 2018). Again, the scarcity of high-mass PMS sources in these studies hampers the statistics.

5.2 Main results

It is thus useful to characterise the set of known Herbig Ae/Be stars and to obtain a large, homogeneous, and unbiased catalogue of new Herbig Ae/Be stars. Gaia Data Release 2 (DR2, Gaia Collaboration et al., 2018b, 2016) provides a five dimensional astrometric solution for up to $G \lesssim 21$ mag to over 1.3 billion objects. This large dataset allows for exploitation with statistical learning techniques and it is the ideal

survey for finding new sources of the elusive Herbig Ae/Be class. Thus, Gaia DR2 is of central importance for this PhD thesis and it was used repeatedly in every chapter.

5.2.1 Analysis of previously known Herbig Ae/Be stars

Before attempting to find new Herbig Ae/Be stars, it was necessary to first characterise and homogeneously study the properties of the set of known HAeBes.

For all previously known Herbig Ae/Be stars with parallaxes in Gaia DR2, I collected their atmospheric parameters and photometric and extinction values from the literature. To these data I added catalogued near-infrared and mid-infrared photometry, and collected Hα emission line properties such as equivalent widths and line profiles, and their binarity status. In addition, I developed a photometric variability indicator from Gaia's DR2 information; a proxy that enables for an objective measurement of the optical photometric variability.

The Gaia parallaxes were used to determine distances and from these I derived luminosities, which made it possible to place 252 Herbig Ae/Be stars in the Hertzsprung-Russell (HR) diagram. From their location in the HR diagram stellar masses and ages were derived and from the spectral energy distributions I determined the infrared (IR) excess of each source. Thus, I homogeneously derived luminosities, distances, masses, ages, variabilities and IR excesses for the most complete sample of Herbig Ae/Be stars to date. This is a tenfold increase in number on earlier studies using Hipparcos parallax data. I investigated the various observed properties of the set and the relations between parameters and reached the following conclusions.

1. The Gaia photometric variability indicator shows that 48 out of 193 or $\sim 25\%$ of all Herbig Ae/Be stars are strongly variable. The presence of variability correlates very well with the Hα line profile. The variable objects display doubly peaked profiles, indicating an edge-on disc. Most sources catalogued as UX Ori type stars (UXORs) in the sample appear as strongly variable

with double-peaked profiles. In addition, the fraction of strongly variable A-type objects is close to that found for the A-type objects with the UXOR phenomenon. It had been suggested that the UXOR variability is in most cases due to asymmetric dusty disc structures seen edge-on. The observation here is the most compelling confirmation of this hypothesis. Moreover, 31 new UX Ori type sources were proposed among the previously known HAeBes.

2. High-mass Herbig Ae/Be stars do not display an IR excess and show no strong photometric variability. Several suggestions have been put forward to explain this. These include fast evolutionary timescales and fast dust dispersion timescales for high-mass objects. The break is found around 7 $M_{\odot}$, which is intriguingly similar to other statistical studies related to dusty discs around Herbig Ae/Be stars which signpost a different or more efficient disc dispersal mechanism for high-mass objects.

3. Various UV, optical, and Hα line properties including mass accretion rates, spectropolarimetric properties, and emission line variability seem to differ at a lower mass of $3-4$ $M_{\odot}$. The latter has been linked to different accretion mechanisms at work; magnetospheric accretion for the lower mass objects and another mechanism, possibly boundary layer accretion, for the higher mass objects. The differing IR and variability properties detected in this work are related to different or differently acting (dust-) disc-dispersal mechanisms, which occur at higher masses and larger size scales than the accretion traced by hydrogen recombination line emission.

This characterisation and homogeneous study of all previously known Herbig Ae/Be stars greatly assist to develop the methodology for identifying new objects of the class. The findings presented in the previous points signal the potential of Gaia for studying Herbig Ae/Be stars and high-mass star formation. Thus, Gaia is an excellent survey to search and identify new HAeBes.

5.2.2 Catalogue of new Herbig Ae/Be and classical Be stars

In this thesis I have used machine learning techniques (mainly artificial neural networks or ANNs) to produce a catalogue of new PMS candidates and a catalogue of new classical Be (CBe) stars from the Gaia survey. From the resulting catalogue of new PMS objects, new Herbig Ae/Be stars were selected by their location in the HR diagram. The pipeline, algorithms, and methodology used are described in detail in Sect. 3.3 of Chapter 3.

Herbig Ae/Be and classical Be stars have been historically a source of confusion and misclassification, and are often indistinguishable by using classical techniques. Therefore, one of the main goals of the methodology was to effectively separate these two types of objects, and as a consequence new CBe candidates were found.

The ANN was trained with the set of known Herbig Ae/Be stars previously described, and with a set of known T Tauri and CBe stars. The features that feed the ANN need to be relevant for identifying PMS sources. Hence, I wanted the features to trace the main observational characteristics of PMS sources, which are: infrared (IR) excesses, because of the radiation of the heated up protoplanetary disc, emission lines, that trace the surrounding material close to the forming star, and photometric variability. This PMS variability is caused by the presence of the disc in the line of sight (e.g. dippers, Bouvier et al., 1999, or UX Ori type sources, Grinin, 2000), because of episodic accretion events (EX Lup or FU Ori type sources, Cody et al., 2017), or pulsations due to internal instability (Zwintz et al., 2014). To feed the algorithm with these characteristics, I used observables belonging to five different surveys: Gaia DR2 for variability, 2MASS and WISE for near- and mid-IR excess respectively, and IPHAS and VPHAS+ for Hα emission.

4,150,983 sources resulted from the cross-match of Gaia DR2, 2MASS, WISE, and IPHAS or VPHAS+. To each of the 4,150,983 sources the algorithm assigned a PMS and a classical Be probability. The entire set of sources is available in electronic

form[1] (together with the catalogues described below) so the users can choose the probability thresholds (p) that fit their needs.

In Chapter 3 I presented the methodology used, evaluated the quality of the classification, and conducted an analysis of its flaws and biases. For this assessment, I made use of observables that have not been accounted for by the algorithm and hence are selection-independent, such as coordinates and parallax based distances. I have given independent evidence that the categorisation is accurate and consistent, having a high efficiency at separating PMS sources from classical Be stars. The main remarks and conclusions are:

1. At $p \geq 50\%$ the catalogue of PMS candidates is: 8470 sources, recall (completeness) of $78.8 \pm 1.4\%$ and lower limit to precision of $40.7 \pm 1.5\%$. Independent analyses indicate that the real precision is around double this value. The PMS candidates are distributed all over the Galactic plane, tend to be associated with nebulosities and appear mostly in PMS locations in the HR diagram. This catalogue (Table 3.3) is available in electronic form independently.

2. 2052 PMS candidates have a good astrometric solution in Gaia DR2 (RUWE $<$ 1.4 and $\varpi/\sigma(\varpi) \geq 10$). Of those, 1361 occupy a location in the HR diagram compatible with that of known Herbig Ae/Be stars. Many more new Herbig Ae/Be candidates can be obtained from the set of PMS candidates by relaxing the constraint on the parallax quality. This comes at a price, as the larger errors on the absolute magnitudes make it more difficult to distinguish low- and high-mass objects from each other.

3. At $p \geq 50\%$ the catalogue of classical Be candidates is: 693 sources, recall (completeness) of $85.5 \pm 1.2\%$ and lower limit to precision of $88.6 \pm 1.1\%$. In contrast to the PMS candidates, the classical Be candidates are distributed all over the Galactic plane and appear mostly in classical Be locations in

[1] https://vizier.u-strasbg.fr

the HR diagram. This catalogue (Table 3.4) is available in electronic form independently.

4. There are 1309 sources that have a combined probability of larger than 50% of belonging to either of these categories but each individual category has a probability of below 50% (the 'Either' catalogue). In general these objects have characteristics of classical Be stars or the less extreme PMS sources in the observables used. These sources are listed in a table equivalent to Tables 3.3 and 3.4 which is only available in electronic form.

5. I have made a thorough analysis of the possible biases and contaminants present in the selection. The biases can be summarised in that I have retrieved the most extreme PMS and classical Be sources in the observables used. The contaminants are mostly giants, with the special case of planetary nebulae appearing as PMS. These contaminants are sparse and easy to avoid. Instructions are given to minimise their impact (in Sects. 3.4.6, 3.4, 3.5.1, 3.5.2). The new HAeBe candidates are little affected by these contaminants, mainly because they have a good astrometric solution.

6. 3436 PMS candidates (at $p \geq 50\%$) show strong irregular photometric variabilities. For the HAeBe candidates the UXOR phenomenon is the most likely explanation. The proportion of variable HAeBe candidates is consistent with the idea that in order to see UX Ori type variability, the objects need to be oriented favourably.

7. An analysis of the relative importance of the different observables used shows that irregular photometric variability is extremely important for identifying PMS sources in general and that W3 [12 μm] and W4 [22 μm] are very powerful to separate high-mass PMS sources from classical Be stars. On the

other hand, perhaps surprisingly, $r - H\alpha$ is not very relevant for the selection of these two types of objects.

5.2.3 Independent spectroscopic observations

The catalogues of new PMS and classical Be candidates were subjected to follow up studies using independent observations. A sample of 145 PMS and 14 CBe candidates were observed with low resolution spectroscopy at the INT, CAHA2.2m, and NTT telescopes (with IDS, CAFOS, and EFOSC2 instruments respectively). The goal of the observations was to determine their spectral types and total extinction values. These make it possible to accurately place them in the HR diagram and derive their masses, luminosities, and ages. In addition, the Hα line was observed and characterised for 107 of the PMS sources and 9 of the CBe stars. Below I summarise the key conclusions of the observations:

- Evidence to support the Herbig Ae/Be nature of 120 of the 145 observed PMS sources is provided. Only 25 were detected as possible contaminants. This number agrees with the precision proposed for the PMS catalogue ($\sim 81\%$). The 120 new Herbig Ae/Be stars proposed appear in the HR diagram in the same locations as the previously known HAeBes. However, some of the newly observed objects are younger (further away from the ZAMS) than most of the previously known HAeBes of the same mass. All objects observed in Hα show the line in emission. In addition, the 120 objects present the trends with mass, photometric variability, Hα EW and Hα line profile that were found for the previously known HAeBes.

- All 14 CBe candidates observed are consistent with being a classical Be star. They are all of B-type, all of those observed in Hα have emission, and they appear in classical Be locations in the HR diagram. In addition, eight of the

nine Hα emitting objects show absorption line broadening that likely implies a fast rotational velocity, consistent with a classical Be nature.

- An additional 187 sources from the catalogues (103 PMS, 56 CBe, and 28 Either) have spectral types tabulated in the SIMBAD database at the time of writing and could be placed in the HR diagram. Although the heterogeneity and unknown biases of this set make it inferior to the spectroscopically observed sample, judging by their resulting location in the HR diagram most of the objects in this set seem to be correctly classified. These sources were essentially a random subset of the catalogues. Therefore, this is another important assessment of the quality of the catalogues as this independent set is not biased towards preferred, well-behaved, sources.

- Although the observed sources are a representative subset of the full catalogues, because of sensitivity limitations only sources at the bright end of the catalogues were observed. However, as the catalogues are distance and position independent, the results of these observations can be extrapolated to the entirety of the catalogues.

- The warning flags of the catalogues are efficient at tagging contaminants. In particular, the 'PN' flag is very efficient at identifying extreme emission line objects. In the mass range of the Herbig Ae/Be stars, these are mostly B[e] (FS CMa) stars.

- Intrinsic Gaia $G_{BP} - G_{RP}$ colours were obtained for spectral types O9/O9.5V to K4V. This is the first empirical determination of the intrinsic Gaia colours. Evidence is provided for the adequate quality of this derivation.

Therefore, these observations provide a large set of new objects of the Herbig Ae/Be and classical Be class, and further support the categorisation allowing extending the analysis to the full sample.

5.2.4 Analysis of the catalogues

Mass, temperature and luminosity lower limits are derived for a large fraction of the PMS, CBe, and Either catalogues by means of the HR diagram. In addition, Hα equivalent widths (EW_{obs}) were derived from the $r-i$ and $r-H\alpha$ colours. In the case of the stellar parameters, $\sim 49\%$ of the lower limit determinations are within 20% deviations of the value obtained with data from spectroscopic observations. This correspondence is higher for low mass objects ($M \lesssim 4$ $M_{\odot}$).

For the catalogue of new PMS sources, the break in variability properties is found to occur at the same mass (7 $M_{\odot}$) as for the previously known HAeBes. However, a larger number of highly variable high-mass objects is present. It is proposed that these are new massive objects that are younger than most previously known Herbig Be stars. Their variability can be explained by the fact that they still have not dispersed their discs. In addition, the same trend in the median amount of Hα EW per mass bin detected for the previously known HAeBes is observed for the new PMS candidates.

The mass distribution of the catalogue of PMS objects is consistent with the IMF down to $0.5-1$ $M_{\odot}$. Therefore, the catalogue of new PMS candidates has no artificial gap between Herbig Ae/Be stars and T Tauri stars, contrary to the set of previously known HAeBes.

Regarding the catalogue of classical Be candidates, most sources have masses that correspond to main-sequence B stars. As expected for the CBe group, the mid-IR excess as traced by $W3-W4$ is not large for any CBe candidate.

Regarding the Either group, I demonstrated that it contains the high-mass PMS stars that are more similar to CBe stars. This means that the high-mass sources in the PMS catalogue have larger IR excess, Hα emission and photometric variability than is typical for the previously known high-mass PMS sources. It is proposed that this is because they are less evolved or have suffered from a different disc evolution.

The fact that most of the sources beyond 7 $M_{\odot}$ are in this Either group is a strong support for the fast disc dispersal mechanisms proposed for the high-mass Herbig Ae/Be stars.

Summarising, the study of a well-characterised sample of 8470 new pre-main sequence objects that were homogeneously selected (of which, at least, 1361 are strong Herbig Ae/Be candidates) validate the results that were obtained for the previously known set of HAeBes. The current sample corrects the heterogeneous nature and the biases of the set of previously known HAeBes, and allows for a much better defined, constrained, robust, and rigorous studies.

5.3 Future work

This section of future work is divided into two different subsections. First, in Sect. 5.3.1 I discuss some aspects of the work presented in this thesis that could have been done differently or from another perspective. Secondly, in Sect. 5.3.2 I present ideas and concepts for future work from the results presented in the previous sections of this chapter.

5.3.1 Different approach to some results of this thesis

In Chapter 3 a set of observables was chosen to feed the machine learning algorithm and classify sources into the PMS and CBe categories. By demanding to have detections in all these observables, the set of sources available for the training reduced significantly. In addition, the objects that could be classified with the trained algorithm also reduced considerably. In Table 5.1 I present the known sources available for training for different combinations of the observables used. The total number of sources with those observables is also included (i.e. the corresponding 'Sample of Study').

Table 5.1 Size of the sets of known PMS stars, known classical Be sources, and Sample of Study for different chosen observables.

	PMS			
	Herbig Ae/Be	T Tauri	CBe	Sample of Study
All considered sources	255	3171	1992	51,548,230
Using all observables	**163**	**685**	**775**	**4,151,538**
All but r and Hα	206	1622	1450	4,707,553
All but B_p, G, and R_p	163	685	775	4,151,538
All but J, H, and K_s	166	695	779	4,711,861
All but W1, W2, W3, and W4	197	1130	976	35,463,237
All but W3 and W4	168	1124	969	35,236,079
All but G_{var} and V_{htg}	163	685	775	4,151,538

Notes: In bold font the selection of observables used in Chapter 3 to train the ANN. Note that the Sample of Study column includes the sources of the other columns.

The goal of Chapter 3 was to find new high-mass PMS sources and to differentiate them from CBe stars. Therefore, these observables were selected *a priori* because of the known properties of Herbig Ae/Be and CBe sources. However, in Sect. 3.5.4 of Chapter 3 it was discussed what was their effective relevance for separating Herbig Ae/Be from classical Be stars and in general for identifying sources of these two categories among other types of sources. Future studies may want to use different observables based on what was learnt here.

In addition, among the sources that are not in the Either or PMS catalogue but have a significant probability of being PMS (e.g. $p(PMS) > 0.40$ but $p(CBe) < 0.10$) there are also interesting PMS objects. However, this set likely does not contain many massive objects and, as discussed in Chapter 3, it brings complications to use sources with low probabilities. That is the reason why these sources were not considered in this dissertation. However, future studies might be aimed at using some of those sources with the intention to search for the pre-main sequence objects that were not the main target of this thesis, like debris discs around low-mass T Tauris or PMS sources experiencing outbursts.

5.3.2 Research from the results of this thesis

Most of the uncertainties of this thesis come from the parallax error. Therefore, future Gaia data releases will expand the conclusions and improve the derivations of this work, and will allow to better study the properties of Herbig Ae/Be stars. In particular, smaller parallax uncertainties will allow to study SED evolution along PMS tracks, an analysis that with current Gaia DR2 parallaxes proved to be impossible.

The most immediate future work is the derivation of the accretion rates from the Hα EWs of the new Herbig Ae/Be stars spectroscopically observed. Independently, accretion rates can be directly measured from the UV excess emission over the Balmer jump. This is a work in preparation led by Mr. C. Wichittanakom in collaboration with myself. Similarly, an estimation of the accretion rate can be obtained for the whole catalogue of PMS candidates, given that it was demonstrated in this work that the Hα EW derived from $r-i$ and $r-H\alpha$ colours is adequate. In addition, accretion rates can be independently estimated for the full set of PMS candidates by using catalogues with U band information to trace the UV excess, like IGAPS (INT Galactic Plane Survey, Monguió et al., 2020).

Moreover, there are 22 sources that were removed from the SIMBAD set of Chapter 4 (14 CBe, 6 PMS, and 2 Either). They were excluded because of their spectral type 'O' or 'OB' which I could not characterise because of the limitations of the Gaia intrinsic colour derivations (see Chapter 4). However, these very massive sources deserve a study in their own right.

In order to probe the new PMS sources from a different perspective, it would be of interest to study in detail the sky associations found in Chapter 3. In particular, to study how many new Herbig Ae/Be stars appear related to known open clusters and whether there are new high-mass PMS candidates in isolation. As the selection was independent from the objects' coordinates and distances, a study of the clustering

properties of the new PMS stars would not be biased towards any preferred direction or star forming region.

Separately, a detailed study of the newly spectroscopically observed CBe sources and of the general catalogue of classical Be candidates is also missing in this doctoral thesis and it is planned for the future. In addition, the elusive B[e] stars, that were treated here as contaminants, deserve more attention. Very few of these objects are known at the moment and the 'PN' warning flag of the catalogues has proved to be rather efficient at identifying sources of this class. In this work eight B[e] stars were identified.

Finally, it would be interesting to measure the disc mass of some of the near-by new Herbig Ae/Be stars with ALMA. This, in addition to confirm the existence of discs around these objects, will allow for a detailed study of protoplanetary disc evolution in high-mass objects, which in turn will enable a study of the possible planet formation mechanisms of these massive stars as a function of their age and mass.

I conclude this section of future work by stating that the catalogue of new PMS candidates was accepted as target list for the WEAVE survey, a wide-field spectroscopic survey which will be carried out at the William Herschel Telescope in the forthcoming years (Dalton et al., 2018).

5.4 Final remarks

In this dissertation, a study of the general properties of the previously known and proposed Herbig Ae/Be stars is presented. This study evidences that high-mass stars ($M \gtrsim 7$ M$_\odot$) suffer from faster or more efficient disc dispersal mechanisms, and provides the necessary information to find new objects of the class.

By using machine learning techniques a catalogue of 8470 new pre-main sequence candidates was obtained. Evidence has been provided for the accuracy

and high-quality of the catalogue. 1361 of those pre-main sequence sources have been proposed as strong new Herbig Ae/Be candidates from their position in the Hertzsprung-Russell diagram. However, many more Herbig Ae/Be candidates can be obtained by relaxing the constraints on the parallax quality.

In addition, a different catalogue of 693 new classical Be candidates was obtained. The algorithm used to construct the new catalogues has efficiently separated the high-mass PMS sources from the classical Be stars. This could not be achieved in previous studies that used classical techniques and it is the first time these two types of objects are effectively separated in a large scale categorisation.

The analysis of the general properties of the new catalogue of PMS candidates confirms the results obtained for the set of previously known Herbig Ae/Be stars. However, the newly discovered sources make it possible to correct for the biases and heterogeneity of the set of previously known HAeBes. For example, most previously known high-mass PMS stars are very close to the main sequence, whereas among the proposed new Herbig Ae/Be stars there are many high-mass objects that are younger. In addition, the previous set of HAeBes has an artificial gap between Herbig Ae and T Tauri stars, where it is incomplete. This gap does not exist in the new catalogue of PMS sources.

The catalogue of new Herbig Ae/Be stars that is presented in this thesis increases the number of known objects of the class by an order of magnitude, and it is of paramount importance for providing new insights to the nature of these intermediate-mass pre-main sequence objects. As an example of the studies that this catalogue enables, in this dissertation 120 new Herbig Ae/Be stars are proposed and studied from independent spectroscopic observations.

I conclude the final remarks of this dissertation acknowledging the STARRY project, of which this PhD thesis is part, which has received funding from the European Union's Horizon 2020 research and innovation programme under MSCA ITN-EID grant agreement No 676036.

References

Aarnio, A. N., Monnier, J. D., Harries, T. J., Kraus, S., Calvet, N., Acreman, D., and Che, X. (2017). High-cadence, High-resolution Spectroscopic Observations of Herbig Stars HD 98922 and V1295 Aquila. *ApJ*, 848:18.

Ababakr, K. M., Fairlamb, J. R., Oudmaijer, R. D., and van den Ancker, M. E. (2015). Spectroscopy and linear spectropolarimetry of the early Herbig Be stars PDS 27 and PDS 37. *MNRAS*, 452:2566–2578.

Ababakr, K. M., Oudmaijer, R. D., and Vink, J. S. (2016). Linear spectropolarimetry across the optical spectrum of Herbig Ae/Be stars. *MNRAS*, 461:3089–3110.

Ababakr, K. M., Oudmaijer, R. D., and Vink, J. S. (2017). A statistical spectropolarimetric study of Herbig Ae/Be stars. *MNRAS*, 472:854–868.

Acke, B. and van den Ancker, M. E. (2006). A survey for nanodiamond features in the 3 micron spectra of Herbig Ae/Be stars. *A&A*, 457:171–181.

Acke, B., van den Ancker, M. E., and Dullemond, C. P. (2005). [O I] 6300 Å emission in Herbig Ae/Be systems: Signature of Keplerian rotation. *A&A*, 436:209–230.

Akras, S., Guzman-Ramirez, L., and Gonçalves, D. R. (2019a). Compact planetary nebulae: improved IR diagnostic criteria based on classification tree modelling. *MNRAS*, 488(3):3238–3250.

Akras, S., Leal-Ferreira, M. L., Guzman-Ramirez, L., and Ramos-Larios, G. (2019b). A machine learning approach for identification and classification of symbiotic stars using 2MASS and WISE. *MNRAS*, 483(4):5077–5104.

Alecian, E., Wade, G. A., Catala, C., Grunhut, J. H., Landstreet, J. D., Bagnulo, S., Böhm, T., Folsom, C. P., Marsden, S., and Waite, I. (2013). A high-resolution spectropolarimetric survey of Herbig Ae/Be stars - I. Observations and measurements. *MNRAS*, 429:1001–1026.

Alencar, S. H. P., Bouvier, J., Walter, F. M., Dougados, C., Donati, J. F., Kurosawa, R., Romanova, M., Bonfils, X., Lima, G. H. R. A., Massaro, S., Ibrahimov, M., and Poretti, E. (2012). Accretion dynamics in the classical T Tauri star V2129 Ophiuchi. *A&A*, 541:A116.

Alonso-Albi, T., Fuente, A., Bachiller, R., Neri, R., Planesas, P., Testi, L., Berné, O., and Joblin, C. (2009). Circumstellar disks around Herbig Be stars. *A&A*, 497:117–136.

References

Ambartsumian, V. A. (1947). The Evolution of Stars and Astrophysics (Rus.). *Yerevan: Armenian SSR Academy of Sciences Press*.

Anders, F., Khalatyan, A., Chiappini, C., Queiroz, A. B., Santiago, B. X., Jordi, C., Girardi, L., Brown, A. G. A., Matijevič, G., Monari, G., Cantat-Gaudin, T., Weiler, M., Khan, S., Miglio, A., Carrillo, I., Romero-Gómez, M., Minchev, I., de Jong, R. S., Antoja, T., Ramos, P., Steinmetz, M., and Enke, H. (2019). Photo-astrometric distances, extinctions, and astrophysical parameters for Gaia DR2 stars brighter than G = 18. *A&A*, 628:A94.

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., and Bailer-Jones, C. A. L. (2018). Gaia Data Release 2. First stellar parameters from Apsis. *A&A*, 616:A8.

André, P. (2002). The Initial Conditions for Protostellar Collapse: Observational Constraints. In Bouvier, J. and Zahn, J.-P., editors, *EAS Publications Series*, volume 3 of *EAS Publications Series*, pages 1–38.

André, P., Di Francesco, J., Ward-Thompson, D., Inutsuka, S. I., Pudritz, R. E., and Pineda, J. E. (2014). From Filamentary Networks to Dense Cores in Molecular Clouds: Toward a New Paradigm for Star Formation. In Beuther, H., Klessen, R. S., Dullemond, C. P., and Henning, T., editors, *Protostars and Planets VI*, page 27.

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., et al. (2010). From filamentary clouds to prestellar cores to the stellar IMF: Initial highlights from the Herschel Gould Belt Survey. *A&A*, 518:L102.

Andrews, S. M. (2020). Observations of Protoplanetary Disk Structures. *arXiv e-prints*, page arXiv:2001.05007.

Andrews, S. M., Huang, J., Pérez, L. M., Isella, A., Dullemond, C. P., Kurtovic, N. T., Guzmán, V. V., Carpenter, J. M., Wilner, D. J., Zhang, S., Zhu, Z., Birnstiel, T., Bai, X.-N., Benisty, M., Hughes, A. M., Öberg, K. I., and Ricci, L. (2018). The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, Sample, Calibration, and Overview. *ApJ*, 869(2):L41.

Arellano Ferro, A. and Giridhar, S. (2003). HD 34700: A new T Tauri double-lined spectroscopic binary. *A&A*, 408:L29–L32.

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., et al. (2018). Gaia Data Release 2. Catalogue validation. *A&A*, 616:A17.

Argiroffi, C., Drake, J. J., Bonito, R., Orlando, S., Peres, G., and Miceli, M. (2017). Redshifted X-rays from the material accreting onto TW Hydrae: Evidence of a low-latitude accretion spot. *A&A*, 607:A14.

Arias, M. L., Cidale, L. S., Kraus, M., Torres, A. F., Aidelman, Y., Zorec, J., and Granada, A. (2018). Near-infrared Spectra of a Sample of Galactic Unclassified B[e] Stars. *PASP*, 130(993):114201.

Arun, R., Mathew, B., Manoj, P., Ujjwal, K., Kartha, S. S., Viswanath, G., Narang, M., and Paul, K. T. (2019). On the Mass Accretion Rate and Infrared Excess in Herbig Ae/Be Stars. *AJ*, 157(4):159.

Aspin, C. (1998). Discovery of multiple non-axially symmetric near-IR bow shocks around the pre-main sequence binary {AFGL 961}. *A&A*, 335:1040–1048.

Avenhaus, H., Quanz, S. P., Garufi, A., Perez, S., Casassus, S., Pinte, C., Bertrang, G. H. M., Caceres, C., Benisty, M., and Dominik, C. (2018). Disks around T Tauri Stars with SPHERE (DARTTS-S). I. SPHERE/IRDIS Polarimetric Imaging of Eight Prominent T Tauri Disks. *ApJ*, 863(1):44.

Baade, D., Pigulski, A., Rivinius, T., Carciofi, A. C., Panoglou, D., Ghoreyshi, M. R., Hand ler, G., Kuschnig, R., Moffat, A. F. J., Pablo, H., Popowicz, A., Wade, G. A., Weiss, W. W., and Zwintz, K. (2018). Short-term variability and mass loss in Be stars. III. BRITE and SMEI satellite photometry of 28 Cygni. *A&A*, 610:A70.

Bailer-Jones, C. A. L. (2015). Estimating Distances from Parallaxes. *Publications of the Astronomical Society of the Pacific*, 127:994.

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., and Andrae, R. (2018). Estimating Distance from Parallaxes. IV. Distances to 1.33 Billion Stars in Gaia Data Release 2. *AJ*, 156:58.

Baines, D., Oudmaijer, R. D., Porter, J. M., and Pozzo, M. (2006). On the binarity of Herbig Ae/Be stars. *MNRAS*, 367:737–753.

Banda-Barragán, W. E., Parkin, E. R., Federrath, C., Crocker, R. M., and Bicknell, G. V. (2016). Filament formation in wind-cloud interactions - I. Spherical clouds in uniform magnetic fields. *MNRAS*, 455(2):1309–1333.

Banko, M. and Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 26–33, Toulouse, France. Association for Computational Linguistics.

Banzatti, A., Garufi, A., Kama, M., Benisty, M., Brittain, S., Pontoppidan, K. M., and Rayner, J. (2018). Observing the linked depletion of dust and CO gas at 0.1-10 au in disks of intermediate-mass stars. *A&A*, 609:L2.

Baraffe, I., Chabrier, G., and Gallardo, J. (2009). Episodic Accretion at Early Stages of Evolution of Low-Mass Stars and Brown Dwarfs: A Solution for the Observed Luminosity Spread in H-R Diagrams? *ApJ*, 702(1):L27–L31.

Barentsen, G., Farnhill, H. J., Drew, J. E., González-Solares, E. A., Greimel, R., Irwin, M. J., Miszalski, B., Ruhland, C., Groot, P., Mampaso, A., Sale, S. E., Henden, A. A., Aungwerojwit, A., Barlow, M. J., Carter, P. J., Corradi, R. L. M., Drake, J. J., Eislöffel, J., Fabregat, J., Gänsicke, B. T., Gentile Fusillo, N. P., Greiss, S., Hales, A. S., Hodgkin, S., Huckvale, L., Irwin, J., King, R., Knigge, C., Kupfer, T., Lagadec, E., et al. (2014). The second data release of the INT Photometric Hα Survey of the Northern Galactic Plane (IPHAS DR2). *MNRAS*, 444:3230–3257.

Baron, D. (2019). Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, page arXiv:1904.07248.

Baug, T., Dewangan, L. K., Ojha, D. K., Tachihara, K., Pandey, A. K., Sharma, S., Tamura, M., Ninan, J. P., and Ghosh, S. K. (2018). Star Formation in the Sh 2-53 Region Influenced by Accreting Molecular Filaments. *ApJ*, 852(2):119.

Bertout, C., Basri, G., and Bouvier, J. (1988). Accretion Disks around T Tauri Stars. *ApJ*, 330:350.

Bessel, F. W. (1838). Bestimmung der Entfernung des 61sten Sterns des Schwans. *Astronomische Nachrichten*, 16(5):65.

Beuther, H., Ragan, S. E., Johnston, K., Henning, T., Hacar, A., and Kainulainen, J. T. (2015). Filament fragmentation in high-mass star formation. *A&A*, 584:A67.

Bhandare, A., Kuiper, R., Henning, T., Fendt, C., Marleau, G.-D., and Kölligan, A. (2018). First core properties: from low- to high-mass star formation. *A&A*, 618:A95.

Biller, B., Lacour, S., Juhász, A., Benisty, M., Chauvin, G., Olofsson, J., Pott, J.-U., Müller, A., Sicilia- Aguilar, A., Bonnefoy, M., Tuthill, P., Thebault, P., Henning, T., and Crida, A. (2012). A Likely Close-in Low-mass Stellar Companion to the Transitional Disk Star HD 142527. *ApJ*, 753:L38.

Blondel, P. F. C. and Djie, H. R. E. T. A. (2006). Modeling of PMS Ae/Fe stars using UV spectra. *A&A*, 456(3):1045–1068.

Bodensteiner, J., Shenar, T., Mahy, L., Fabry, M., Marchant, P., Abdul-Masih, M., Banyard, G., Bowman, D. M., Dsilva, K., Frost, A. J., Hawcroft, C., Reggiani, M., and Sana, H. (2020a). Is HR 6819 a triple system containing a black hole?. An alternative explanation. *A&A*, 641:A43.

Bodensteiner, J., Shenar, T., and Sana, H. (2020b). Investigating the lack of main-sequence companions to massive Be stars. *A&A*, 641:A42.

Bodman, E. H. L., Quillen, A. C., Ansdell, M., Hippke, M., Boyajian, T. S., Mamajek, E. E., Blackman, E. G., Rizzuto, A., and Kastner, J. H. (2017). Dippers and dusty disc edges: new diagnostics and comparison to model predictions. *MNRAS*, 470(1):202–223.

Boehm, T. and Catala, C. (1995). Rotation, winds and active phenomena in Herbig Ae/Be stars. *A&A*, 301:155.

Boersma, C., Peeters, E., Martín-Hernández, N. L., van der Wolk, G., Verhoeff, A. P., Tielens, A. G. G. M., Waters, L. B. F. M., and Pel, J. W. (2009). A spatial study of the mid-IR emission features in four Herbig Ae/Be stars. *A&A*, 502:175–187.

Boissay, R., Parker, Q. A., Frew, D. J., and Bojicic, I. (2012). Planetary nebulae and their mimics: The MASH-MEN Project. In *IAU Symposium*, volume 283 of *IAU Symposium*, pages 316–317.

Bolatto, A. D., Wong, T., Utomo, D., Blitz, L., Vogel, S. N., Sánchez, S. F., Barrera-Ballesteros, J., Cao, Y., Colombo, D., Dannerbauer, H., García-Benito, R., Herrera-Camus, R., Husemann, B., Kalinova, V., Leroy, A. K., Leung, G., Levy, R. C., Mast, D., Ostriker, E., Rosolowsky, E., Sandstrom, K. M., Teuben, P., van de Ven, G., and Walter, F. (2017). The EDGE-CALIFA Survey: Interferometric Observations of 126 Galaxies with CARMA. *ApJ*, 846(2):159.

Booth, A. S. and Ilee, J. D. (2020). $^{13}C^{17}O$ suggests gravitational instability in the HL Tau disc. *MNRAS*, 493(1):L108–L113.

Borges Fernandes, M., Kraus, M., Lorenz Martins, S., and de Araújo, F. X. (2007). On the evolutionary stage of the unclassified B[e] star CD-42°11721†. *MNRAS*, 377:1343–1362.

Borre, C. C., Baade, D., Pigulski, A., Panoglou, D., Weiss, A., Rivinius, T., Handler, G., Moffat, A. F. J., Popowicz, A., Wade, G. A., Weiss, W. W., and Zwintz, K. (2020). Short-term variability and mass loss in Be stars. V. Space photometry and ground-based spectroscopy of γ Cas. *A&A*, 635:A140.

Bouvier, J., Alencar, S. H. P., Harries, T. J., Johns- Krull, C. M., and Romanova, M. M. (2007). Magnetospheric Accretion in Classical T Tauri Stars. In Reipurth, B., Jewitt, D., and Keil, K., editors, *Protostars and Planets V*, page 479.

Bouvier, J., Chelli, A., Allain, S., Carrasco, L., Costero, R., Cruz-Gonzalez, I., Dougados, C., Fernández, M., Martín, E. L., Ménard, F., Mennessier, C., Mujica, R., Recillas, E., Salas, L., Schmidt, G., and Wichmann, R. (1999). Magnetospheric accretion onto the T Tauri star AA Tauri. I. Constraints from multisite spectrophotometric monitoring. *A&A*, 349:619–635.

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

Brickhouse, N. S., Cranmer, S. R., Dupree, A. K., Luna, G. J. M., and Wolk, S. (2010). A Deep Chandra X-Ray Spectrum of the Accreting Young Star TW Hydrae. *ApJ*, 710(2):1835–1847.

Buder, S., Asplund, M., Duong, L., Kos, J., Lind, K., Ness, M. K., Sharma, S., Bland -Hawthorn, J., Casey, A. R., de Silva, G. M., D'Orazi, V., Freeman, K. C., Lewis, G. F., Lin, J., Martell, S. L., Schlesinger, K. J., Simpson, J. D., Zucker, D. B., Zwitter, T., Amarsi, A. M., Anguiano, B., Carollo, D., Casagrande, L., Čotar, K., Cottrell, P. L., da Costa, G., Gao, X. D., Hayden, M. R., Horner, J., Ireland, M. J., et al. (2018). The GALAH Survey: second data release. *MNRAS*, 478(4):4513–4552.

Cannon, A. J. and Pickering, E. C. (1901). Spectra of bright southern stars photographed with the 13-inch Boyden telescope as part of the Henry Draper Memorial. *Annals of Harvard College Observatory*, 28:129–P.6.

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

Canovas, H., Hardy, A., Zurlo, A., Wahhaj, Z., Schreiber, M. R., Vigan, A., Villaver, E., Olofsson, J., Meeus, G., Ménard, F., Caceres, C., Cieza, L. A., and Garufi, A. (2017). Constraining the mass of the planet(s) sculpting a disk cavity. The intriguing case of 2MASS J16042165-2130284. *A&A*, 598:A43.

Cardelli, J. A., Clayton, G. C., and Mathis, J. S. (1989). The Relationship between Infrared, Optical, and Ultraviolet Extinction. *ApJ*, 345:245.

Carmona, A., van den Ancker, M. E., Audard, M., Henning, T., Setiawan, J., and Rodmann, J. (2010). New Herbig Ae/Be stars confirmed via high-resolution optical spectroscopy. *A&A*, 517:A67.

Carrasco-González, C., Torrelles, J. M., Cantó, J., Curiel, S., Surcis, G., Vlemmings, W. H. T., van Langevelde, H. J., Goddi, C., Anglada, G., Kim, S. W., Kim, J. S., and Gómez, J. F. (2015). Observing the onset of outflow collimation in a massive protostar. *Science*, 348(6230):114–117.

Casagrande, L. and VandenBerg, D. A. (2018). On the use of Gaia magnitudes and new tables of bolometric corrections. *MNRAS*, 479(1):L102–L107.

Castelli, F. and Kurucz, R. L. (2003). New Grids of ATLAS9 Model Atmospheres. In Piskunov, N., Weiss, W. W., and Gray, D. F., editors, *Modelling of Stellar Atmospheres*, volume 210 of *IAU Symposium*, page A20.

Castro-Ginard, A., Jordi, C., Luri, X., Cantat-Gaudin, T., and Balaguer-Núñez, L. (2019). Hunting for open clusters in Gaia DR2: the Galactic anticentre. *A&A*, 627:A35.

Castro-Ginard, A., Jordi, C., Luri, X., Julbe, F., Morvan, M., Balaguer-Núñez, L., and Cantat-Gaudin, T. (2018). A new method for unveiling open clusters in Gaia. New nearby open clusters confirmed by DR2. *A&A*, 618:A59.

Cauley, P. W. and Johns-Krull, C. M. (2014). Diagnosing Mass Flows around Herbig Ae/Be Stars Using the He I λ10830 Line. *ApJ*, 797:112.

Cauley, P. W. and Johns-Krull, C. M. (2015). Optical Mass Flow Diagnostics in Herbig Ae/Be Stars. *ApJ*, 810:5.

Chandrasekhar, S. (1939). *An introduction to the study of stellar structure*. The University of Chicago press.

Chelli, A., Cruz-Gonzalez, I., and Reipurth, B. (1995). Near infrared observations of southern young multiple systems. *Astronomy and Astrophysics Supplement Series*, 114:135.

Chen, P. S., Shan, H. G., and Zhang, P. (2016). A new photometric study of Herbig Ae/Be stars in the infrared. *New A*, 44:1–11.

Choi, J., Dotter, A., Conroy, C., Cantiello, M., Paxton, B., and Johnson, B. D. (2016). Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled Models. *ApJ*, 823(2):102.

Chollet, F. et al. (2015). Keras. Self-publishing: https://keras.io.

Clarke, S. D., Whitworth, A. P., Duarte-Cabral, A., and Hubber, D. A. (2017). Filamentary fragmentation in a turbulent medium. *MNRAS*, 468(2):2489–2505.

Clem, J. L. and Landolt, A. U. (2016). Faint UBVRI Standard Star Fields at +50° Declination. *AJ*, 152:91.

Cochetti, Y. R., Arcos, C., Kanaan, S., Meilland, A., Cidale, L. S., and Curé, M. (2019). Spectro-interferometric observations of a sample of Be stars. Setting limits to the geometry and kinematics of stable Be disks. *A&A*, 621:A123.

Cody, A. M., Hillenbrand, L. A., David, T. J., Carpenter, J. M., Everett, M. E., and Howell, S. B. (2017). A Continuum of Accretion Burst Behavior in Young Stars Observed by K2. *ApJ*, 836(1):41.

Connelley, M. S., Reipurth, B., and Tokunaga, A. T. (2008). The Evolution of the Multiplicity of Embedded Protostars. I. Sample Properties and Binary Detections. *AJ*, 135:2496–2525.

Contreras, Y., Garay, G., Rathborne, J. M., and Sanhueza, P. (2016). Fragmentation in filamentary molecular clouds. *MNRAS*, 456(2):2041–2051.

Contreras Peña, C., Naylor, T., and Morrell, S. (2019). Determining the recurrence time-scale of long-lasting YSO outbursts. *MNRAS*, 486(4):4590–4611.

Corporon, P. and Lagrange, A. M. (1999). A search for spectroscopic binaries among Herbig Ae/Be stars. *Astronomy and Astrophysics Supplement Series*, 136:429–444.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. In *Machine Learning*, pages 273–297.

Costigan, G., Vink, J. S., Scholz, A., Ray, T., and Testi, L. (2014). Temperaments of young stars: rapid mass accretion rate changes in T Tauri and Herbig Ae stars. *MNRAS*, 440:3444–3461.

Cote, J. and Waters, L. B. F. M. (1987). IRAS observations of Be stars. I. Statistical study of the IR excess of 101 Be stars. *A&A*, 176:93–106.

Coulson, I. M. and Walther, D. M. (1995). SAO 206462 - a solar-type star with a dusty, organically rich environment. *MNRAS*, 274:977–986.

Crowther, P. A., Caballero-Nieves, S. M., Bostroem, K. A., Maíz Apellániz, J., Schneider, F. R. N., Walborn, N. R., Angus, C. R., Brott, I., Bonanos, A., de Koter, A., de Mink, S. E., Evans, C. J., Gräfener, G., Herrero, A., Howarth, I. D., Langer, N., Lennon, D. J., Puls, J., Sana, H., and Vink, J. S. (2016). The R136 star cluster dissected with Hubble Space Telescope/STIS. I. Far-ultraviolet spectroscopic census and the origin of He II λ1640 in young star clusters. *MNRAS*, 458(1):624–659.

Curran, R. L., Argiroffi, C., Sacco, G. G., Orland o, S., Peres, G., Reale, F., and Maggio, A. (2011). Multiwavelength diagnostics of accretion in an X-ray selected sample of CTTSs. *A&A*, 526:A104.

Cutri, R. M., Wright, E. L., Conrow, T., Fowler, J. W., Eisenhardt, P. R. M., Grillmair, C., Kirkpatrick, J. D., Masci, F., McCallon, H. L., Wheelock, S. L., Fajardo-Acosta, S., Yan, L., Benford, D., Harbut, M., Jarrett, T., Lake, S., Leisawitz, D., Ressler, M. E., Stanford, S. A., Tsai, C. W., Liu, F., Helou, G., Mainzer, A., Gettings, D., Gonzalez, A., Hoffman, D., Marsh, K. A., Padgett, D., Skrutskie, M. F., Beck, R. P., et al. (2013). Explanatory Supplement to the AllWISE Data Release Products. Technical report.

Dalton, G., Trager, S., Abrams, D. C., Bonifacio, P., Aguerri, J. A. L., Vallenari, A., Middleton, K., Benn, C., Dee, K., Sayède, F., Lewis, I., Pragt, J., Picó, S., Walton, N., Rey, J., Allende, C., Lhomé, É., Terrett, D., Brock, M., Gilbert, J., Ridings, A., Verheijen, M., Tosh, I., Steele, I., Stuik, R., Kroes, G., Tromp, N., Kragt, J., Lesman, D., Mottram, C., et al. (2018). Construction progress of WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope. In *Proc. SPIE*, volume 10702 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 107021B.

Davies, B., Hoare, M. G., Lumsden, S. L., Hosokawa, T., Oudmaijer, R. D., Urquhart, J. S., Mottram, J. C., and Stead, J. (2011). The Red MSX Source survey: critical tests of accretion models for the formation of massive stars. *MNRAS*, 416:972–990.

Davies, B., Oudmaijer, R. D., and Sahu, K. C. (2007). Integral-Field Spectroscopy of the Post-Red Supergiant IRC +10420: Evidence for an Axisymmetric Wind. *ApJ*, 671:2059–2067.

de Wit, W. J., Testi, L., Palla, F., and Zinnecker, H. (2005). The origin of massive O-type field stars: II. Field O stars as runaways. *A&A*, 437(1):247–255.

Deason, A. J., Belokurov, V., Erkal, D., Koposov, S. E., and Mackey, D. (2017). The Clouds are breaking: tracing the Magellanic system with Gaia DR1 Mira variables. *MNRAS*, 467:2636–2647.

Delgado, A., Hodgkin, S., Evans, D. W., Harrison, D. L., Rixon, G., van Leeuwen, F., van Leeuwen, M., and Yoldas, A. (2019). *Gaia Photometric Science Alerts Data Flow*, volume 523 of *Astronomical Society of the Pacific Conference Series*, page 261.

Devorkin, D. H. (2000). *Henry Norris Russell: dean of American astronomers*. Princeton University Press.

Dhabal, A., Mundy, L. G., Rizzo, M. J., Storm, S., and Teuben, P. (2018). Morphology and Kinematics of Filaments in the Serpens and Perseus Molecular Clouds. *ApJ*, 853(2):169.

Dipierro, G., Price, D., Laibe, G., Hirsh, K., Cerioli, A., and Lodato, G. (2015). On planet formation in HL Tau. *MNRAS*, 453(1):L73–L77.

Doering, R. L. and Meixner, M. (2009). Near-Infrared Photometric Survey of Herbig Ae/Be Candidate Stars. *AJ*, 138:780–786.

Donati, J. F., Gregory, S. G., Alencar, S. H. P., Hussain, G., Bouvier, J., Jardine, M. M., Ménard, F., Dougados, C., Romanova, M. M., and MaPP Collaboration (2013). Magnetospheric accretion on the fully convective classical T Tauri star DN Tau. *MNRAS*, 436(1):881–897.

Donati, J. F., Gregory, S. G., Montmerle, T., Maggio, A., Argiroffi, C., Sacco, G., Hussain, G., Kastner, J., Alencar, S. H. P., Audard, M., Bouvier, J., Damiani, F., Güdel, M., Huenemoerder, D., and Wade, G. A. (2011). The close classical T Tauri binary V4046 Sgr: complex magnetic fields and distributed mass accretion. *MNRAS*, 417(3):1747–1759.

Donehew, B. and Brittain, S. (2011). Measuring the Stellar Accretion Rates of Herbig Ae/Be Stars. *AJ*, 141:46.

Dong, R., Najita, J. R., and Brittain, S. (2018). Spiral Arms in Disks: Planets or Gravitational Instability? *ApJ*, 862(2):103.

Drew, J. E., Gonzalez-Solares, E., Greimel, R., Irwin, M. J., Küpcü Yoldas, A., Lewis, J., Barentsen, G., Eislöffel, J., Farnhill, H. J., Martin, W. E., Walsh, J. R., Walton, N. A., Mohr-Smith, M., Raddi, R., Sale, S. E., Wright, N. J., Groot, P., Barlow, M. J., Corradi, R. L. M., Drake, J. J., Fabregat, J., Frew, D. J., Gänsicke, B. T., Knigge, C., Mampaso, A., Morris, R. A. H., Naylor, T., Parker, Q. A., Phillipps, S., Ruhland, C., et al. (2014). The VST Photometric Hα Survey of the Southern Galactic Plane and Bulge (VPHAS+). *MNRAS*, 440:2036–2058.

Drew, J. E., Greimel, R., Irwin, M. J., Aungwerojwit, A., Barlow, M. J., Corradi, R. L. M., Drake, J. J., Gänsicke, B. T., Groot, P., Hales, A., Hopewell, E. C., Irwin, J., Knigge, C., Leisy, P., Lennon, D. J., Mampaso, A., Masheder, M. R. W., Matsuura, M., Morales-Rueda, L., Morris, R. A. H., Parker, Q. A., Phillipps, S., Rodriguez-Gil, P., Roelofs, G., Skillen, I., Sokoloski, J. L., Steeghs, D., Unruh, Y. C., Viironen, K., Vink, J. S., et al. (2005). The INT Photometric Hα Survey of the Northern Galactic Plane (IPHAS). *MNRAS*, 362:753–776.

Dubus, G., Otulakowska-Hypka, M., and Lasota, J.-P. (2018). Testing the disk instability model of cataclysmic variables. *A&A*, 617:A26.

Duchêne, G. (2015). Herbig AeBe stars: multiplicity and consequences. *Ap&SS*, 355:291–301.

Duez, V. and Mathis, S. (2010). Relaxed equilibrium configurations to model fossil fields . I. A first family. *A&A*, 517:A58.

Dullemond, C. P. and Dominik, C. (2004a). Flaring vs. self-shadowed disks: The SEDs of Herbig Ae/Be stars. *A&A*, 417:159–168.

Dullemond, C. P. and Dominik, C. (2004b). The effect of dust settling on the appearance of protoplanetary disks. *A&A*, 421:1075–1086.

Dullemond, C. P. and Dominik, C. (2005). Dust coagulation in protoplanetary disks: A rapid depletion of small grains. *A&A*, 434:971–986.

Dunhill, A. C., Cuadra, J., and Dougados, C. (2015). Precession and accretion in circumbinary discs: the case of HD 104237. *MNRAS*, 448:3545–3554.

Dunkin, S. K., Barlow, M. J., and Ryan, S. G. (1997). High-resolution spectroscopy of Vega-like stars - II. Age indicators, activity and circumstellar gas. *MNRAS*, 290:165–185.

Eddington, A. S. (1926). *The Internal Constitution of the Stars*. Cambridge University Press.

Eiroa, C., Oudmaijer, R. D., Davies, J. K., de Winter, D., Garzón, F., Palacios, J., Alberdi, A., Ferlet, R., Grady, C. A., Collier Cameron, A., Deeg, H. J., Harris, A. W., Horne, K., Merín, B., Miranda, L. F., Montesinos, B., Mora, A., Penny, A., Quirrenbach, A., Rauer, H., Schneider, J., Solano, E., Tsapras, Y., and Wesselius, P. R. (2002). On the simultaneous optical and near-infrared variability of pre-main sequence stars. *A&A*, 384:1038–1049.

El-Badry, K. and Quataert, E. (2020). A stripped-companion origin for Be stars: clues from the putative black holes HR 6819 and LB-1. *arXiv e-prints*, page arXiv:2006.11974.

Emeriau, C. and Mathis, S. (2015). Impact of rotation on the geometrical configurations of fossil magnetic fields. In Meynet, G., Georgy, C., Groh, J., and Stee, P., editors, *New Windows on Massive Stars*, volume 307 of *IAU Symposium*, pages 373–374.

Engels, D. and Bunzel, F. (2015). A database of circumstellar OH masers. *A&A*, 582:A68.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press.

Evans, D. W., Riello, M., De Angeli, F., Carrasco, J. M., Montegriffo, P., Fabricius, C., Jordi, C., Palaversa, L., Diener, C., Busso, G., Cacciari, C., van Leeuwen, F., Burgess, P. W., Davidson, M., Harrison, D. L., Hodgkin, S. T., Pancino, E., Richards, P. J., Altavilla, G., Balaguer-Núñez, L., Barstow, M. A., Bellazzini, M., Brown, A. G. A., Castellani, M., Cocozza, G., De Luise, F., Delgado, A., Ducourant, C., Galleti, S., Gilmore, G., et al. (2018). Gaia Data Release 2. Photometric content and validation. *A&A*, 616:A4.

Fairlamb, J. R., Oudmaijer, R. D., Mendigutía, I., Ilee, J. D., and 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.

Fairlamb, J. R., Oudmaijer, R. D., Mendigutia, I., Ilee, J. D., and van den Ancker, M. E. (2017). A spectroscopic survey of Herbig Ae/Be stars with X-Shooter - II. Accretion diagnostic lines. *MNRAS*, 464:4721–4735.

Fang, M., Kim, J. S., van Boekel, R., Sicilia-Aguilar, A., Henning, T., and Flaherty, K. (2013). Young Stellar Objects in Lynds 1641: Disks, Accretion, and Star Formation History. *The Astrophysical Journal Supplement Series*, 207:5.

Federrath, C. (2016). On the universality of interstellar filaments: theory meets simulations and observations. *MNRAS*, 457(1):375–388.

Federrath, C., Schrön, M., Banerjee, R., and Klessen, R. S. (2014). Modeling Jet and Outflow Feedback during Star Cluster Formation. *ApJ*, 790(2):128.

Feiden, G. A. (2016). Magnetic inhibition of convection and the fundamental properties of low-mass stars. III. A consistent 10 Myr age for the Upper Scorpius OB association. *A&A*, 593:A99.

Finkenzeller, U. and Mundt, R. (1984). The Herbig Ae/Be stars associated with nebulosity. *Astronomy and Astrophysics Supplement Series*, 55:109–141.

Folsom, C. P., Bouvier, J., Petit, P., Lèbre, A., Amard, L., Palacios, A., Morin, J., Donati, J. F., and Vidotto, A. A. (2018). The evolution of surface magnetic fields in young solar-type stars II: the early main sequence (250-650 Myr). *MNRAS*, 474(4):4956–4987.

Folsom, C. P., Petit, P., Bouvier, J., Lèbre, A., Amard, L., Palacios, A., Morin, J., Donati, J. F., Jeffers, S. V., Marsden, S. C., and Vidotto, A. A. (2016). The evolution of surface magnetic fields in young solar-type stars - I. The first 250 Myr. *MNRAS*, 457(1):580–607.

Folsom, C. P., Wade, G. A., Kochukhov, O., Alecian, E., Catala, C., Bagnulo, S., Landstreet, J. D., and Hanes, D. (2008). Magnetic, chemical and rotational properties of the Herbig Ae/Be binary system HD72106. *Contributions of the Astronomical Observatory Skalnate Pleso*, 38:245–250.

Fontani, F., Commerçon, B., Giannetti, A., Beltrán, M. T., Sánchez-Monge, A., Testi, L., Brand, J., Caselli, P., Cesaroni, R., Dodson, R., Longmore, S., Rioja, M., Tan, J. C., and Walmsley, C. M. (2016). Magnetically regulated fragmentation of a massive, dense, and turbulent clump. *A&A*, 593:L14.

Frasca, A., Miroshnichenko, A. S., Rossi, C., Friedjung, M., Marilli, E., Muratorio, G., and Busà, I. (2016). Interpreting the spectral behavior of MWC 314. *A&A*, 585:A60.

Frew, D. J. and Parker, Q. A. (2010). Planetary Nebulae: Observational Properties, Mimics and Diagnostics. *PASA*, 27(2):129–148.

Friedemann, C., Guertler, J., and Loewe, M. (1996). Eclipsing binaries as IRAS sources. *Astronomy and Astrophysics Supplement Series*, 117:205–225.

Frost, A. J. (2020). *The circumstellar environment of massive young stellar objects - a multi-scale analysis*. PhD thesis, University of Leeds.

Fuente, A., Martin-Pintado, J., Bachiller, R., Neri, R., and Palla, F. (1998). Progressive dispersal of the dense gas in the environment of early-type and late-type Herbig Ae-Be stars. *A&A*, 334:253–263.

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., Barstow, M. A., Jordi, C., Vallenari, A., Bossini, D., Bressan, A., Cantat-Gaudin, T., van Leeuwen, M., Brown, A. G. A., Prusti, T., de Bruijne, J. H. J., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., Walton, N. A., et al. (2018a). Gaia Data Release 2. Observational Hertzsprung-Russell diagrams. *A&A*, 616:A10.

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., van Leeuwen, F., Walton, N. A., Arenou, F., Bastian, U., Cropper, M., Drimmel, R., Katz, D., et al. (2018b). Gaia Data Release 2. Summary of the contents and survey properties. *A&A*, 616:A1.

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., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Milligan, D. J., Panem, C., Poinsignon, V., Pourbaix, D., Randich, S., Sarri, G., Sartoretti, P., Siddiqui, H. I., Soubiran, C., Valette, V., van Leeuwen, F., Walton, N. A., et al. (2016). The Gaia mission. *A&A*, 595:A1.

Galli, P. A. B., Loinard, L., Bouy, H., Sarro, L. M., Ortiz-León, G. N., Dzib, S. A., Olivares, J., Heyer, M., Hernandez, J., Román-Zúñiga, C., Kounkel, M., and Covey, K. (2019). Structure and kinematics of the Taurus star-forming region from Gaia-DR2 and VLBI astrometry. *A&A*, 630:A137.

Garufi, A., Benisty, M., Pinilla, P., Tazzari, M., Dominik, C., Ginski, C., Henning, T., Kral, Q., Langlois, M., Ménard, F., Stolker, T., Szulagyi, J., Villenave, M., and van der Plas, G. (2018). Evolution of protoplanetary disks from their taxonomy in scattered light: spirals, rings, cavities, and shadows. *A&A*, 620:A94.

Geen, S., Soler, J. D., and Hennebelle, P. (2017). Interpreting the star formation efficiency of nearby molecular clouds with ionizing radiation. *MNRAS*, 471(4):4844–4855.

Gilmore, G., Randich, S., Asplund, M., Binney, J., Bonifacio, P., Drew, J., Feltzing, S., Ferguson, A., Jeffries, R., Micela, G., Negueruela, I., Prusti, T., Rix, H. W., Vallenari, A., Alfaro, E., Allende-Prieto, C., Babusiaux, C., Bensby, T., Blomme, R., Bragaglia, A., Flaccomio, E., François, P., Irwin, M., Koposov, S., Korn, A., Lanzafame, A., Pancino, E., Paunzen, E., Recio-Blanco, A., Sacco, G., et al. (2012). The Gaia-ESO Public Spectroscopic Survey. *The Messenger*, 147:25–31.

Gkouvelis, L., Fabregat, J., Zorec, J., Steeghs, D., Drew, J. E., Raddi, R., Wright, N. J., and Drake, J. J. (2016). Physical parameters of IPHAS-selected classical Be stars. I. Determination procedure and evaluation of the results. *A&A*, 591:A140.

Gorti, U., Dullemond, C. P., and Hollenbach, D. (2009). Time Evolution of Viscous Circumstellar Disks due to Photoevaporation by Far-Ultraviolet, Extreme-Ultraviolet, and X-ray Radiation from the Central Star. *ApJ*, 705:1237–1251.

Grady, C. A., Hamaguchi, K., Schneider, G., Stecklum, B., Woodgate, B. E., McCleary, J. E., Williger, G. M., Sitko, M. L., Ménard, F., Henning, T., Brittain, S., Troutmann, M., Donehew, B., Hines, D., Wisniewski, J. P., Lynch, D. K., Russell, R. W., Rudy, R. J., Day, A. N., Shenoy, A., Wilner, D., Silverstone, M., Bouret, J. C., Meusinger, H., Clampin, M., Kim, S., Petre, R., Sahu, M., Endres, M., and Collins, K. A. (2010). Locating the Accretion Footprint on a Herbig Ae Star: MWC 480. *ApJ*, 719:1565–1581.

Gray, R. O. and Corbally, Christopher, J. (2009). *Stellar Spectral Classification*. Princeton University Press.

Grinin, V. P. (2000). Photopolarimetric Activity of Pre-Main-Sequence Stars (Invited Review). In Garzón, G., Eiroa, C., de Winter, D., and Mahoney, T. J., editors, *Disks, Planetesimals, and Planets*, volume 219, page 216.

Grinin, V. P. and Rostopchina, A. N. (1996). Orientation of circumstellar disks and the statistics of Hα profiles of Ae/Be Herbig stars. *Astronomy Reports*, 40:171–178.

Gritschneder, M., Heigl, S., and Burkert, A. (2017). Oscillating Filaments. I. Oscillation and Geometrical Fragmentation. *ApJ*, 834(2):202.

Groenewegen, M. A. T., Sevenster, M., Spoon, H. W. W., and Pérez, I. (2002). Millimetre observations of infrared carbon stars. II. Mass loss rates and expansion velocities. *A&A*, 390:511–522.

Grundstrom, E. D. and Gies, D. R. (2006). Estimating Be Star Disk Radii using Hα Emission Equivalent Widths. *ApJ*, 651:L53–L56.

Gürtler, J., Friedemann, C., Reimann, H. G., Splittgerber, E., and Rudolph, E. (1999). A comparative study of the long-term light variations of six young irregular variables. *A&AS*, 140:293–307.

Guszejnov, D. and Hopkins, P. F. (2015). Mapping the core mass function to the initial mass function. *MNRAS*, 450(4):4137–4149.

Gvaramadze, V. V., Weidner, C., Kroupa, P., and Pflamm-Altenburg, J. (2012). Field O stars: formed in situ or as runaways? *MNRAS*, 424(4):3037–3049.

Hacar, A., Tafalla, M., and Alves, J. (2017). Fibers in the NGC 1333 proto-cluster. *A&A*, 606:A123.

Haffert, S. Y., Bohn, A. J., de Boer, J., Snellen, I. A. G., Brinchmann, J., Girard, J. H., Keller, C. U., and Bacon, R. (2019). Two accreting protoplanets around the young star PDS 70. *Nature Astronomy*, 3:749–754.

Hales, A. S., Corder, S. A., Dent, W. R. D., Andrews, S. M., Eisner, J. A., and Cieza, L. A. (2015). The Early ALMA View of the FU Ori Outburst System. *ApJ*, 812(2):134.

Hales, A. S., Pérez, S., Saito, M., Pinte, C., Knee, L. B. G., de Gregorio-Monsalvo, I., Dent, B., López, C., Plunkett, A., Cortés, P., Corder, S., and Cieza, L. (2018). The Circumstellar Disk and Asymmetric Outflow of the EX Lup Outburst System. *ApJ*, 859(2):111.

Hamaguchi, K., Choi, M., Corcoran, M. F., Choi, C.-S., Tatematsu, K., and Petre, R. (2008). Resolving a Class I Protostar Binary System with Chandra. *ApJ*, 687:425–432.

Hampton, E. J., Medling, A. M., Groves, B., Kewley, L., Dopita, M., Davies, R., Ho, I. T., Kaasinen, M., Leslie, S., Sharp, R., Sweet, S. M., Thomas, A. D., Allen, J., Bland-Hawthorn, J., Brough, S., Bryant, J. J., Croom, S., Goodwin, M., Green, A., Konstantantopoulos, I. S., Lawrence, J., López-Sánchez, Á. R., Lorente, N. P. F., McElroy, R., Owers, M. S., Richards, S. N., and Shastri, P. (2017). Using an artificial neural network to classify multicomponent emission lines with integral field spectroscopy from SAMI and S7. *MNRAS*, 470(3):3395–3416.

Hartmann, L., Herczeg, G., and Calvet, N. (2016). Accretion onto Pre-Main-Sequence Stars. *ARA&A*, 54:135–180.

Hayashi, C. (1966). Evolution of Protostars. *ARA&A*, 4:171.

Hedges, C., Hodgkin, S., and Kennedy, G. (2018). Discovery of new dipper stars with K2: a window into the inner disc region of T Tauri stars. *MNRAS*, 476(3):2968–2998.

Heinze, A. N., Tonry, J. L., Denneau, L., Flewelling, H., Stalder, B., Rest, A., Smith, K. W., Smartt, S. J., and Weiland, H. (2018). A First Catalog of Variable Stars Measured by the Asteroid Terrestrial-impact Last Alert System (ATLAS). *AJ*, 156(5):241.

Henden, A. A., Levine, S., Terrell, D., Welch, D. L., Munari, U., and Kloppenborg, B. K. (2018). APASS Data Release 10. In *American Astronomical Society Meeting Abstracts #232*, volume 232 of *American Astronomical Society Meeting Abstracts*, page 223.06.

Hennebelle, P., Commerçon, B., Joos, M., Klessen, R. S., Krumholz, M., Tan, J. C., and Teyssier, R. (2011). Collapse, outflows and fragmentation of massive, turbulent and magnetized prestellar barotropic cores. *A&A*, 528:A72.

Hennebelle, P. and Falgarone, E. (2012). Turbulent molecular clouds. *A&A Rev.*, 20:55.

Henning, T. and Semenov, D. (2013). Chemistry in Protoplanetary Disks. *Chemical Reviews*, 113(12):9016–9042.

Henyey, L. G., Lelevier, R., and Levée, R. D. (1955). The Early Phases of Stellar Evolution. *PASP*, 67(396):154.

Herbig, G. (2002). Historical introduction. Star formation: the early history. In Palla, F., Zinnecker, H., Maeder, A., and Meynet, G., editors, *Physics of Star Formation in Galaxies*, pages 1–7.

Herbig, G. H. (1960). The Spectra of Be- and Ae-Type Stars Associated with Nebulosity. *The Astrophysical Journal Supplement Series*, 4:337.

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

Herbig, G. H. (2008). History and Spectroscopy of EXor Candidates. *AJ*, 135(2):637–648.

Herbig, G. H. and Bell, K. R. (1988). *Third Catalog of Emission-Line Stars of the Orion Population : 3 : 1988*. Lick Observatory Bulletin 1111.

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

Hernández, J., Calvet, N., Hartmann, L., Briceño, C., Sicilia-Aguilar, A., and Berlind, P. (2005). Herbig Ae/Be Stars in nearby OB Associations. *AJ*, 129:856–871.

Hertzsprung, E. (1911). Über die Verwendung photographischer effektiver Wellenlaengen zur Bestimmung von Farbenäquivalenten. *Publikationen des Astrophysikalischen Observatoriums zu Potsdam*, 63.

Hillenbrand, L. A., Meyer, M. R., Strom, S. E., and Skrutskie, M. F. (1995). Isolated Star-Forming Regions Containing Herbig AE/BE Stars. I. The Young Stellar Aggregate Associated With BD +40 degrees 4124. *AJ*, 109:280.

Ho, T. K. (1995). C4.5 decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pages 278–282.

Hoffleit, D. and Jaschek, C. (1991). *The Bright star catalogue*. Yale University Observatory, 5th rev.ed.

Holl, B., Audard, M., Nienartowicz, K., Jevardat de Fombelle, G., Marchal, O., Mowlavi, N., Clementini, G., De Ridder, J., Evans, D. W., Guy, L. P., Lanzafame, A. C., Lebzelter, T., Rimoldini, L., Roelens, M., Zucker, S., Distefano, E., Garofalo, A., Lecoeur-Taïbi, I., Lopez, M., Molinaro, R., Muraveva, T., Panahi, A., Regibo, S., Ripepi, V., Sarro, L. M., Aerts, C., Anderson, R. I., Charnas, J., Barblan, F., Blanco-Cuaresma, S., et al. (2018). Gaia Data Release 2. Summary of the variability processing and analysis results. *A&A*, 618:A30.

Honda, M., Maaskant, K., Okamoto, Y. K., Kataza, H., Yamashita, T., Miyata, T., Sako, S., Fujiyoshi, T., Sakon, I., Fujiwara, H., Kamizuka, T., Mulders, G. D., Lopez-Rodriguez, E., Packham, C., and Onaka, T. (2015). High-resolution 25 μm Imaging of the Disks around Herbig Ae/Be Stars. *ApJ*, 804:143.

Hou, W., Luo, A. L., Hu, J.-Y., Yang, H.-F., Du, C.-D., Liu, C., Lee, C.-D., Lin, C.-C., Wang, Y.-F., Zhang, Y., Cao, Z.-H., and Hou, Y.-H. (2016). A catalog of early-type emission-line stars and Hα line profiles from LAMOST DR2. *Research in Astronomy and Astrophysics*, 16:138.

Hoyle, F. and Lyttleton, R. A. (1939). The evolution of the stars. *Proceedings of the Cambridge Philosophical Society*, 35(4):592.

Hubrig, S., Schöller, M., Savanov, I., González, J. F., Cowley, C. R., Schütz, O., Arlt, R., and Rüdiger, G. (2010). The exceptional Herbig Ae star HD 101412: The first detection of resolved magnetically split lines and the presence of chemical spots in a Herbig star. *Astronomische Nachrichten*, 331(4):361.

Ilee, J. D., Fairlamb, J., Oudmaijer, R. D., Mendigutía, I., van den Ancker, M. E., Kraus, S., and Wheelwright, H. E. (2014). Investigating the inner discs of Herbig Ae/Be stars with CO bandhead and Brγ emission. *MNRAS*, 445:3723–3736.

Inoue, T., Hennebelle, P., Fukui, Y., Matsumoto, T., Iwasaki, K., and Inutsuka, S.-i. (2018). The formation of massive molecular filaments and massive stars triggered by a magnetohydrodynamic shock wave. *PASJ*, 70:S53.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., Shappee, B. J., Holoien, T. W. S., Thompson, T. A., Prieto, J. L., Dong, S., Pawlak, M., Pejcha, O., Shields, J. V., Pojmanski, G., Otero, S., Hurst, N., Britt, C. A., and Will, D. (2019). The ASAS-SN catalogue of variable stars III: variables in the southern TESS continuous viewing zone. *MNRAS*, 485(1):961–971.

Jeans, J. (1928). *Astronomy and Cosmogony*. Cambridge University Press.

Jiang, D., Han, Z., Ge, H., Yang, L., and Li, L. (2012). The short-period limit of contact binaries. *MNRAS*, 421(4):2769–2773.

Jiang, Y.-F., Cantiello, M., Bildsten, L., Quataert, E., Blaes, O., and Stone, J. (2018). Outbursts of luminous blue variable stars from variations in the helium opacity. *Nature*, 561:498–501.

Joner, M. D. and Hintz, E. G. (2015). Standard Stars and Empirical Calibrations for Hα and Hβ Photometry. *AJ*, 150:204.

Joy, A. H. (1942). Spectral Criteria in the Classification of Variable Stars. *PASP*, 54(317):15.

Joy, A. H. (1945). T Tauri Variable Stars. *ApJ*, 102:168.

Joy, A. H. (1949). Bright-Line Stars among the Taurus Dark Clouds. *ApJ*, 110:424.

Kainulainen, J., Hacar, A., Alves, J., Beuther, H., Bouy, H., and Tafalla, M. (2016). Gravitational fragmentation caught in the act: the filamentary Musca molecular cloud. *A&A*, 586:A27.

Kashi, A. and Soker, N. (2010). Periastron Passage Triggering of the 19th Century Eruptions of Eta Carinae. *ApJ*, 723(1):602–611.

Kennedy, G. M. and Kenyon, S. J. (2008). Planet Formation around Stars of Various Masses: The Snow Line and the Frequency of Giant Planets. *ApJ*, 673(1):502–512.

Kerber, F., Mignani, R. P., Guglielmetti, F., and Wicenec, A. (2003). Galactic Planetary Nebulae and their central stars. I. An accurate and homogeneous set of coordinates. *A&A*, 408:1029–1035.

Khokhlov, S. A., Miroshnichenko, A. S., Zharikov, S. V., Manset, N., Arkharov, A. A., Efimova, N., Klimanov, S., Larionov, V. M., Kusakin, A. V., Kokumbaeva, R. I., Omarov, C. T., Kuratov, K. S., Kuratova, A. K., Rudy, R. J., Laag, E. A., Crawford, K. B., Swift, T. K., Puetter, R. C., Perry, R. B., Chojnowski, S. D., Agishev, A., Caton, D. B., Hawkins, R. L., Smith, A. B., Reichart, D. E., Kouprianov, V. V., and Haislip, J. B. (2018). Toward Understanding the B[e] Phenomenon. VII. AS 386, a Single-lined Binary with a Candidate Black Hole Component. *ApJ*, 856:158.

Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv e-prints*, page arXiv:1412.6980.

Kirk, H., Myers, P. C., Bourke, T. L., Gutermuth, R. A., Hedden, A., and Wilson, G. W. (2013). Filamentary Accretion Flows in the Embedded Serpens South Protocluster. *ApJ*, 766(2):115.

Klement, R., Carciofi, A. C., Rivinius, T., Ignace, R., Matthews, L. D., Torstensson, K., Gies, D., Vieira, R. G., Richardson, N. D., Domiciano de Souza, A., Bjorkman, J. E., Hallinan, G., Faes, D. M., Mota, B., Gullingsrud, A. D., de Breuck, C., Kervella, P., Curé, M., and Gunawan, D. (2019). Prevalence of SED Turndown among Classical Be Stars: Are All Be Stars Close Binaries? *ApJ*, 885(2):147.

Klement, R., Carciofi, A. C., Rivinius, T., Matthews, L. D., Vieira, R. G., Ignace, R., Bjorkman, J. E., Mota, B. C., Faes, D. M., Bratcher, A. D., Curé, M., and Štefl, S. (2017). Revealing the structure of the outer disks of Be stars. *A&A*, 601:A74.

Kobayashi, H., Kimura, H., Watanabe, S. i., Yamamoto, T., and Müller, S. (2011). Sublimation temperature of circumstellar dust particles and its importance for dust ring formation. *Earth, Planets, and Space*, 63(10):1067–1075.

Koenig, X. P. and Leisawitz, D. T. (2014). A Classification Scheme for Young Stellar Objects Using the Wide-field Infrared Survey Explorer AllWISE Catalog: Revealing Low-density Star Formation in the Outer Galaxy. *ApJ*, 791(2):131.

Kounkel, M. and Covey, K. (2019). Untangling the Galaxy. I. Local Structure and Star Formation History of the Milky Way. *AJ*, 158(3):122.

Kraus, S. (2015). The interferometric view of Herbig Ae/Be stars. *Ap&SS*, 357:97.

Kraus, S., Calvet, N., Hartmann, L., Hofmann, K.-H., Kreplin, A., Monnier, J. D., and Weigelt, G. (2012). On the Nature of the Herbig B[e] Star Binary System V921 Scorpii: Discovery of a Close Companion and Relation to the Large-scale Bipolar Nebula. *ApJ*, 746:L2.

Kraus, S., Hofmann, K.-H., Menten, K. M., Schertl, D., Weigelt, G., Wyrowski, F., Meilland, A., Perraut, K., Petrov, R., Robbe-Dubois, S., Schilke, P., and Testi, L. (2010). A hot compact dust disk around a massive young stellar object. *Nature*, 466(7304):339–342.

Kroupa, P. (2001). On the variation of the initial mass function. *MNRAS*, 322(2):231–246.

Kruijssen, J. M. D. (2012). On the fraction of star formation occurring in bound stellar clusters. *MNRAS*, 426(4):3008–3040.

Krumholz, M. R., Myers, A. T., Klein, R. I., and McKee, C. F. (2016). What physics determines the peak of the IMF? Insights from the structure of cores in radiation-magnetohydrodynamic simulations. *MNRAS*, 460(3):3272–3283.

Ksoll, V. F., Gouliermis, D. A., Klessen, R. S., Grebel, E. K., Sabbi, E., Anderson, J., Lennon, D. J., Cignoni, M., de Marchi, G., Smith, L. J., Tosi, M., and van der Marel, R. P. (2018). Hubble Tarantula Treasury Project - VI. Identification of pre-main- sequence stars using machine-learning techniques. *MNRAS*, 479:2389–2414.

Kubát, J., Saad, S. M., Kawka, A., Nouh, M. I., Iliev, L., Uytterhoeven, K., Korčáková, D., Hadrava, P., Škoda, P., Votruba, V., Dovčiak, M., and Šlechta, M. (2010). Spectroscopic analysis of the B/Be visual binary HR 1847. *A&A*, 520:A103.

Kurosawa, R., Harries, T. J., and Symington, N. H. (2006). On the formation of Hα line emission around classical T Tauri stars. *MNRAS*, 370(2):580–596.

Kurosawa, R., Romanova, M. M., and Harries, T. J. (2008). 3D simulations of rotationally-induced line variability from a classical T Tauri star with a misaligned magnetic dipole. *MNRAS*, 385:1931–1945.

Kurosawa, R., Romanova, M. M., and Harries, T. J. (2011). Multidimensional models of hydrogen and helium emission line profiles for classical T Tauri stars: method, tests and examples. *MNRAS*, 416(4):2623–2639.

Kurtz, D. W., Shibahashi, H., Murphy, S. J., Bedding, T. R., and Bowman, D. M. (2015). A unifying explanation of complex frequency spectra of γ Dor, SPB and Be stars: combination frequencies and highly non-sinusoidal light curves. *MNRAS*, 450(3):3015–3029.

Kučerová, B., Korčáková, D., Polster, J., Wolf, M., Votruba, V., Kubát, J., Škoda, P., Šlechta, M., and Křížek, M. (2013). Time-dependent spectral-feature variations of stars displaying the B[e] phenomenon. II. MWC 342. *A&A*, 554:A143.

Labadie-Bartz, J., Pepper, J., McSwain, M. V., Bjorkman, J. E., Bjorkman, K. S., Lund, M. B., Rodriguez, J. E., Stassun, K. G., Stevens, D. J., James, D. J., Kuhn, R. B., Siverd, R. J., and Beatty, T. G. (2017). Photometric Variability of the Be Star Population. *AJ*, 153:252.

Lallement, R., Babusiaux, C., Vergely, J. L., Katz, D., Arenou, F., Valette, B., Hottier, C., and Capitanio, L. (2019). Gaia-2MASS 3D maps of Galactic interstellar dust within 3 kpc. *A&A*, 625:A135.

Lamers, H. J. G. L. M., Zickgraf, F.-J., de Winter, D., Houziaux, L., and Zorec, J. (1998). An improved classification of B[e]-type stars. *A&A*, 340:117–128.

Landolt, A. U. (2009). UBVRI Photometric Standard Stars Around the Celestial Equator: Updates and Additions. *AJ*, 137:4186–4269.

Larson, R. B. (1981). Turbulence and star formation in molecular clouds. *MNRAS*, 194:809–826.

Larson, R. B. (2003). The physics of star formation. *Reports on Progress in Physics*, 66(10):1651–1697.

Lazareff, B., Berger, J. P., Kluska, J., Le Bouquin, J. B., Benisty, M., Malbet, F., Koen, C., Pinte, C., Thi, W. F., Absil, O., Baron, F., Delboulbé, A., Duvert, G., Isella, A., Jocou, L., Juhasz, A., Kraus, S., Lachaume, R., Ménard, F., Millan-Gabet, R., Monnier, J. D., Moulin, T., Perraut, K., Rochat, S., Soulez, F., Tallon, M., Thiébaut, E., Traub, W., and Zins, G. (2017). Structure of Herbig AeBe disks at the milliarcsecond scale . A statistical survey in the H band using PIONIER-VLTI. *A&A*, 599:A85.

Lee, C.-D., Chen, W.-P., and Liu, S.-Y. (2016). Evolutionary status of isolated B[e] stars. *A&A*, 592:A130.

Lee, C.-F. (2020). Molecular jets from low-mass young protostellar objects. *A&A Rev.*, 28(1):1.

Lee, Y.-N., Hennebelle, P., and Chabrier, G. (2017). Analytical Core Mass Function (CMF) from Filaments: Under Which Circumstances Can Filament Fragmentation Reproduce the CMF? *ApJ*, 847(2):114.

Leinert, C., Richichi, A., and Haas, M. (1997). Binaries among Herbig Ae/Be stars. *A&A*, 318:472–484.

Licquia, T. C. and Newman, J. A. (2015). Improved Estimates of the Milky Way's Stellar Mass and Star Formation Rate from Hierarchical Bayesian Meta-Analysis. *ApJ*, 806(1):96.

Lima, G. H. R. A., Alencar, S. H. P., Calvet, N., Hartmann, L., and Muzerolle, J. (2010). Modeling the Hα line emission around classical T Tauri stars using magnetospheric accretion and disk wind models. *A&A*, 522:A104.

Lindegren, L., Hernández, J., Bombrun, A., Klioner, S., Bastian, U., Ramos-Lerate, M., de Torres, A., Steidelmüller, H., Stephenson, C., Hobbs, D., Lammers, U., Biermann, M., Geyer, R., Hilger, T., Michalik, D., Stampa, U., McMillan, P. J., Castañeda, J., Clotet, M., Comoretto, G., Davidson, M., Fabricius, C., Gracia, G., Hambly, N. C., Hutton, A., Mora, A., Portell, J., van Leeuwen, F., Abbas, U., Abreu, A., et al. (2018). Gaia Data Release 2. The astrometric solution. *A&A*, 616:A2.

Liu, Q. Z., van Paradijs, J., and van den Heuvel, E. P. J. (2000). A catalogue of high-mass X-ray binaries. *Astronomy and Astrophysics Supplement Series*, 147:25–49.

Liu, Y., Dipierro, G., Ragusa, E., Lodato, G., Herczeg, G. J., Long, F., Harsono, D., Boehler, Y., Menard, F., Johnstone, D., Pascucci, I., Pinilla, P., Salyk, C., van der Plas, G., Cabrit, S., Fischer, W. J., Hendler, N., Manara, C. F., Nisini, B., Rigliaco, E., Avenhaus, H., Banzatti, A., and Gully-Santiago, M. (2019). Ring structure in the MWC 480 disk revealed by ALMA. *A&A*, 622:A75.

Long, F., Pinilla, P., Herczeg, G. J., Harsono, D., Dipierro, G., Pascucci, I., Hendler, N., Tazzari, M., Ragusa, E., Salyk, C., Edwards, S., Lodato, G., van de Plas, G., Johnstone, D., Liu, Y., Boehler, Y., Cabrit, S., Manara, C. F., Menard, F., Mulders, G. D., Nisini, B., Fischer, W. J., Rigliaco, E., Banzatti, A., Avenhaus, H., and Gully-Santiago, M. (2018). Gaps and Rings in an ALMA Survey of Disks in the Taurus Star-forming Region. *ApJ*, 869(1):17.

Lucas, W. E., Rybak, M., Bonnell, I. A., and Gieles, M. (2018). A clustered origin for isolated massive stars. *MNRAS*, 474(3):3582–3592.

Lumsden, S. L., Hoare, M. G., Urquhart, J. S., Oudmaijer, R. D., Davies, B., Mottram, J. C., Cooper, H. D. B., and Moore, T. J. T. (2013). The Red MSX Source Survey: The Massive Young Stellar Population of Our Galaxy. *The Astrophysical Journal Supplement Series*, 208:11.

Luo, A. L., Zhao, Y. H., Zhao, G., and et al. (2019). VizieR Online Data Catalog: LAMOST DR5 catalogs (Luo+, 2019). *VizieR Online Data Catalog*, page V/164.

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., and Delgado, H. E. (2018). Gaia Data Release 2. Using Gaia parallaxes. *A&A*, 616:A9.

Maaskant, K. M., Honda, M., Waters, L. B. F. M., Tielens, A. G. G. M., Dominik, C., Min, M., Verhoeff, A., Meeus, G., and van den Ancker, M. E. (2013). Identifying gaps in flaring Herbig Ae/Be disks using spatially resolved mid-infrared imaging. Are all group I disks transitional? *A&A*, 555:A64.

Maheswar, G., Manoj, P., and Bhatt, H. C. (2002). Circumstellar disks around Herbig Ae/Be stars: Polarization, outflows and binary orbits. *A&A*, 387:1003–1012.

Maíz Apellániz, J. and Weiler, M. (2018). Reanalysis of the Gaia Data Release 2 photometric sensitivity curves using HST/STIS spectrophotometry. *A&A*, 619:A180.

Małek, K., Solarz, A., Pollo, A., Fritz, A., Garilli, B., Scodeggio, M., Iovino, A., Granett, B. R., Abbas, U., Adami, C., Arnouts, S., Bel, J., Bolzonella, M., Bottini, D., Branchini, E., Cappi, A., Coupon, J., Cucciati, O., Davidzon, I., De Lucia, G., de la Torre, S., Franzetti, P., Fumana, M., Guzzo, L., Ilbert, O., Krywult, J., Le Brun, V., Le Fevre, O., Maccagni, D., Marulli, F., et al. (2013). The VIMOS Public Extragalactic Redshift Survey (VIPERS). A support vector machine classification of galaxies, stars, and AGNs. *A&A*, 557:A16.

Malkov, O. Y., Oblak, E., Snegireva, E. A., and Torra, J. (2006). A catalogue of eclipsing variables. *A&A*, 446:785–789.

Manara, C. F., Morbidelli, A., and Guillot, T. (2018). Why do protoplanetary disks appear not massive enough to form the known exoplanet population? *A&A*, 618:L3.

Manara, C. F., Natta, A., Rosotti, G. P., Alcalá, J. M., Nisini, B., Lodato, G., Testi, L., Pascucci, I., Hillenbrand, L., Carpenter, J., Scholz, A., Fedele, D., Frasca, A., Mulders, G., Rigliaco, E., Scardoni, C., and Zari, E. (2020). X-shooter survey of disk accretion in Upper Scorpius. I. Very high accretion rates at age > 5 Myr. *A&A*, 639:A58.

Manara, C. F., Testi, L., Herczeg, G. J., Pascucci, I., Alcalá, J. M., Natta, A., Antoniucci, S., Fedele, D., Mulders, G. D., Henning, T., Mohanty, S., Prusti, T., and Rigliaco, E. (2017). X-shooter study of accretion in Chamaeleon I. II. A steeper increase of accretion with stellar mass for very low-mass stars? *A&A*, 604:A127.

Manoj, P., Bhatt, H. C., Maheswar, G., and Muneer, S. (2006). Evolution of Emission-Line Activity in Intermediate-Mass Young Stars. *ApJ*, 653:657–674.

Marconi, M. and Palla, F. (1998). The Instability Strip for Pre-main-sequence Stars. *ApJ*, 507:L141–L144.

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., and Wood, P. R. (2017). A New Generation of PARSEC-COLIBRI Stellar Isochrones Including the TP- AGB Phase. *ApJ*, 835:77.

Marrese, P. M., Marinoni, S., Fabrizio, M., and Altavilla, G. (2019). Gaia Data Release 2. Cross-match with external catalogues: algorithms and results. *A&A*, 621:A144.

Marston, A. P. and McCollum, B. (2008). Extended shells around B[e] stars. Implications for B[e] star evolution. *A&A*, 477(1):193–202.

Martayan, C., Lobel, A., Baade, D., Mehner, A., Rivinius, T., Boffin, H. M. J., Girard, J., Mawet, D., Montagnier, G., Blomme, R., Kervella, P., Sana, H., Štefl, S., Zorec, J., Lacour, S., Le Bouquin, J.-B., Martins, F., Mérand , A., Patru, F., Selman, F., and Frémat, Y. (2016). Luminous blue variables: An imaging perspective on their binarity and near environment. *A&A*, 587:A115.

Marton, G., Ábrahám, P., Szegedi-Elek, E., Varga, J., Kun, M., Kóspál, Á., Varga-Verebélyi, E., Hodgkin, S., Szabados, L., Beck, R., and Kiss, C. (2019). Identification of Young Stellar Object candidates in the Gaia DR2 x AllWISE catalogue with machine learning methods. *MNRAS*, 487(2):2522–2537.

Marton, G., Tóth, L. V., Paladini, R., Kun, M., Zahorecz, S., McGehee, P., and Kiss, C. (2016). An all-sky support vector machine selection of WISE YSO candidates. *MNRAS*, 458(4):3479–3488.

Masci, F. J., Laher, R. R., Rusholme, B., Shupe, D. L., Groom, S., Surace, J., Jackson, E., Monkewitz, S., Beck, R., Flynn, D., Terek, S., Landry, W., Hacopians, E., Desai, V., Howell, J., Brooke, T., Imel, D., Wachter, S., Ye, Q.-Z., Lin, H.-W., Cenko, S. B., Cunningham, V., Rebbapragada, U., Bue, B., Miller, A. A., Mahabal, A., Bellm, E. C., Patterson, M. T., Jurić, M., Golkhou, V. Z., et al. (2019). The Zwicky Transient Facility: Data Processing, Products, and Archive. *PASP*, 131(995):018003.

Masunaga, H. and Inutsuka, S.-i. (2000). A Radiation Hydrodynamic Model for Protostellar Collapse. II. The Second Collapse and the Birth of a Protostar. *ApJ*, 531(1):350–365.

Mathew, B., Manoj, P., Narang, M., Banerjee, D. P. K., Nayak, P., Muneer, S., Vig, S., Pramod Kumar, S., Paul, K. T., and Maheswar, G. (2018). Excitation Mechanism of O I Lines in Herbig Ae/Be Stars. *ApJ*, 857:30.

Mathew, B., Subramaniam, A., and Bhatt, B. C. r. (2008). Be phenomenon in open clusters: results from a survey of emission-line stars in young open clusters. *MNRAS*, 388(4):1879–1888.

Matt, S. P., Pinzón, G., Greene, T. P., and Pudritz, R. E. (2012). Spin Evolution of Accreting Young Stars. II. Effect of Accretion-powered Stellar Winds. *ApJ*, 745(1):101.

Matthews, T. G., Ade, P. A. R., Angilè, F. E., Benton, S. J., Chapin, E. L., Chapman, N. L., Devlin, M. J., Fissel, L. M., Fukui, Y., Gandilo, N. N., Gundersen, J. O., Hargrave, P. C., Klein, J., Korotkov, A. L., Moncelsi, L., Mroczkowski, T. K., Netterfield, C. B., Novak, G., Nutter, D., Olmi, L., Pascale, E., Poidevin, F., Savini, G., Scott, D., Shariff, J. A., Soler, J. D., Tachihara, K., Thomas, N. E., Truch, M. D. P., Tucker, C. E., et al. (2014). Lupus I Observations from the 2010 Flight of the Balloon-borne Large Aperture Submillimeter Telescope for Polarimetry. *ApJ*, 784(2):116.

Mayer, A., Deschamps, R., and Jorissen, A. (2016). Search for systemic mass loss in Algols with bow shocks. *A&A*, 587:A30.

Mayer, J. R. (1842). Bemerkungen über die kräfte der unbelebten natur. *Justus Liebigs Annalen der Chemie*, 42(2):233–240.

Mayer, J. R. (1848). Beiträge zur Dynamik des Himmels in populärer Darstellung. *Published by J. U. Landherr*.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv e-prints*, page arXiv:1802.03426.

Meeus, G., Waters, L. B. F. M., Bouwman, J., van den Ancker, M. E., Waelkens, C., and Malfait, K. (2001). ISO spectroscopy of circumstellar dust in 14 Herbig Ae/Be systems: Towards an understanding of dust processing. *A&A*, 365:476–490.

Mendigutía, I. (2020). On the Mass Accretion Rates of Herbig Ae/Be Stars. Magnetospheric Accretion or Boundary Layer? *Galaxies*, 8(2):39.

Mendigutía, I., Calvet, N., Montesinos, B., Mora, A., Muzerolle, J., Eiroa, C., Oudmaijer, R. D., and Merín, B. (2011a). Accretion rates and accretion tracers of Herbig Ae/Be stars. *A&A*, 535:A99.

Mendigutía, I., Eiroa, C., Montesinos, B., Mora, A., Oudmaijer, R. D., Merín, B., and Meeus, G. (2011b). Optical spectroscopic variability of Herbig Ae/Be stars. *A&A*, 529:A34.

Mendigutía, I., Mora, A., Montesinos, B., Eiroa, C., Meeus, G., Merín, B., and Oudmaijer, R. D. (2012). Accretion-related properties of Herbig Ae/Be stars. Comparison with T Tauris. *A&A*, 543:A59.

Mendigutía, I., Oudmaijer, R. D., Garufi, A., Lumsden, S. L., Huélamo, N., Cheetham, A., de Wit, W. J., Norris, B., Olguin, F. A., and Tuthill, P. (2017). The protoplanetary system HD 100546 in Hα polarized light from SPHERE/ZIMPOL. A bar-like structure across the disk gap? *A&A*, 608:A104.

Mendigutía, I., Oudmaijer, R. D., Rigliaco, E., Fairlamb, J. R., Calvet, N., Muzerolle, J., Cunningham, N., and Lumsden, S. L. (2015). On the origin of the correlations between the accretion luminosity and emission line luminosities in pre-main-sequence stars. *MNRAS*, 452:2837–2844.

Mendigutía, I., Oudmaijer, R. D., Schneider, P. C., Huélamo, N., Baines, D., Brittain, S. D., and Aberasturi, M. (2018). Spectro-astrometry of the pre-transitional star LkCa 15 does not reveal an accreting planet but extended Hα emission. *A&A*, 618:L9.

Millour, F., Chesneau, O., Borges Fernandes, M., Meilland , A., Mars, G., Benoist, C., Thiébaut, E., Stee, P., Hofmann, K. H., Baron, F., Young, J., Bendjoya, P., Carciofi, A., Domiciano de Souza, A., Driebe, T., Jankov, S., Kervella, P., Petrov, R. G., Robbe-Dubois, S., Vakili, F., Waters, L. B. F. M., and Weigelt, G. (2009). A binary engine fuelling HD 87643's complex circumstellar environment. Determined using AMBER/VLTI imaging. *A&A*, 507(1):317–326.

Miroshnichenko, A. S. (2007). Toward Understanding the B[e] Phenomenon. I. Definition of the Galactic FS CMa Stars. *ApJ*, 667:497–504.

Miroshnichenko, A. S. (2017). FS CMa Type Objects - Products of Intermediate-Mass Non-Conservative Evolution. In Miroshnichenko, A., Zharikov, S., Korčáková, D., and Wolf, M., editors, *The B[e] Phenomenon: Forty Years of Studies*, volume 508 of *Astronomical Society of the Pacific Conference Series*, page 285.

Miroshnichenko, A. S., Bjorkman, K. S., Chentsov, E. L., Klochkova, V. G., Manset, N., García-Lario, P., Perea Calderón, J. V., Rudy, R. J., Lynch, D. K., Wilson, J. C., and Gandet, T. L. (2002). V669 Cep: A new binary system with a B[e] star. *A&A*, 388:563–572.

Miroshnichenko, A. S., Chentsov, E. L., Klochkova, V. G., Kuratov, K. S., Sheikina, T. A., Mukanov, D. B., Bjorkman, K. S., Gray, R. O., Rudy, R. J., Lynch, D. K., Mazuk, S., Puetter, R., García-Lario, P., Perea, J. V., and Bergner, Y. K. (2000). Spectroscopy and photometry of the emission-line B-type stars AS 78 and MWC 657. *A&AS*, 147:5–24.

Miroshnichenko, A. S., Fremat, Y., Houziaux, L., Andrillat, Y., Chentsov, E. L., and Klochkova, V. G. (1998). High resolution spectroscopy of the galactic candidate LBV MWC 314. *A&AS*, 131:469–478.

Miroshnichenko, A. S., Gray, R. O., Vieira, S. L. A., Kuratov, K. S., and Bergner, Y. K. (1999). Observations of recently recognized candidate Herbig Ae/Be stars. *A&A*, 347:137–150.

Miroshnichenko, A. S., Levato, H., Bjorkman, K. S., Grosso, M., Manset, N., Men'shchikov, A. B., Rudy, R. J., Lynch, D. K., Mazuk, S., Venturini, C. C., Puetter, R. C., and Perry, R. B. (2004). Properties of galactic B[e] supergiants. III. MWC 300. *A&A*, 417:731–743.

Miroshnichenko, A. S., Manset, N., Kusakin, A. V., Chentsov, E. L., Klochkova, V. G., Zharikov, S. V., Gray, R. O., Grankin, K. N., Gandet, T. L., Bjorkman, K. S., Rudy, R. J., Lynch, D. K., Venturini, C. C., Mazuk, S., Puetter, R. C., Perry, R. B., Levato, H., Grosso, M., Bernabei, S., Polcaro, V. F., Viotti, R. F., Norci, L., and Kuratov, K. S. (2007). Toward Understanding the B[e] Phenomenon. II. New Galactic FS CMa Stars. *ApJ*, 671:828–841.

Miroshnichenko, A. S., Manset, N., Zharikov, S. V., Zsargo, J., Juarez Jimenez, J. A., Groh, J. H., Levato, H., Grosso, M., Rudy, R. J., Laag, E. A., Crawford, K. B.,

Puetter, R. C., Reichart, D. E., Ivarsen, K. M., Haislip, J. B., Nysewander, M. C., and Lacluyze, A. P. (2014). Confirmation of the luminous blue variable status of MWC 930. *Advances in Astronomy*, 2014:E7.

Miroshnichenko, A. S., Polcaro, V. F., Rossi, C., Zharikov, S. V., Rudy, R. J., Kusakin, A. V., Kuratov, K. S., Naurzbayeva, A. Z., Alimgazinova, N. S., Manapbayeva, A. B., and Manset, N. (2017). Observations of Newly Found Objects With the B[e] Phenomenon. In Miroshnichenko, A., Zharikov, S., Korčáková, D., and Wolf, M., editors, *The B[e] Phenomenon: Forty Years of Studies*, volume 508 of *Astronomical Society of the Pacific Conference Series*, page 387.

Miroshnichenko, A. S. and Zharikov, S. V. (2015). FS CMa Type Binaries. In *EAS Publications Series*, volume 71-72 of *EAS Publications Series*, pages 181–186.

Molinari, S., Swinyard, B., Bally, J., Barlow, M., Bernard, J. P., Martin, P., Moore, T., Noriega-Crespo, A., Plume, R., Testi, L., Zavagno, A., Abergel, A., Ali, B., Anderson, L., André, P., Baluteau, J. P., Battersby, C., Beltrán, M. T., Benedettini, M., Billot, N., Blommaert, J., Bontemps, S., Boulanger, F., Brand, J., Brunt, C., Burton, M., Calzoletti, L., Carey, S., Caselli, P., Cesaroni, R., et al. (2010). Clouds, filaments, and protostars: The Herschel Hi-GAL Milky Way. *A&A*, 518:L100.

Monguió, M., Greimel, R., Drew, J. E., Barentsen, G., Groot, P. J., Irwin, M. J., Casares, J., Gänsicke, B. T., Carter, P. J., Corral-Santana, J. M., Gentile-Fusillo, N. P., Greiss, S., van Haaften, L. M., Hollands, M., Jones, D., Kupfer, T., Manser, C. J., Murphy, D. N. A., McLeod, A. F., Oosting, T., Parker, Q. A., Pyrzas, S., Rodríguez-Gil, P., van Roestel, J., Scaringi, S., Schellart, P., Toloza, O., Vaduvescu, O., van Spaand onk, L., Verbeek, K., et al. (2020). IGAPS: the merged IPHAS and UVEX optical surveys of the northern Galactic plane. *A&A*, 638:A18.

Monnier, J. D., Millan-Gabet, R., Billmeier, R., Akeson, R. L., Wallace, D., Berger, J. P., Calvet, N., D'Alessio, P., Danchi, W. C., Hartmann, L., Hillenbrand, L. A., Kuchner, M., Rajagopal, J., Traub, W. A., Tuthill, P. G., Boden, A., Booth, A., Colavita, M., Gathright, J., Hrynevych, M., Le Mignant, D., Ligon, R., Neyman, C., Swain, M., Thompson, R., Vasisht, G., Wizinowich, P., Beichman, C., Beletic, J., Creech-Eakman, M., et al. (2005). The Near-Infrared Size-Luminosity Relations for Herbig Ae/Be Disks. *ApJ*, 624:832–840.

Montesinos, B., Eiroa, C., Mora, A., and Merín, B. (2009). Parameters of Herbig Ae/Be and Vega-type stars. *A&A*, 495:901–917.

Morrell, N. and Levato, H. (1991). Spectroscopic Binaries in the Orion OB1 Association. *The Astrophysical Journal Supplement Series*, 75:965.

Motte, F., Nony, T., Louvet, F., Marsh, K. A., Bontemps, S., Whitworth, A. P., Men'shchikov, A., Nguyen Luong, Q., Csengeri, T., Maury, A. J., Gusdorf, A., Chapillon, E., Könyves, V., Schilke, P., Duarte-Cabral, A., Didelon, P., and Gaudel, M. (2018). The unexpectedly large proportion of high-mass star-forming cores in a Galactic mini-starburst. *Nature Astronomy*, 2:478–482.

Mottram, J. C., Hoare, M. G., Urquhart, J. S., Lumsden, S. L., Oudmaijer, R. D., Robitaille, T. P., Moore, T. J. T., Davies, B., and Stead, J. (2011). The Red MSX Source survey: the bolometric fluxes and luminosity distributions of young massive stars. *A&A*, 525:A149.

Mottram, J. C., Vink, J. S., Oudmaijer, R. D., and Patel, M. (2007). On the difference between Herbig Ae and Herbig Be stars. *MNRAS*, 377:1363–1374.

Mowlavi, N., Lecoeur-Taïbi, I., Lebzelter, T., Rimoldini, L., Lorenz, D., Audard, M., De Ridder, J., Eyer, L., Guy, L. P., Holl, B., Jevardat de Fombelle, G., Marchal, O., Nienartowicz, K., Regibo, S., Roelens, M., and Sarro, L. M. (2018). Gaia Data Release 2. The first Gaia catalogue of long-period variable candidates. *A&A*, 618:A58.

Munari, U., Joshi, V., Banerjee, D. P. K., Čotar, K., Shugarov, S. Y., Jurdana-Šepić, , R., Belligoli, R., Bergamini, A., Graziani, M., Righetti, G. L., Vagnozzi, A., and Valisa, P. (2019). The 2018 eruption and long-term evolution of the new high-mass Herbig Ae/Be object Gaia-18azl = VES 263. *MNRAS*, 488(4):5536–5550.

Muzerolle, J., Calvet, N., and Hartmann, L. (2001). Emission-Line Diagnostics of T Tauri Magnetospheric Accretion. II. Improved Model Tests and Insights into Accretion Physics. *ApJ*, 550(2):944–961.

Muzerolle, J., D'Alessio, P., Calvet, N., and Hartmann, L. (2004). Magnetospheres and Disk Accretion in Herbig Ae/Be Stars. *ApJ*, 617(1):406–417.

Nakano, M., Sugitani, K., Watanabe, M., Fukuda, N., Ishihara, D., and Ueno, M. (2012). Wide-field Survey of Emission-line Stars in IC 1396. *AJ*, 143:61.

Natta, A., Grinin, V. P., Mannings, V., and Ungerechts, H. (1997). The Evolutionary Status of UX Orionis-Type Stars. *ApJ*, 491:885–890.

Natta, A. and Whitney, B. A. (2000). Models of scattered light in UXORs. *A&A*, 364:633–640.

Neiner, C., de Batz, B., Cochard, F., Floquet, M., Mekkas, A., and Desnoux, V. (2011). The Be Star Spectra (BeSS) Database. *AJ*, 142(5):149.

Neiner, C., Lee, U., Mathis, S., Saio, H., Lovekin, C. C., and Augustson, K. C. (2020). Transport of angular momentum by stochastically excited waves as an explanation for the outburst of the rapidly rotating Be star HD49330. *arXiv e-prints*, page arXiv:2007.08977.

Ng, A. (2017). *Machine Learning Yearning*. Self-publishing: https://www.deeplearning.ai/machine-learning-yearning.

Offner, S. S. R. and Arce, H. G. (2014). Investigations of Protostellar Outflow Launching and Gas Entrainment: Hydrodynamic Simulations and Molecular Emission. *ApJ*, 784(1):61.

Offner, S. S. R. and Chaban, J. (2017). Impact of Protostellar Outflows on Turbulence and Star Formation Efficiency in Magnetized Dense Cores. *ApJ*, 847(2):104.

Offner, S. S. R., Clark, P. C., Hennebelle, P., Bastian, N., Bate, M. R., Hopkins, P. F., Moraux, E., and Whitworth, A. P. (2014). The Origin and Universality of the Stellar Initial Mass Function. In Beuther, H., Klessen, R. S., Dullemond, C. P., and Henning, T., editors, *Protostars and Planets VI*, page 53.

Oudmaijer, R. D. (2017). The B[e] Phenomenon in Pre-Main-Sequence Herbig Ae/Be Stars. In Miroshnichenko, A., Zharikov, S., Korčáková, D., and Wolf, M., editors, *The B[e] Phenomenon: Forty Years of Studies*, volume 508, page 175.

Oudmaijer, R. D., Busfield, G., and Drew, J. E. (1997). Diffuse interstellar bands in the spectra of massive young stellar objects. *MNRAS*, 291:797–804.

Oudmaijer, R. D. and Drew, J. E. (1997). An Hα outburst in the B emission line star HD 76534. *A&A*, 318:198–203.

Oudmaijer, R. D. and Drew, J. E. (1999). Hα spectropolarimetry of B[e] and Herbig Be stars. *MNRAS*, 305:166–180.

Oudmaijer, R. D. and Miroshnichenko, A. S. (2017). Introduction to the B[e] Phenomenon. In Miroshnichenko, A., Zharikov, S., Korčáková, D., and Wolf, M., editors, *The B[e] Phenomenon: Forty Years of Studies*, volume 508 of *Astronomical Society of the Pacific Conference Series*, page 3.

Oudmaijer, R. D., Palacios, J., Eiroa, C., Davies, J. K., de Winter, D., Ferlet, R., Garzón, F., Grady, C. A., Collier Cameron, A., Deeg, H. J., Harris, A. W., Horne, K., Merín, B., Miranda, L. F., Montesinos, B., Mora, A., Penny, A., Quirrenbach, A., Rauer, H., Schneider, J., Solano, E., Tsapras, Y., and Wesselius, P. R. (2001). EXPORT: Optical photometry and polarimetry of Vega-type and pre-main sequence stars. *A&A*, 379:564–578.

Oudmaijer, R. D., Proga, D., Drew, J. E., and de Winter, D. (1998). The evolved B[e] star HD 87643: observations and a radiation-driven disc wind model for B[e] stars. *MNRAS*, 300:170–182.

Palla, F. and Zinnecker, H. (1999). *Physics of Star Formation in Galaxies*. Springer.

Palmeirim, P., André, P., Kirk, J., Ward-Thompson, D., Arzoumanian, D., Könyves, V., Didelon, P., Schneider, N., Benedettini, M., Bontemps, S., Di Francesco, J., Elia, D., Griffin, M., Hennemann, M., Hill, T., Martin, P. G., Men'shchikov, A., Molinari, S., Motte, F., Nguyen Luong, Q., Nutter, D., Peretto, N., Pezzuto, S., Roy, A., Rygl, K. L. J., Spinoglio, L., and White, G. L. (2013). Herschel view of the Taurus B211/3 filament and striations: evidence of filamentary growth? *A&A*, 550:A38.

Panić, O. and Min, M. (2017). Effects of disc mid-plane evolution on CO snowline location. *MNRAS*, 467(1):1175–1185.

Pascucci, I., Testi, L., Herczeg, G. J., Long, F., Manara, C. F., Hendler, N., Mulders, G. D., Krijt, S., Ciesla, F., Henning, T., Mohanty, S., Drabek-Maunder, E., Apai, D., Szűcs, L., Sacco, G., and Olofsson, J. (2016). A Steeper than Linear Disk Mass-Stellar Mass Scaling Relation. *ApJ*, 831(2):125.

Pashchenko, I. N., Sokolovsky, K. V., and Gavras, P. (2018). Machine learning search for variable stars. *MNRAS*, 475(2):2326–2343.

Patel, P., Sigut, T. A. A., and Landstreet, J. D. (2017). Photoionization Models for the Inner Gaseous Disks of Herbig Be Stars: Evidence against Magnetospheric Accretion? *ApJ*, 836:214.

Payne, C. H. (1925). *Stellar Atmospheres; a Contribution to the Observational Study of High Temperature in the Reversing Layers of Stars.* PhD thesis, Radcliffe College.

Pecaut, M. J. and Mamajek, E. E. (2013). Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main- sequence Stars. *The Astrophysical Journal Supplement Series*, 208:9.

Pellegrini, E. W., Baldwin, J. A., and Ferland, G. J. (2011). Structure and Feedback in 30 Doradus. II. Structure and Chemical Abundances. *ApJ*, 738(1):34.

Peretto, N., Fuller, G. A., André, P., Arzoumanian, D., Rivilla, V. M., Bardeau, S., Duarte Puertas, S., Guzman Fernandez, J. P., Lenfestey, C., Li, G. X., Olguin, F. A., Röck, B. R., de Villiers, H., and Williams, J. (2014). SDC13 infrared dark clouds: Longitudinally collapsing filaments? *A&A*, 561:A83.

Peretto, N., Fuller, G. A., Duarte-Cabral, A., Avison, A., Hennebelle, P., Pineda, J. E., André, P., Bontemps, S., Motte, F., Schneider, N., and Molinari, S. (2013). Global collapse of molecular clouds as a formation mechanism for the most massive stars. *A&A*, 555:A112.

Pérez-Ortiz, M. F., García-Varela, A., Quiroz, A. J., Sabogal, B. E., and Hernández, J. (2017). Machine learning techniques to select Be star candidates. An application in the OGLE-IV Gaia south ecliptic pole field. *A&A*, 605:A123.

Pillai, T., Kauffmann, J., Tan, J. C., Goldsmith, P. F., Carey, S. J., and Menten, K. M. (2015). Magnetic Fields in High-mass Infrared Dark Clouds. *ApJ*, 799(1):74.

Pineda, J. E., Offner, S. S. R., Parker, R. J., Arce, H. G., Goodman, A. A., Caselli, P., Fuller, G. A., Bourke, T. L., and Corder, S. A. (2015). The formation of a quadruple star system with wide separation. *Nature*, 518(7538):213–215.

Pinte, C., van der Plas, G., Ménard, F., Price, D. J., Christiaens, V., Hill, T., Mentiplay, D., Ginski, C., Choquet, E., Boehler, Y., Duchêne, G., Perez, S., and Casassus, S. (2019). Kinematic detection of a planet carving a gap in a protoplanetary disk. *Nature Astronomy*, 3:1109–1114.

Pogodin, M. A., Hubrig, S., Yudin, R. V., Schöller, M., González, J. F., and Stelzer, B. (2012). Measuring the mass accretion rates of Herbig Ae/Be stars with X-shooter. *Astronomische Nachrichten*, 333:594–612.

Polster, J., Korčáková, D., and Manset, N. (2018). Time-dependent spectral-feature variations of stars displaying the B[e] phenomenon. IV. V2028 Cygni: modelling of Hα bisector variability. *A&A*, 617:A79.

Polster, J., Korčáková, D., Votruba, V., Škoda, P., Šlechta, M., Kučerová, B., and Kubát, J. (2012). Time-dependent spectral-feature variations of stars displaying the B[e] phenomenon. I. V2028 Cygni. *A&A*, 542:A57.

Pomohaci, R., Oudmaijer, R. D., Lumsden, S. L., Hoare, M. G., and Mendigutía, I. (2017). Medium-resolution near-infrared spectroscopy of massive young stellar objects. *MNRAS*, 472:3624–3636.

Portegies Zwart, S. F., McMillan, S. L. W., and Gieles, M. (2010). Young Massive Star Clusters. *ARA&A*, 48:431–493.

Povich, M. S., Maldonado, J. T., Haze Nuñez, E., and Robitaille, T. P. (2019). The Duration of Star Formation in Galactic Giant Molecular Clouds. I. The Great Nebula in Carina. *ApJ*, 881(1):37.

Poxon, M. (2015). UXOR Hunting among Algol Variables. *Journal of the American Association of Variable Star Observers (JAAVSO)*, 43:35.

Pudritz, R. E., Ouyed, R., Fendt, C., and Brandenburg, A. (2007). Disk Winds, Jets, and Outflows: Theoretical and Computational Foundations. In Reipurth, B., Jewitt, D., and Keil, K., editors, *Protostars and Planets V*, page 277.

Purser, S. J. D., Lumsden, S. L., Hoare, M. G., Urquhart, J. S., Cunningham, N., Purcell, C. R., Brooks, K. J., Garay, G., Gúzman, A. E., and Voronkov, M. A. (2016). A search for ionized jets towards massive young stellar objects. *MNRAS*, 460(1):1039–1053.

Quanz, S. P., Amara, A., Meyer, M. R., Girard, J. H., Kenworthy, M. A., and Kasper, M. (2015). Confirmation and Characterization of the Protoplanet HD 100546 b—Direct Evidence for Gas Giant Planet Formation at 50 AU. *ApJ*, 807(1):64.

Raddi, R., Drew, J. E., Fabregat, J., Steeghs, D., Wright, N. J., Sale, S. E., Farnhill, H. J., Barlow, M. J., Greimel, R., Sabin, L., Corradi, R. M. L., and Drake, J. J. (2013). First results of an Hα based search of classical Be stars in the Perseus Arm and beyond. *MNRAS*, 430(3):2169–2187.

Raddi, R., Drew, J. E., Steeghs, D., Wright, N. J., Drake, J. J., Barentsen, G., Fabregat, J., and Sale, S. E. (2015). A deep catalogue of classical Be stars in the direction of the Perseus Arm: spectral types and interstellar reddenings. *MNRAS*, 446:274–298.

Reipurth, B. (2016). George Herbig and Early Stellar Evolution. Institute for Astronomy Special Publications.

Reiter, M., Calvet, N., Thanathibodee, T., Kraus, S., Cauley, P. W., Monnier, J., Rubinstein, A., Aarnio, A., and Harries, T. J. (2018). Linking Signatures of Accretion with Magnetic Field Measurements-Line Profiles are not Significantly Different in Magnetic and Non- magnetic Herbig Ae/Be Stars. *ApJ*, 852:5.

Ribas, Á., Bouy, H., and Merín, B. (2015). Protoplanetary disk lifetimes vs. stellar mass and possible implications for giant planet populations. *A&A*, 576:A52.

Riello, M., De Angeli, F., Evans, D. W., Busso, G., Hambly, N. C., Davidson, M., Burgess, P. W., Montegriffo, P., Osborne, P. J., Kewley, A., Carrasco, J. M., Fabricius, C., Jordi, C., Cacciari, C., van Leeuwen, F., and Holland, G. (2018). Gaia Data Release 2. Processing of the photometric data. *A&A*, 616:A3.

Rimoldini, L., Holl, B., Audard, M., Mowlavi, N., Nienartowicz, K., Evans, D. W., Guy, L. P., Lecoeur-Taïbi, I., Jevardat de Fombelle, G., Marchal, O., Roelens, M., De Ridder, J., Sarro, L. M., Regibo, S., Lopez, M., Clementini, G., Ripepi, V., Molinaro, R., Garofalo, A., Molnár, L., Plachy, E., Juhász, Á., Szabados, L.,

Lebzelter, T., Teyssier, D., and Eyer, L. (2019). Gaia Data Release 2. All-sky classification of high-amplitude pulsating stars. *A&A*, 625:A97.

Rivinius, T., Carciofi, A. C., and Martayan, C. (2013). Classical Be stars. Rapidly rotating B stars with viscous Keplerian decretion disks. *Astronomy and Astrophysics Review*, 21:69.

Robin, A. C., Luri, X., Reylé, C., Isasi, Y., Grux, E., Blanco-Cuaresma, S., Arenou, F., Babusiaux, C., Belcheva, M., Drimmel, R., Jordi, C., Krone-Martins, A., Masana, E., Mauduit, J. C., Mignard, F., Mowlavi, N., Rocca- Volmerange, B., Sartoretti, P., Slezak, E., and Sozzetti, A. (2012). Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue. *A&A*, 543:A100.

Robitaille, T. P., Meade, M. R., Babler, B. L., Whitney, B. A., Johnston, K. G., Indebetouw, R., Cohen, M., Povich, M. S., Sewilo, M., Benjamin, R. A., and Churchwell, E. (2008). Intrinsically Red Sources Observed by Spitzer in the Galactic Midplane. *AJ*, 136(6):2413–2440.

Romanova, M. M., Ustyugova, G. V., Koldoba, A. V., and Lovelace, R. V. E. (2013). Warps, bending and density waves excited by rotating magnetized stars: results of global 3D MHD simulations. *MNRAS*, 430(1):699–724.

Rosen, A. L., Lopez, L. A., Krumholz, M. R., and Ramirez-Ruiz, E. (2014). Gone with the wind: Where is the missing stellar wind energy from massive star clusters? *MNRAS*, 442(3):2701–2716.

Rosenberg, H. (1910). Über den Zusammenhang von Helligkeit und Spektraltypus in den Plejaden. *Astronomische Nachrichten*, 186(5):71.

Russell, H. N. (1913). "Giant" and "dwarf" stars. *The Observatory*, 36:324–329.

Russell, H. N. (1914). Relations Between the Spectra and other Characteristics of the Stars. II. Brightness and Spectral Class. *Nature*, 93(2323):252–258.

Sabin, L., Parker, Q. A., Corradi, R. L. M., Guzman-Ramirez, L., Morris, R. A. H., Zijlstra, A. A., Bojičić, I. S., Frew, D. J., Guerrero, M., Stupar, M., Barlow, M. J., Cortés Mora, F., Drew, J. E., Greimel, R., Groot, P., Irwin, J. M., Irwin, M. J., Mampaso, A., Miszalski, B., Olguín, L., Phillipps, S., Santander García, M., Viironen, K., and Wright, N. J. (2014). First release of the IPHAS catalogue of new extended planetary nebulae. *MNRAS*, 443(4):3388–3401.

Sahman, D. I., Dhillon, V. S., Knigge, C., and Marsh, T. R. (2015). Searching for nova shells around cataclysmic variables. *MNRAS*, 451(3):2863–2876.

Sallum, S., Follette, K. B., Eisner, J. A., Close, L. M., Hinz, P., Kratter, K., Males, J., Skemer, A., Macintosh, B., Tuthill, P., Bailey, V., Defrère, D., Morzinski, K., Rodigas, T., Spalding, E., Vaz, A., and Weinberger, A. J. (2015). Accreting protoplanets in the LkCa 15 transition disk. *Nature*, 527(7578):342–344.

Salpeter, E. E. (1954). Reactions of Light Nuclei and Young Contracting Stars. In *Liege International Astrophysical Colloquia*, volume 5 of *Liege International Astrophysical Colloquia*, page 116.

Santiago-García, J., Tafalla, M., Johnstone, D., and Bachiller, R. (2009). Shells, jets, and internal working surfaces in the molecular outflow from IRAS 04166+2706. *A&A*, 495(1):169–181.

Sartori, M. J., Gregorio-Hetem, J., Rodrigues, C. V., Hetem, Annibal, J., and Batalha, C. (2010). Analysis of the Pico Dos Dias Survey Herbig Ae/Be Candidates. *AJ*, 139:27–38.

Scaringi, S., Knigge, C., Drew, J. E., Monguió, M., Breedt, E., Fratta, M., Gänsicke, B., Maccarone, T. J., Pala, A. F., and Schill, C. (2018). The Gaia/IPHAS and Gaia/KIS value-added catalogues. *MNRAS*, 481:3357–3369.

Schmidt, E. G., Johnston, D., Lee, K. M., Langan, S., Newman, P. R., and Snedden, S. A. (2004). The Spectra of Type II Cepheids. III. The Hα Line and Helium Emission in Long-Period Stars. *AJ*, 128(6):2988–2996.

Schneider, P. C., Günther, H. M., Robrade, J., Schmitt, J. H. M. M., and Güdel, M. (2018). Multiepoch, multiwavelength study of accretion onto T Tauri. X-ray versus optical and UV accretion tracers. *A&A*, 618:A55.

Schöller, M., Pogodin, M. A., Cahuasquí, J. A., Drake, N. A., Hubrig, S., Petr-Gotzens, M. G., Savanov, I. S., Wolff, B., González, J. F., Mysore, S., Ilyin, I., Järvinen, S. P., and Stelzer, B. (2016). Spectroscopic signatures of magnetospheric accretion in Herbig Ae/Be stars. I. The case of HD 101412. *A&A*, 592:A50.

Schütz, O., Meeus, G., Carmona, A., Juhász, A., and Sterzik, M. F. (2011). The young B-star quintuple system HD 155448. *A&A*, 533:A54.

Seifried, D., Pudritz, R. E., Banerjee, R., Duffin, D., and Klessen, R. S. (2012). Magnetic fields during the early stages of massive star formation - II. A generalized outflow criterion. *MNRAS*, 422(1):347–366.

Shokry, A., Rivinius, T., Mehner, A., Martayan, C., Hummel, W., Townsend, R. H. D., Mérand, A., Mota, B., Faes, D. M., Hamdy, M. A., Beheary, M. M., Gadallah, K. A. K., and Abo-Elazm, M. S. (2018). Stellar parameters of Be stars observed with X-shooter. *A&A*, 609:A108.

Sicilia-Aguilar, A., Kim, J. S., Sobolev, A., Getman, K., Henning, T., and Fang, M. (2013). The low-mass stellar population in the young cluster Tr 37. Disk evolution, accretion, and environment. *A&A*, 559:A3.

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., et al. (2006). The Two Micron All Sky Survey (2MASS). *AJ*, 131:1163–1183.

Smith, R. J., Glover, S. C. O., Klessen, R. S., and Fuller, G. A. (2016). On the nature of star-forming filaments - II. Subfilaments and velocities. *MNRAS*, 455(4):3640–3655.

Snider, S., Allende Prieto, C., von Hippel, T., Beers, T. C., Sneden, C., Qu, Y., and Rossi, S. (2001). Three-dimensional Spectral Classification of Low-Metallicity Stars Using Artificial Neural Networks. *ApJ*, 562(1):528–548.

Sokol, A. D., Gutermuth, R. A., Pokhrel, R., Gómez-Ruiz, A. I., Wilson, G. W., Offner, S. S. R., Heyer, M., Luna, A., Schloerb, F. P., and Sánchez, D. (2019). Early science with the Large Millimetre Telescope: An LMT/AzTEC 1.1 mm Survey of dense cores in the Monoceros R2 giant molecular cloud. *MNRAS*, 483(1):407–424.

Solarz, A., Bilicki, M., Gromadzki, M., Pollo, A., Durkalec, A., and Wypych, M. (2017). Automated novelty detection in the WISE survey with one-class support vector machines. *A&A*, 606:A39.

Spezzi, L., Alcalá, J. M., Covino, E., Frasca, A., Gandolfi, D., Oliveira, I., Chapman, N., Evans, Neal J., I., Huard, T. L., Jørgensen, J. K., Merín, B., and Stapelfeldt, K. R. (2008). The Young Population of the Chamaeleon II Dark Cloud. *ApJ*, 680:1295–1318.

Stahler, S. (2017). EXors and the stellar birthline. *Mem. Soc. Astron. Italiana*, 88:616.

Stahler, S. W. and Palla, F. (2004). *The Formation of Stars*. Wiley-VCH.

Stanghellini, L., Shaw, R. A., and Villaver, E. (2008). The Magellanic Cloud Calibration of the Galactic Planetary Nebula Distance Scale. *ApJ*, 689(1):194–202.

Steinmetz, M., Guiglion, G., McMillan, P. J., Matijevič, G., Enke, H., Kordopatis, G., Zwitter, T., Valentini, M., Chiappini, C., Casagrande, L., Wojno, J., Anguiano, B., Bienaymé, O., Bijaoui, A., Binney, J., Burton, D., Cass, P., de Laverny, P., Fiegert, K., Freeman, K., Fulbright, J. P., Gibson, B. K., Gilmore, G., Grebel, E. K., Helmi, A., Kunder, A., Munari, U., Navarro, J. F., Parker, Q., Ruchti, G. R., et al. (2020). The Sixth Data Release of the Radial Velocity Experiment (RAVE). II. Stellar Atmospheric Parameters, Chemical Abundances, and Distances. *AJ*, 160(2):83.

Stephens, I. W., Gouliermis, D., Looney, L. W., Gruendl, R. A., Chu, Y.-H., Weisz, D. R., Seale, J. P., Chen, C. H. R., Wong, T., Hughes, A., Pineda, J. L., Ott, J., and Muller, E. (2017). Stellar Clusterings around "Isolated" Massive YSOs in the LMC. *ApJ*, 834(1):94.

Suh, K.-W. and Hong, J. (2017). A New Catalog of AGB Stars Based on Infrared Two-Color Diagrams. *Journal of Korean Astronomical Society*, 50(4):131–138.

Szczerba, R., Siódmiak, N., Stasińska, G., and Borkowski, J. (2007). An evolutionary catalogue of galactic post-AGB and related objects. *A&A*, 469(2):799–806.

Tackenberg, J., Beuther, H., Henning, T., Linz, H., Sakai, T., Ragan, S. E., Krause, O., Nielbock, M., Hennemann, M., Pitann, J., and Schmiedeke, A. (2014). Kinematic structure of massive star-forming regions. I. Accretion along filaments. *A&A*, 565:A101.

Tafalla, M., Su, Y. N., Shang, H., Johnstone, D., Zhang, Q., Santiago-García, J., Lee, C. F., Hirano, N., and Wang, L. Y. (2017). Anatomy of the internal bow shocks in the IRAS 04166+2706 protostellar jet. *A&A*, 597:A119.

Tambovtseva, L. V., Grinin, V. P., and Weigelt, G. (2014). Hydrogen lines as a diagnostic tool for studying multicomponent emitting regions in hot young stars: magnetosphere, X-wind, and disk wind. *A&A*, 562:A104.

Testi, L., Palla, F., and Natta, A. (1999). The onset of cluster formation around Herbig Ae/Be stars. *A&A*, 342:515–523.

The, P. S., de Winter, D., and Perez, M. R. (1994). A new catalogue of members and candidate members of the Herbig Ae/Be (HAEBE) stellar group. *Astronomy and Astrophysics Supplement Series*, 104:315–339.

Tielens, A. G. G. M. (2005). *The Physics and Chemistry of the Interstellar Medium*. Cambridge University Press.

Torres, C. A. O., Quast, G. R., da Silva, L., and de La Reza, R. (2000). Binaries in two surveys of young stars and the case of PDS 1. In *IAU Symposium*, volume 200, page 118.

Tritsis, A. and Tassis, K. (2016). Striations in molecular clouds: streamers or MHD waves? *MNRAS*, 462(4):3602–3615.

Uemura, M., Kato, T., Ishioka, R., Yoshida, S., Kadota, K.-I., Ohkura, N., Henden, A., Pejcha, O., Kinugasa, K., Fujii, M., Simonsen, M., Greaves, J., Dubovsky, P. A., Poyner, G., West, D., Stine, R. J., Taylor, D., Poxon, M., Muyllaert, E., Ripero, J., Reszelski, M., and Jones, C. P. (2004). Deep Fading of the New Herbig Be Star MisV1147. *PASJ*, 56:S183–S192.

Valdivia, V., Hennebelle, P., Gérin, M., and Lesaffre, P. (2016). H_2 distribution during the formation of multiphase molecular clouds. *A&A*, 587:A76.

van den Ancker, M. E., de Winter, D., and Tjin A Djie, H. R. E. (1998). HIPPARCOS photometry of Herbig Ae/Be stars. *A&A*, 330:145–154.

van den Ancker, M. E., The, P. S., Feinstein, A., Vazquez, R. A., de Winter, D., and Perez, M. R. (1997). A multiwavelength study of star formation in the very young open cluster NGC 6530. *Astronomy and Astrophysics Supplement Series*, 123:63–82.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.

van der Marel, N., Dong, R., di Francesco, J., Williams, J. P., and Tobin, J. (2019). Protoplanetary Disk Rings and Gaps across Ages and Luminosities. *ApJ*, 872(1):112.

Venuti, L., Bouvier, J., Flaccomio, E., Alencar, S. H. P., Irwin, J., Stauffer, J. R., Cody, A. M., Teixeira, P. S., Sousa, A. P., Micela, G., Cuillandre, J. C., and Peres, G. (2014). Mapping accretion and its variability in the young open cluster NGC 2264: a study based on u-band photometry. *A&A*, 570:A82.

Vieira, R. G., Gregorio-Hetem, J., Hetem, A., Stasińska, G., and Szczerba, R. (2011). Distinguishing post-AGB impostors in a sample of pre-main sequence stars. *A&A*, 526:A24.

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., and da Silva, L. (2003). Investigation of 131 Herbig Ae/Be Candidate Stars. *AJ*, 126:2971–2987.

Villebrun, F., Alecian, E., Hussain, G., Bouvier, J., Folsom, C. P., Lebreton, Y., Amard, L., Charbonnel, C., Gallet, F., Haemmerlé, L., Böhm, T., Johns-Krull, C., Kochukhov, O., Marsden, S. C., Morin, J., and Petit, P. (2019). Magnetic fields of intermediate-mass T Tauri stars. I. Magnetic detections and fundamental stellar parameters. *A&A*, 622:A72.

Vink, J. S., Drew, J. E., Harries, T. J., and Oudmaijer, R. D. (2002). Probing the circumstellar structure of Herbig Ae/Be stars. *MNRAS*, 337:356–368.

Vioque, M., Oudmaijer, R. D., Baines, D., Mendigutía, I., and Pérez-Martínez, R. (2018). Gaia DR2 study of Herbig Ae/Be stars. *A&A*, 620:A128.

Vioque, M., Oudmaijer, R. D., Schreiner, M., Mendigutía, I., Baines, D., Mowlavi, N., and 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.

Vorobyov, E. I. and Basu, S. (2005). The Origin of Episodic Accretion Bursts in the Early Stages of Star Formation. *ApJ*, 633(2):L137–L140.

Walch, S., Girichidis, P., Naab, T., Gatto, A., Glover, S. C. O., Wünsch, R., Klessen, R. S., Clark, P. C., Peters, T., Derigs, D., and Baczynski, C. (2015). The SILCC (SImulating the LifeCycle of molecular Clouds) project - I. Chemical evolution of the supernova-driven ISM. *MNRAS*, 454(1):238–268.

Ward-Thompson, D. and Whitworth, A. P. (2011). *An Introduction to Star Formation*. Cambridge University Press.

Waters, L. B. F. M., Cote, J., and Geballe, T. R. (1988). 51 Ophiuchi (B9.5Ve) : a Be star in the class of beta Pictoris stars ? *A&A*, 203:348–354.

Waters, L. B. F. M., Cote, J., and Lamers, H. J. G. L. M. (1987). IRAS observations of Be stars. II. Far-IR characteristics and mass loss rates. *A&A*, 185:206–224.

Watkins, E. (2012). *Kant: Natural Science*. Cambridge University Press.

Watson, C. L., Henden, A. A., and Price, A. (2006). The International Variable Star Index (VSX). *Society for Astronomical Sciences Annual Symposium*, 25:47.

Weiler, M. (2018). Revised Gaia Data Release 2 passbands. *A&A*, 617:A138.

Wheelwright, H. E., Oudmaijer, R. D., and Goodwin, S. P. (2010). The mass ratio and formation mechanisms of Herbig Ae/Be star binary systems. *MNRAS*, 401:1199–1218.

Wheelwright, H. E., Vink, J. S., Oudmaijer, R. D., and Drew, J. E. (2011). On the alignment between the circumstellar disks and orbital planes of Herbig Ae/Be binary systems. *A&A*, 532:A28.

Whitworth, A. P. and Jaffa, S. E. (2018). A simple approach to CO cooling in molecular clouds. *A&A*, 611:A20.

Wichittanakom, C., Oudmaijer, R. D., Fairlamb, J. R., Mendigutía, I., Vioque, M., and Ababakr, K. M. (2020). The accretion rates and mechanisms of Herbig Ae/Be stars. *MNRAS*, 493(1):234–249.

Wilson, T. J. and Naylor, T. (2018). A contaminant-free catalogue of Gaia DR2-WISE Galactic plane matches: including the effects of crowding in the cross-matching of photometric catalogues. *MNRAS*, 481:2148–2167.

Wright, C. O., Egan, M. P., Kraemer, K. E., and Price, S. D. (2003). The Tycho-2 Spectral Type Catalog. *AJ*, 125(1):359–363.

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier, Thomas N., I., McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., et al. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *AJ*, 140(6):1868–1881.

Wu, B., Tan, J. C., Nakamura, F., Van Loo, S., Christie, D., and Collins, D. (2017). GMC Collisions as Triggers of Star Formation. II. 3D Turbulent, Magnetized Simulations. *ApJ*, 835(2):137.

Wuchterl, G. and Tscharnuter, W. M. (2003). From clouds to stars. Protostellar collapse and the evolution to the pre-main sequence I. Equations and evolution in the Hertzsprung-Russell diagram. *A&A*, 398:1081–1090.

Yuan, J., Li, J.-Z., Wu, Y., Ellingsen, S. P., Henkel, C., Wang, K., Liu, T., Liu, H.-L., Zavagno, A., Ren, Z., and Huang, Y.-F. (2018). High-mass Star Formation through Filamentary Collapse and Clump-fed Accretion in G22. *ApJ*, 852(1):12.

Zanni, C., Ferrari, A., Rosner, R., Bodo, G., and Massaglia, S. (2007). MHD simulations of jet acceleration from Keplerian accretion disks. The effects of disk resistivity. *A&A*, 469(3):811–828.

Zari, E., Brown, A. G. A., and de Zeeuw, P. T. (2019). Structure, kinematics, and ages of the young stellar populations in the Orion region. *A&A*, 628:A123.

Zhang, K., Blake, G. A., and Bergin, E. A. (2015). Evidence of Fast Pebble Growth Near Condensation Fronts in the HL Tau Protoplanetary Disk. *ApJ*, 806(1):L7.

Zhang, P., Yang, H. T., and Liu, J. (2006). An infrared study of Herbig Ae/Be stars. *Ap&SS*, 305:11–23.

Zhang, Q., Claus, B., Watson, L., and Moran, J. (2017). Angular Momentum in Disk Wind Revealed in the Young Star MWC 349A. *ApJ*, 837(1):53.

Zhang, Y., Tan, J. C., and Hosokawa, T. (2014). Radiation Transfer of Models of Massive Star Formation. III. The Evolutionary Sequence. *ApJ*, 788:166.

Zhu, Z., Hartmann, L., Calvet, N., Hernand ez, J., Muzerolle, J., and Tannirkulam, A.-K. (2007). The Hot Inner Disk of FU Orionis. *ApJ*, 669(1):483–492.

Zuckerman, B., Melis, C., Song, I., Meier, D. S., Perrin, M. D., Macintosh, B., Marois, C., Weinberger, A. J., Rhee, J. H., Graham, J. R., Kastner, J. H., Palmer, P., Forveille, T., Becklin, E. E., Wilner, D. J., Barman, T. S., Marcy, G. W., and Bessell, M. S. (2008). Gas and Dust Associated with the Strange, Isolated Star BP Piscium. *ApJ*, 683(2):1085–1103.

Zwintz, K., Fossati, L., and Ryabchikova, T. (2015). *Pre-Main-Sequence Stars and the Power of Asteroseismology*, volume 494 of *Astronomical Society of the Pacific Conference Series*, page 157.

Zwintz, K., Fossati, L., Ryabchikova, T., Guenther, D., Aerts, C., Barnes, T. G., Themeßl, N., Lorenz, D., Cameron, C., Kuschnig, R., Pollack-Drs, S., Moravveji, E., Baglin, A., Matthews, J. M., Moffat, A. F. J., Poretti, E., Rainer, M., Rucinski, S. M., Sasselov, D., and Weiss, W. W. (2014). Echography of young stars reveals their evolution. *Science*, 345(6196):550–553.

Zwintz, K., Kallinger, T., Guenther, D. B., Gruberbauer, M., Kuschnig, R., Weiss, W. W., Auvergne, M., Jorda, L., Favata, F., Matthews, J., and Fischer, M. (2011). Pulsational Analysis of V 588 Mon and V 589 Mon Observed with the MOST and CoRoT Satellites. *ApJ*, 729(1):20.

Appendix A

Parameters of known Herbig Ae/Be stars

This appendix contains Tables A.1, A.2, A.3, A.4, A.5, and A.6 which were introduced and discussed in Chapter 2. These tables tabulate the main compiled and derived parameters of the 252 known Herbig Ae/Be stars considered in that chapter. Tables A.1, A.2, and A.3 contain the data for the 218 sources of the high-quality sample (Sect. A.1) and Tables A.4, A.5, and A.6 contain the data for the 34 sources of the low-quality sample (Sect. A.2):

1. Tables A.1 and A.4 tabulate the main parameters: coordinates, parallax, distance, T_{eff}, luminosity, A_V, V magnitude, and binarity information.

2. Tables A.2 and A.5 tabulate other parameters: near- and mid-IR excess, Hα EW, Hα line shape, variability indicator, mass, and age.

3. Tables A.3 and A.6 tabulate the IR excess at each bandpass (J, H, K_s, W1, W2, W3, and W4) defined as $F_{observed}/F_{CK}$.

A.1 High-quality sample parameters

Table A.1 Main parameters of each known Herbig Ae/Be star belonging to the high-quality sample of 218 sources.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
AB Aur	04:55:45.9	+30:33:04	6.140 ± 0.057	$162.9^{+2.6}_{-2.4}$	9500^{+750}_{-790}	$1.61^{+0.19}_{-0.21}$	$0.43^{+0.28}_{-0.35}$	7.32	Yes[1]
AK Sco	16:54:44.8	-36:53:19	7.113 ± 0.062	$140.6^{+2.1}_{-2.0}$	6250^{+250}_{-250}	$0.623^{+0.028}_{-0.005}$	$0.000^{+0.050}_{-0.000}$	8.90	Yes[6]
AS 310	18:33:21.2	-04:58:06	0.390 ± 0.046	2110^{+350}_{-240}	24500^{+4500}_{-5000}	$4.17^{+0.39}_{-0.44}$	$4.13^{+0.26}_{-0.29}$	12.49	-
AS 470	21:36:14.2	+57:21:31	0.137 ± 0.027	4040^{+620}_{-440}	8200^{+1600}_{-800}	$3.01^{+0.47}_{-0.27}$	$2.27^{+0.62}_{-0.42}$	12.44	-
AS 477	21:52:34.1	+47:13:44	1.290 ± 0.029	773^{+30}_{-27}	10000^{+1700}_{-500}	$2.22^{+0.36}_{-0.12}$	$1.19^{+0.46}_{-0.12}$	10.05	Yes[4]
BD+30 549	03:29:19.8	+31:24:57	3.384 ± 0.083	295^{+13}_{-11}	11500^{+1500}_{-800}	$1.54^{+0.25}_{-0.14}$	$1.73^{+0.24}_{-0.12}$	10.51	-
BD+41 3731	20:24:15.7	+42:18:01	0.987 ± 0.035	1003^{+64}_{-53}	17000^{+1000}_{-1000}	$2.96^{+0.31}_{-0.31}$	$1.06^{+0.50}_{-0.50}$	9.90	-
BF Ori	05:37:13.3	-06:35:01	2.570 ± 0.053	389^{+14}_{-12}	8970^{+200}_{-200}	$1.287^{+0.060}_{-0.051}$	$0.330^{+0.030}_{-0.020}$	9.82	-
BH Cep	22:01:42.9	+69:44:36	2.984 ± 0.018	$335.1^{+3.4}_{-3.3}$	6630^{+210}_{-300}	$0.76^{+0.10}_{-0.14}$	$0.83^{+0.24}_{-0.35}$	11.23	-
BO Cep	22:16:54.1	+70:03:45	2.670 ± 0.020	$374.5^{+4.7}_{-4.5}$	6650^{+100}_{-100}	$0.467^{+0.088}_{-0.056}$	$0.12^{+0.20}_{-0.12}$	11.52	-
CO Ori	05:27:38.3	+11:25:39	2.474 ± 0.042	404^{+12}_{-11}	6250^{+150}_{-80}	$1.50^{+0.10}_{-0.09}$	$2.14^{+0.20}_{-0.19}$	11.13	Yes[4]
CPM 25	06:23:56.3	+14:30:28	0.293 ± 0.066	2130^{+470}_{-290}	19500^{+5000}_{-3000}	$2.85^{+0.51}_{-0.39}$	$3.83^{+0.29}_{-0.23}$	14.94	-
CQ Tau	05:35:58.5	+24:44:54	6.131 ± 0.082	$163.1^{+3.7}_{-3.5}$	6750^{+370}_{-200}	$0.87^{+0.18}_{-0.12}$	$0.41^{+0.42}_{-0.28}$	8.98	Yes[5]
DG Cir	15:03:23.8	-63:22:59	1.191 ± 0.042	833^{+52}_{-43}	11000^{+3000}_{-3000}	$1.58^{+0.33}_{-0.45}$	$3.94^{+0.13}_{-0.54}$	14.75	-
GSC 1876-0892	06:07:15.4	+29:57:55	0.154 ± 0.046	3000^{+580}_{-390}	19500^{+5000}_{-3000}	$3.85^{+0.49}_{-0.38}$	$4.64^{+0.29}_{-0.23}$	14.00	-
GSC 3975-0579	21:38:08.5	+57:26:48	1.055 ± 0.029	942^{+46}_{-39}	8900^{+600}_{-380}	$1.53^{+0.23}_{-0.13}$	$0.78^{+0.35}_{-0.17}$	11.59	-
GSC 6546-3156	07:24:17.5	-26:16:05	0.701 ± 0.023	1409^{+82}_{-68}	9800^{+900}_{-300}	$1.38^{+0.26}_{-0.11}$	$1.92^{+0.34}_{-0.12}$	14.13	-
GSC 8143-1225	07:59:11.6	-50:22:47	2.588 ± 0.021	$386.3^{+5.3}_{-5.1}$	6750^{+250}_{-100}	$0.47^{+0.13}_{-0.09}$	$0.89^{+0.29}_{-0.20}$	12.35	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
GSC 8581-2002	08:44:23.6	-59:56:58	1.792 ± 0.023	558^{+12}_{-11}	9750^{+250}_{-250}	$1.243^{+0.055}_{-0.038}$	$0.940^{+0.040}_{-0.000}$	11.48	-
GSC 8645-1401	12:17:47.5	-59:43:59	0.538 ± 0.030	1780^{+170}_{-130}	7000^{+250}_{-250}	$2.17^{+0.18}_{-0.18}$	$2.58^{+0.25}_{-0.29}$	13.08	-
GSC 8994-3902	13:19:04.0	-62:34:10	0.346 ± 0.037	2390^{+370}_{-250}	19500^{+5000}_{-3000}	$3.70^{+0.46}_{-0.35}$	$1.51^{+0.29}_{-0.23}$	10.73	-
HBC 217	06:40:42.2	+09:33:37	1.432 ± 0.040	696^{+35}_{-30}	6250^{+150}_{-80}	$0.79^{+0.12}_{-0.06}$	$0.06^{+0.20}_{-0.06}$	12.00	-
HBC 222	06:40:51.2	+09:44:46	1.411 ± 0.035	706^{+31}_{-27}	6250^{+150}_{-80}	$0.82^{+0.12}_{-0.08}$	$0.11^{+0.20}_{-0.11}$	12.01	-
HBC 334	02:16:30.7	+55:23:00	0.535 ± 0.032	1770^{+180}_{-140}	16500^{+3000}_{-800}	$2.18^{+0.35}_{-0.19}$	$2.37^{+0.23}_{-0.15}$	14.32	-
HBC 442	05:34:14.2	-05:36:54	2.592 ± 0.039	386^{+10}_{-9}	6170^{+80}_{-160}	$0.983^{+0.093}_{-0.046}$	$0.07^{+0.19}_{-0.07}$	10.27	-
HBC 7	00:43:25.3	+61:38:23	0.338 ± 0.019	2760^{+250}_{-190}	19500^{+5000}_{-3000}	$3.87^{+0.42}_{-0.32}$	$4.53^{+0.29}_{-0.23}$	13.65	-
HBC 705	20:51:02.7	+43:49:32	0.456 ± 0.027	2070^{+210}_{-160}	19500^{+5000}_{-3000}	$3.71^{+0.42}_{-0.33}$	$5.48^{+0.29}_{-0.23}$	14.37	-
HD 100453	11:33:05.5	-54:19:29	9.597 ± 0.039	$104.20^{+0.70}_{-0.69}$	7250^{+250}_{-250}	$0.790^{+0.024}_{-0.003}$	$0.000^{+0.050}_{-0.000}$	7.78	Yes[7]
HD 100546	11:33:25.3	-70:11:41	9.089 ± 0.051	$110.0^{+1.0}_{-1.0}$	9750^{+500}_{-500}	$1.371^{+0.069}_{-0.046}$	$0.000^{+0.050}_{-0.000}$	6.69	-
HD 101412	11:39:44.4	-60:10:28	2.431 ± 0.028	$411.3^{+8.1}_{-7.6}$	9750^{+250}_{-250}	$1.581^{+0.049}_{-0.036}$	$0.210^{+0.030}_{-0.000}$	9.24	Yes[5]
HD 104237	12:00:04.9	-78:11:35	9.226 ± 0.058	$108.4^{+1.1}_{-1.1}$	8000^{+250}_{-250}	$1.329^{+0.035}_{-0.011}$	$0.000^{+0.050}_{-0.000}$	6.52	Yes[8]
HD 114981	13:14:40.7	-38:39:06	1.405 ± 0.062	705^{+57}_{-44}	16000^{+500}_{-500}	$3.24^{+0.12}_{-0.09}$	$0.000^{+0.050}_{-0.000}$	7.23	-
HD 130437	14:50:50.2	-60:17:10	0.574 ± 0.036	1650^{+180}_{-130}	24500^{+4500}_{-5000}	$4.31^{+0.34}_{-0.40}$	$2.61^{+0.26}_{-0.29}$	10.05	-
HD 132947	15:04:56.0	-63:07:53	2.618 ± 0.057	382^{+15}_{-13}	10250^{+250}_{-250}	$1.606^{+0.073}_{-0.050}$	$0.000^{+0.050}_{-0.000}$	8.91	-
HD 135344	15:15:48.9	-37:08:56	7.036 ± 0.090	$142.1^{+3.1}_{-2.9}$	6750^{+250}_{-250}	$1.119^{+0.015}_{-0.013}$	0.00	7.65	-
HD 135344B	15:15:48.4	-37:09:16	7.365 ± 0.077	$135.8^{+2.4}_{-2.3}$	6380^{+120}_{-120}	$0.786^{+0.033}_{-0.035}$	$0.230^{+0.050}_{-0.060}$	8.63	Yes[9]
HD 139614	15:40:46.4	-42:29:54	7.424 ± 0.053	$134.7^{+1.6}_{-1.6}$	7750^{+250}_{-250}	$0.773^{+0.032}_{-0.010}$	$0.000^{+0.050}_{-0.000}$	8.40	-
HD 141569	15:49:57.7	-03:55:17	9.039 ± 0.044	$110.63^{+0.91}_{-0.88}$	9750^{+250}_{-250}	$1.216^{+0.032}_{-0.027}$	$0.010^{+0.010}_{-0.000}$	7.10	Yes[4]
HD 141926	15:54:21.8	-55:19:44	0.717 ± 0.044	1340^{+150}_{-110}	28000^{+1500}_{-1500}	$4.74^{+0.15}_{-0.14}$	$2.400^{+0.030}_{-0.040}$	8.64	Yes[10]
HD 142527	15:56:41.9	-42:19:24	6.356 ± 0.047	$157.3^{+2.0}_{-1.9}$	6500^{+250}_{-250}	$0.963^{+0.026}_{-0.005}$	$0.000^{+0.050}_{-0.000}$	8.27	Yes[11]

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
HD 142666	15:56:40.0	-22:01:40	6.744 ± 0.053	$148.3^{+2.0}_{-1.9}$	7500^{+250}_{-250}	$0.939^{+0.043}_{-0.045}$	$0.500^{+0.080}_{-0.090}$	8.67	-
HD 143006	15:58:36.9	-22:57:16	6.02 ± 0.15	$166.1^{+7.1}_{-6.2}$	5430^{+38}_{-80}	$0.46^{+0.09}_{-0.12}$	$0.31^{+0.12}_{-0.25}$	10.10	-
HD 144432	16:06:57.9	-27:43:10	6.437 ± 0.058	$155.4^{+2.4}_{-2.2}$	7500^{+250}_{-250}	$0.971^{+0.038}_{-0.011}$	$0.000^{+0.060}_{-0.000}$	8.17	Yes[4]
HD 149914	16:38:28.6	-18:13:14	6.296 ± 0.080	$158.8^{+3.5}_{-3.2}$	10250^{+850}_{-600}	$2.09^{+0.14}_{-0.12}$	$0.95^{+0.14}_{-0.14}$	6.75	-
HD 150193	16:40:17.9	-23:53:45	6.632 ± 0.070	$150.8^{+2.7}_{-2.5}$	9000^{+250}_{-250}	$1.367^{+0.039}_{-0.044}$	$1.550^{+0.020}_{-0.040}$	8.80	Yes[4]
HD 155448	17:12:58.8	-32:14:34	1.024 ± 0.058	950^{+100}_{-70}	10700^{+800}_{-900}	$2.74^{+0.20}_{-0.28}$	$0.47^{+0.12}_{-0.34}$	8.64	Yes[12]
HD 158643	17:31:25.0	-23:57:46	8.15 ± 0.30	$122.8^{+8.2}_{-6.7}$	9800^{+900}_{-300}	$2.22^{+0.26}_{-0.07}$	$0.00^{+0.34}_{-0.00}$	4.81	-
HD 163296	17:56:21.3	-21:57:22	9.85 ± 0.11	$101.5^{+2.0}_{-1.9}$	9250^{+250}_{-250}	$1.199^{+0.055}_{-0.032}$	$0.000^{+0.050}_{-0.000}$	6.85	-
HD 169142	18:24:29.8	-29:46:50	8.775 ± 0.064	$114.0^{+1.4}_{-1.3}$	10700^{+800}_{-900}	$1.31^{+0.12}_{-0.22}$	$1.02^{+0.12}_{-0.34}$	8.16	-
HD 17081	02:44:07.3	-13:51:32	9.38 ± 0.37	$106.7^{+7.9}_{-6.3}$	13000^{+1000}_{-1500}	$2.58^{+0.19}_{-0.17}$	$0.00^{+0.14}_{-0.00}$	4.24	-
HD 174571	18:50:47.2	+08:42:10	0.879 ± 0.059	1100^{+130}_{-90}	19500^{+5000}_{-3000}	$4.23^{+0.43}_{-0.34}$	$2.50^{+0.29}_{-0.23}$	8.70	-
HD 176386	19:01:38.9	-36:53:27	6.281 ± 0.061	$159.2^{+2.6}_{-2.5}$	10700^{+800}_{-900}	$1.58^{+0.12}_{-0.22}$	$0.38^{+0.12}_{-0.34}$	7.56	Yes[13]
HD 179218	19:11:11.3	+15:47:15	3.759 ± 0.047	$266.0^{+5.6}_{-5.2}$	9500^{+200}_{-200}	$2.05^{+0.09}_{-0.14}$	$0.53^{+0.12}_{-0.26}$	7.38	Yes[2]
HD 199603	20:58:41.8	-14:29:00	11.20 ± 0.10	$89.3^{+1.4}_{-1.3}$	7380^{+220}_{-130}	$1.391^{+0.036}_{-0.012}$	$0.000^{+0.059}_{-0.000}$	5.96	Yes[14]
HD 200775	21:01:36.9	+68:09:48	2.771 ± 0.045	361^{+10}_{-9}	16500^{+3000}_{-800}	$3.07^{+0.29}_{-0.14}$	$1.05^{+0.23}_{-0.15}$	7.33	Yes[4]
HD 235495	21:21:27.5	+50:59:48	1.907 ± 0.032	524^{+15}_{-14}	9800^{+900}_{-300}	$1.51^{+0.23}_{-0.08}$	$0.09^{+0.34}_{-0.09}$	9.82	-
HD 244314	05:30:19.0	+11:20:20	2.313 ± 0.059	432^{+19}_{-17}	8500^{+250}_{-250}	$1.153^{+0.058}_{-0.063}$	$0.100^{+0.020}_{-0.050}$	10.10	-
HD 244604	05:31:57.3	+11:17:41	2.374 ± 0.062	421^{+19}_{-17}	9000^{+250}_{-250}	$1.461^{+0.057}_{-0.066}$	$0.140^{+0.000}_{-0.040}$	9.38	-
HD 245185	05:35:09.6	+10:01:51	2.32 ± 0.11	429^{+37}_{-29}	10000^{+500}_{-500}	$1.29^{+0.13}_{-0.10}$	$0.000^{+0.050}_{-0.000}$	9.91	Yes[2]
HD 249879	05:58:55.8	+16:39:57	1.476 ± 0.078	669^{+66}_{-49}	11500^{+1500}_{-800}	$1.56^{+0.29}_{-0.18}$	$0.20^{+0.24}_{-0.12}$	10.69	-
HD 250550	06:02:00.0	+16:30:57	1.401 ± 0.098	697^{+94}_{-64}	11000^{+500}_{-500}	$1.94^{+0.17}_{-0.12}$	$0.000^{+0.050}_{-0.000}$	9.54	-
HD 259431	06:33:05.2	+10:19:20	1.379 ± 0.047	721^{+44}_{-37}	14000^{+2100}_{-2900}	$2.97^{+0.27}_{-0.40}$	$1.11^{+0.21}_{-0.30}$	8.72	Yes[2]

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
HD 287823	05:24:08.0	+02:27:47	2.784 ± 0.053	359^{+12}_{-11}	8380^{+120}_{-120}	$1.116^{+0.053}_{-0.031}$	$0.000^{+0.050}_{-0.000}$	9.68	Yes[6]
HD 288012	05:33:04.8	+02:28:10	2.524 ± 0.056	396^{+15}_{-13}	9800^{+900}_{-300}	$1.66^{+0.24}_{-0.10}$	$0.57^{+0.34}_{-0.12}$	9.32	-
HD 290380	05:23:31.0	-01:04:24	2.821 ± 0.048	354^{+10}_{-9}	6400^{+150}_{-150}	$0.84^{+0.13}_{-0.04}$	$0.06^{+0.27}_{-0.06}$	10.40	-
HD 290409	05:27:05.5	+00:25:08	2.191 ± 0.082	455^{+31}_{-25}	9750^{+500}_{-500}	$1.27^{+0.12}_{-0.09}$	$0.000^{+0.050}_{-0.000}$	10.02	Yes[6]
HD 290500	05:29:48.1	-00:23:43	2.277 ± 0.068	438^{+24}_{-20}	9500^{+500}_{-500}	$0.81^{+0.11}_{-0.08}$	$0.000^{+0.050}_{-0.000}$	11.04	-
HD 290764	05:38:05.3	-01:15:22	2.510 ± 0.063	398^{+18}_{-15}	7880^{+380}_{-380}	$1.178^{+0.093}_{-0.091}$	$0.16^{+0.12}_{-0.14}$	9.88	-
HD 290770	05:37:02.4	-01:37:21	2.502 ± 0.074	399^{+21}_{-18}	10500^{+250}_{-250}	$1.522^{+0.085}_{-0.060}$	$0.000^{+0.050}_{-0.000}$	9.27	-
HD 305298	10:33:05.0	-60:19:51	0.134 ± 0.027	4040^{+630}_{-440}	34000^{+1000}_{-1000}	$4.55^{+0.15}_{-0.14}$	$1.300^{+0.000}_{-0.020}$	10.86	-
HD 313571	18:01:07.2	-22:15:04	0.616 ± 0.077	1400^{+280}_{-170}	16500^{+3000}_{-800}	$3.47^{+0.42}_{-0.23}$	$1.80^{+0.23}_{-0.15}$	10.01	-
HD 31648	04:58:46.3	+29:50:37	6.182 ± 0.076	$161.8^{+3.4}_{-3.2}$	8250^{+200}_{-200}	$1.27^{+0.14}_{-0.05}$	$0.06^{+0.28}_{-0.06}$	7.62	-
HD 319896	17:31:05.9	-35:08:29	0.716 ± 0.063	1300^{+200}_{-130}	15750^{+750}_{-750}	$3.19^{+0.24}_{-0.21}$	$2.38^{+0.15}_{-0.17}$	10.99	-
HD 323771	17:34:04.6	-39:23:41	0.908 ± 0.055	1070^{+120}_{-90}	15000^{+800}_{-1000}	$2.51^{+0.20}_{-0.21}$	$1.32^{+0.17}_{-0.16}$	11.11	-
HD 34282	05:16:00.5	-09:48:35	3.210 ± 0.047	$311.5^{+7.9}_{-7.2}$	9500^{+250}_{-250}	$0.980^{+0.049}_{-0.039}$	$0.010^{+0.020}_{-0.000}$	9.89	Yes[2]
HD 344261	19:21:53.5	+21:31:51	3.328 ± 0.036	$300.5^{+5.6}_{-5.2}$	7000^{+120}_{-250}	$0.72^{+0.11}_{-0.13}$	$0.39^{+0.25}_{-0.29}$	10.65	-
HD 34700	05:19:41.4	+05:38:43	2.805 ± 0.048	356^{+11}_{-10}	5900^{+110}_{-100}	$1.36^{+0.10}_{-0.02}$	$0.00^{+0.21}_{-0.00}$	9.09	Yes[15]
HD 35187	05:24:01.2	+24:57:37	6.13 ± 0.10	$163.0^{+4.6}_{-4.2}$	9800^{+900}_{-300}	$1.45^{+0.30}_{-0.24}$	$0.81^{+0.50}_{-0.50}$	8.17	Yes[4]
HD 35929	05:27:42.8	-08:19:39	2.580 ± 0.052	387^{+13}_{-12}	7000^{+250}_{-250}	$1.789^{+0.046}_{-0.023}$	$0.000^{+0.050}_{-0.000}$	8.12	Yes[2]
HD 36112	05:30:27.5	+25:19:57	6.240 ± 0.067	$160.3^{+2.9}_{-2.8}$	7600^{+220}_{-300}	$1.04^{+0.12}_{-0.08}$	$0.15^{+0.25}_{-0.15}$	8.27	Yes[5]
HD 36408	05:32:14.1	+17:03:29	2.288 ± 0.098	435^{+35}_{-27}	11930^{+830}_{-830}	$3.13^{+0.18}_{-0.22}$	$0.38^{+0.12}_{-0.24}$	6.09	Yes[16]
HD 36917	05:34:47.0	-05:34:15	2.103 ± 0.072	474^{+29}_{-24}	11200^{+1100}_{-1300}	$2.43^{+0.24}_{-0.29}$	$0.52^{+0.24}_{-0.34}$	8.03	-
HD 36982	05:35:09.8	-05:27:53	2.447 ± 0.064	408^{+19}_{-16}	20000^{+1000}_{-1000}	$2.83^{+0.14}_{-0.15}$	$0.82^{+0.12}_{-0.16}$	8.45	-
HD 37258	05:36:59.3	-06:09:16	2.751 ± 0.099	363^{+24}_{-19}	9750^{+500}_{-500}	$1.24^{+0.12}_{-0.10}$	$0.060^{+0.050}_{-0.040}$	9.67	Yes[5]

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
HD 37357	05:37:47.1	-06:42:30	1.27 ± 0.30	650^{+280}_{-120}	9500^{+250}_{-250}	$2.04^{+0.35}_{-0.20}$	$0.000^{+0.050}_{-0.000}$	8.84	Yes[2]
HD 37371	05:38:09.9	-00:11:01	2.428 ± 0.062	411^{+18}_{-16}	10700^{+800}_{-900}	$2.45^{+0.15}_{-0.24}$	$0.83^{+0.12}_{-0.34}$	7.90	-
HD 37490	05:39:11.1	+04:07:17	3.16 ± 0.36	312^{+80}_{-43}	16500^{+3000}_{-800}	$3.69^{+0.46}_{-0.25}$	$0.17^{+0.23}_{-0.15}$	4.59	Yes[17]
HD 37806	05:41:02.3	-02:43:01	2.335 ± 0.058	428^{+19}_{-16}	10500^{+1000}_{-700}	$2.17^{+0.19}_{-0.14}$	$0.13^{+0.19}_{-0.13}$	7.94	Yes[5]
HD 38087	05:43:00.6	-02:18:45	2.95 ± 0.10	338^{+22}_{-18}	13600^{+2900}_{-800}	$2.19^{+0.30}_{-0.22}$	$0.46^{+0.24}_{-0.18}$	8.38	Yes[18]
HD 38120	05:43:11.9	-04:59:50	2.463 ± 0.081	405^{+24}_{-20}	10700^{+800}_{-900}	$1.72^{+0.31}_{-0.20}$	$0.21^{+0.50}_{-0.21}$	9.07	-
HD 39014	05:44:46.3	-65:44:08	22.68 ± 0.30	$44.1^{+1.0}_{-0.9}$	7830^{+160}_{-220}	$1.42^{+0.12}_{-0.02}$	$0.00^{+0.25}_{-0.00}$	4.36	-
HD 41511	06:04:59.1	-16:29:04	4.50 ± 0.17	222^{+16}_{-13}	8900^{+600}_{-380}	$2.83^{+0.24}_{-0.14}$	$0.51^{+0.35}_{-0.17}$	4.93	Yes[19]
HD 45677	06:28:17.4	-13:03:11	1.603 ± 0.058	620^{+41}_{-33}	16500^{+3000}_{-800}	$2.88^{+0.32}_{-0.17}$	$0.57^{+0.23}_{-0.15}$	8.50	Yes[1]
HD 46060	06:30:49.8	-09:39:15	1.049 ± 0.058	930^{+100}_{-70}	21000^{+3400}_{-4200}	$3.89^{+0.33}_{-0.38}$	$1.79^{+0.25}_{-0.22}$	8.64	-
HD 50083	06:51:45.8	+05:05:04	0.897 ± 0.046	1090^{+100}_{-80}	16500^{+3000}_{-800}	$4.04^{+0.34}_{-0.18}$	$0.68^{+0.23}_{-0.15}$	6.91	-
HD 50138	06:51:33.4	-06:57:59	2.630 ± 0.064	380^{+16}_{-14}	9450^{+450}_{-450}	$2.46^{+0.13}_{-0.09}$	$0.03^{+0.12}_{-0.03}$	6.67	Yes[5]
HD 56895B	07:18:31.8	-11:11:34	6.051 ± 0.049	$165.3^{+2.2}_{-2.1}$	7000^{+250}_{-250}	$0.97^{+0.11}_{-0.04}$	$0.08^{+0.25}_{-0.08}$	8.42	-
HD 58647	07:25:56.1	-14:10:44	3.139 ± 0.043	$318.5^{+7.4}_{-6.8}$	10500^{+200}_{-200}	$2.44^{+0.11}_{-0.09}$	$0.37^{+0.19}_{-0.12}$	6.85	Yes[1]
HD 59319	07:28:36.8	-21:57:49	1.489 ± 0.049	668^{+39}_{-33}	12500^{+500}_{-500}	$2.51^{+0.11}_{-0.08}$	$0.000^{+0.050}_{-0.000}$	8.31	-
HD 68695	08:11:44.6	-44:05:09	2.526 ± 0.037	396^{+10}_{-9}	9250^{+250}_{-250}	$1.193^{+0.060}_{-0.036}$	$0.000^{+0.050}_{-0.000}$	9.82	-
HD 76534	08:55:08.7	-43:28:00	1.089 ± 0.037	911^{+55}_{-46}	19000^{+500}_{-500}	$3.549^{+0.087}_{-0.074}$	$0.620^{+0.020}_{-0.010}$	8.07	Yes[4]
HD 85567	09:50:28.5	-60:58:03	0.971 ± 0.028	1023^{+53}_{-45}	13000^{+500}_{-500}	$3.19^{+0.10}_{-0.08}$	$0.890^{+0.030}_{-0.020}$	8.51	Yes[1]
HD 87403	10:02:51.4	-59:16:55	0.465 ± 0.043	1910^{+280}_{-190}	10000^{+250}_{-250}	$2.84^{+0.16}_{-0.11}$	$0.000^{+0.050}_{-0.000}$	9.26	-
HD 87643	10:04:30.3	-58:39:52	0.01 ± 0.12	2010^{+570}_{-350}	19500^{+5000}_{-3000}	$4.60^{+0.64}_{-0.53}$	$1.90^{+0.50}_{-0.50}$	8.51	Yes[30]
HD 94509	10:53:27.2	-58:25:24	0.511 ± 0.034	1830^{+210}_{-150}	11500^{+1000}_{-1000}	$2.99^{+0.19}_{-0.15}$	$0.000^{+0.050}_{-0.000}$	9.12	-
HD 95881	11:01:57.6	-71:30:48	0.844 ± 0.033	1168^{+82}_{-66}	10000^{+250}_{-250}	$2.85^{+0.10}_{-0.07}$	$0.000^{+0.050}_{-0.000}$	8.19	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
HD 96042	11:03:40.5	-59:25:59	0.217 ± 0.034	3100^{+510}_{-350}	25500^{+1500}_{-1500}	$4.81^{+0.20}_{-0.17}$	$0.780^{+0.030}_{-0.010}$	8.47	-
HD 9672	01:34:37.9	-15:40:35	17.52 ± 0.10	$57.07^{+0.55}_{-0.53}$	8900^{+200}_{-200}	$1.174^{+0.093}_{-0.020}$	$0.00^{+0.18}_{-0.00}$	5.61	-
HD 97048	11:08:03.2	-77:39:17	5.411 ± 0.039	$184.8^{+2.2}_{-2.1}$	10500^{+500}_{-500}	$1.544^{+0.072}_{-0.060}$	$0.900^{+0.050}_{-0.020}$	8.44	-
HD 98922	11:22:31.7	-53:22:11	1.448 ± 0.034	689^{+28}_{-25}	10500^{+250}_{-250}	$3.032^{+0.059}_{-0.052}$	$0.090^{+0.010}_{-0.000}$	6.77	Yes[1]
HR 5999	16:08:34.3	-39:06:19	6.207 ± 0.070	$161.1^{+3.1}_{-2.9}$	8500^{+250}_{-250}	$1.717^{+0.048}_{-0.041}$	$0.330^{+0.050}_{-0.040}$	6.78	Yes[3]
HT CMa	07:02:42.5	-11:26:12	0.879 ± 0.036	1121^{+83}_{-66}	10500^{+500}_{-500}	$1.47^{+0.12}_{-0.10}$	$0.230^{+0.050}_{-0.020}$	11.87	-
HU CMa	07:04:06.7	-11:26:08	0.834 ± 0.039	1170^{+100}_{-80}	13000^{+250}_{-250}	$2.06^{+0.10}_{-0.08}$	$0.800^{+0.020}_{-0.010}$	11.55	-
Hen 3-1121	15:58:09.6	-53:51:18	0.310 ± 0.098	1750^{+460}_{-280}	26800^{+5100}_{-4800}	$3.9^{+1.3}_{-1.3}$	$4.1^{+2.3}_{-2.3}$	13.00	-
Hen 3-1121S	15:58:09.7	-53:51:35	0.55 ± 0.17	1150^{+390}_{-210}	29000^{+3900}_{-4500}	$4.12^{+0.44}_{-0.43}$	$3.48^{+0.18}_{-0.26}$	11.02	-
Hen 3-1191	16:27:15.1	-48:39:27	0.487 ± 0.069	1660^{+340}_{-210}	29000^{+3900}_{-4500}	$3.49^{+0.34}_{-0.37}$	$3.84^{+0.18}_{-0.26}$	13.76	-
Hen 3-823	12:48:42.4	-59:54:35	0.805 ± 0.068	1170^{+180}_{-120}	16500^{+3000}_{-800}	$2.93^{+0.39}_{-0.21}$	$1.18^{+0.23}_{-0.15}$	10.35	-
Hen 3-847	13:01:17.8	-48:53:19	1.03 ± 0.23	780^{+290}_{-140}	14000^{+500}_{-500}	$2.07^{+0.32}_{-0.22}$	$0.570^{+0.010}_{-0.030}$	10.61	-
Hen 3-938	13:52:42.8	-63:32:49	0.118 ± 0.032	3850^{+640}_{-450}	32900^{+2000}_{-3900}	$5.03^{+0.25}_{-0.29}$	$5.24^{+0.14}_{-0.18}$	13.40	-
IL Cep	22:53:15.6	+62:08:45	1.238 ± 0.027	805^{+31}_{-27}	16500^{+3000}_{-800}	$3.86^{+0.29}_{-0.15}$	$3.16^{+0.23}_{-0.15}$	9.21	Yes[2]
IP Per	03:40:47.0	+32:31:54	3.250 ± 0.078	308^{+13}_{-11}	7990^{+340}_{-380}	$0.89^{+0.13}_{-0.18}$	$0.52^{+0.22}_{-0.37}$	10.42	-
KK Oph	17:10:08.1	-27:15:19	4.52 ± 0.14	221^{+12}_{-10}	8500^{+500}_{-500}	$0.71^{+0.11}_{-0.11}$	$2.70^{+0.10}_{-0.15}$	12.36	Yes[3]
LKHa 260	18:19:09.4	-13:50:41	0.771 ± 0.056	1230^{+160}_{-110}	14000^{+1000}_{-1000}	$2.09^{+0.24}_{-0.23}$	$3.20^{+0.16}_{-0.18}$	14.16	-
LKHa 338	06:10:47.1	-06:12:51	1.119 ± 0.044	885^{+63}_{-50}	10700^{+800}_{-900}	$1.13^{+0.17}_{-0.26}$	$2.60^{+0.12}_{-0.34}$	14.63	-
LkHa 208	06:07:49.5	+18:39:26	1.49 ± 0.20	630^{+170}_{-90}	7830^{+500}_{-580}	$1.04^{+0.35}_{-0.28}$	$0.36^{+0.34}_{-0.36}$	11.44	Yes[3]
LkHa 215	06:32:41.8	+10:09:34	1.395 ± 0.045	713^{+41}_{-34}	14000^{+1000}_{-1000}	$2.57^{+0.18}_{-0.19}$	$2.02^{+0.16}_{-0.18}$	10.60	Yes[2]
LkHa 257	21:54:18.8	+47:12:10	1.258 ± 0.016	794^{+18}_{-16}	15000^{+800}_{-1000}	$1.91^{+0.13}_{-0.15}$	$2.65^{+0.17}_{-0.16}$	13.29	-
LkHa 259	23:58:41.6	+66:26:13	1.319 ± 0.033	755^{+34}_{-29}	7380^{+220}_{-130}	$1.16^{+0.09}_{-0.13}$	$4.05^{+0.14}_{-0.24}$	15.20	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
LkHa 324	21:03:54.2	+50:15:10	1.651 ± 0.025	605^{+16}_{-14}	11500^{+1500}_{-800}	$2.08^{+0.23}_{-0.13}$	$3.64^{+0.24}_{-0.12}$	12.63	-
LkHa 339	06:10:57.8	-06:14:40	1.163 ± 0.026	857^{+33}_{-29}	10500^{+250}_{-250}	$1.922^{+0.058}_{-0.055}$	$3.540^{+0.010}_{-0.010}$	13.47	-
MQ Cas	00:09:37.6	+58:13:11	1.217 ± 0.089	800^{+110}_{-80}	9800^{+900}_{-300}	$1.10^{+0.32}_{-0.16}$	$0.64^{+0.34}_{-0.12}$	12.31	-
MWC 1021	20:29:26.9	+41:40:44	0.507 ± 0.066	1640^{+320}_{-200}	19800^{+9200}_{-9200}	$5.49^{+0.77}_{-0.87}$	$8.89^{+0.60}_{-0.34}$	12.85	-
MWC 1080	23:17:25.6	+60:50:43	0.729 ± 0.092	1200^{+260}_{-150}	29000^{+3900}_{-4500}	$4.52^{+0.37}_{-0.37}$	$5.03^{+0.18}_{-0.26}$	11.61	Yes[1]
MWC 137	06:18:45.5	+15:16:52	0.110 ± 0.055	2910^{+600}_{-400}	29000^{+3900}_{-4500}	$4.94^{+0.37}_{-0.38}$	$4.63^{+0.18}_{-0.26}$	12.10	-
MWC 297	18:27:39.5	-03:49:52	2.663 ± 0.085	375^{+22}_{-18}	24500^{+1500}_{-1500}	$4.59^{+0.12}_{-0.12}$	$8.470^{+0.040}_{-0.030}$	12.03	Yes[2]
MWC 342	20:23:03.6	+39:29:50	0.536 ± 0.024	1810^{+140}_{-110}	26000^{+5900}_{-4000}	$4.89^{+0.35}_{-0.34}$	$4.25^{+0.23}_{-0.33}$	10.56	-
MWC 593	17:49:10.2	-24:14:21	0.682 ± 0.064	1340^{+220}_{-140}	15750^{+750}_{-750}	$3.58^{+0.25}_{-0.21}$	$2.31^{+0.15}_{-0.17}$	10.03	-
MWC 655	22:38:31.8	+55:50:05	0.424 ± 0.029	2170^{+250}_{-180}	24500^{+4500}_{-5000}	$4.12^{+0.36}_{-0.40}$	$0.94^{+0.26}_{-0.29}$	9.39	-
MWC 657	22:42:41.8	+60:24:01	0.265 ± 0.023	3160^{+400}_{-290}	19800^{+9200}_{-9200}	$4.62^{+0.72}_{-0.83}$	$5.03^{+0.60}_{-0.34}$	12.60	Yes[32]
MWC 878	17:24:44.7	-38:43:51	0.486 ± 0.053	1770^{+300}_{-200}	24500^{+4500}_{-5000}	$4.32^{+0.38}_{-0.44}$	$3.06^{+0.26}_{-0.29}$	10.62	-
MWC 953	18:43:28.4	-03:46:17	0.493 ± 0.058	1720^{+310}_{-200}	19500^{+5000}_{-3000}	$4.21^{+0.48}_{-0.37}$	$3.62^{+0.29}_{-0.23}$	10.84	-
NSV 2968	06:26:53.9	-10:15:35	0.942 ± 0.062	1030^{+120}_{-90}	29000^{+3900}_{-4500}	$3.62^{+0.28}_{-0.33}$	$5.88^{+0.18}_{-0.26}$	14.43	-
NV Ori	05:35:31.4	-05:33:09	2.585 ± 0.057	387^{+15}_{-13}	6750^{+250}_{-200}	$1.19^{+0.15}_{-0.08}$	$0.13^{+0.29}_{-0.13}$	9.78	-
PDS 002	01:17:43.5	-52:33:31	2.411 ± 0.027	$414.6^{+7.8}_{-7.3}$	6750^{+250}_{-100}	$0.79^{+0.12}_{-0.01}$	$0.00^{+0.26}_{-0.00}$	10.81	-
PDS 004	03:39:00.6	+29:41:46	2.505 ± 0.050	399^{+14}_{-12}	9800^{+900}_{-300}	$1.28^{+0.24}_{-0.10}$	$0.91^{+0.34}_{-0.12}$	10.63	-
PDS 021	06:02:14.9	-10:00:59	1.119 ± 0.037	887^{+53}_{-44}	15000^{+1500}_{-1500}	$2.58^{+0.18}_{-0.16}$	$1.170^{+0.060}_{-0.040}$	10.38	-
PDS 022	06:03:37.1	-14:53:03	1.11 ± 0.10	850^{+140}_{-90}	9800^{+900}_{-300}	$1.82^{+0.34}_{-0.17}$	$0.20^{+0.34}_{-0.12}$	10.23	-
PDS 025	06:54:27.9	-25:02:16	1.308 ± 0.019	763^{+19}_{-18}	8150^{+180}_{-160}	$0.60^{+0.10}_{-0.09}$	$1.44^{+0.19}_{-0.16}$	14.03	-
PDS 123	05:50:54.8	+20:14:48	0.649 ± 0.035	1490^{+140}_{-110}	19800^{+9200}_{-9200}	$2.80^{+0.69}_{-0.82}$	$2.64^{+0.60}_{-0.34}$	13.11	-
PDS 124	06:06:58.5	-05:55:07	1.160 ± 0.048	853^{+64}_{-51}	10250^{+250}_{-250}	$1.38^{+0.10}_{-0.07}$	$1.230^{+0.030}_{-0.000}$	12.44	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
PDS 126	06:13:37.3	-06:25:02	1.173 ± 0.040	847^{+51}_{-42}	7830^{+160}_{-220}	$1.37^{+0.15}_{-0.14}$	$0.90^{+0.25}_{-0.25}$	11.78	-
PDS 129	06:31:03.6	+10:01:13	1.447 ± 0.069	684^{+60}_{-46}	6550^{+100}_{-150}	$0.96^{+0.15}_{-0.16}$	$0.67^{+0.20}_{-0.27}$	12.15	-
PDS 130	06:49:58.6	-07:38:52	0.754 ± 0.020	1316^{+62}_{-53}	10500^{+250}_{-250}	$1.734^{+0.064}_{-0.057}$	$2.070^{+0.010}_{-0.000}$	13.40	-
PDS 133	07:25:05.0	-25:45:50	0.668 ± 0.024	1475^{+92}_{-76}	14000^{+2000}_{-2000}	$1.95^{+0.23}_{-0.24}$	$1.43^{+0.09}_{-0.10}$	13.13	-
PDS 134	07:32:26.6	-21:55:36	0.328 ± 0.033	2550^{+370}_{-260}	14000^{+500}_{-500}	$2.72^{+0.17}_{-0.13}$	$1.220^{+0.030}_{-0.020}$	12.20	-
PDS 138	11:53:13.2	-62:05:21	0.043 ± 0.029	4630^{+750}_{-540}	29000^{+3900}_{-4500}	$5.14^{+0.31}_{-0.36}$	$5.29^{+0.18}_{-0.26}$	13.31	-
PDS 174	05:06:55.5	-03:21:13	2.514 ± 0.035	398^{+10}_{-9}	17000^{+2000}_{-2000}	$1.96^{+0.16}_{-0.18}$	$3.510^{+0.070}_{-0.070}$	12.84	-
PDS 211	06:10:17.3	+29:25:17	0.922 ± 0.031	1074^{+65}_{-54}	10700^{+800}_{-900}	$1.79^{+0.16}_{-0.25}$	$2.99^{+0.12}_{-0.34}$	13.79	-
PDS 24	06:48:41.7	-16:48:06	0.881 ± 0.019	1130^{+42}_{-37}	10500^{+500}_{-500}	$1.274^{+0.088}_{-0.082}$	$1.110^{+0.040}_{-0.030}$	13.26	-
PDS 241	07:08:38.8	-04:19:05	0.118 ± 0.054	2890^{+600}_{-400}	26000^{+1500}_{-1500}	$4.05^{+0.27}_{-0.19}$	$2.600^{+0.040}_{-0.010}$	12.06	-
PDS 27	07:19:35.9	-17:39:18	0.275 ± 0.044	2550^{+460}_{-310}	17500^{+3500}_{-3500}	$4.15^{+0.37}_{-0.39}$	$5.03^{+0.13}_{-0.13}$	13.00	-
PDS 277	08:23:11.8	-39:07:01	2.891 ± 0.036	$345.8^{+7.3}_{-6.8}$	6750^{+250}_{-100}	$0.948^{+0.074}_{-0.016}$	$0.00^{+0.15}_{-0.00}$	10.02	-
PDS 286	09:06:00.0	-47:18:58	0.515 ± 0.034	1820^{+210}_{-150}	30000^{+3000}_{-3000}	$5.21^{+0.20}_{-0.18}$	$6.270^{+0.050}_{-0.040}$	12.15	-
PDS 290	09:26:11.1	-52:42:27	1.146 ± 0.027	869^{+36}_{-31}	13000^{+1000}_{-1500}	$1.60^{+0.18}_{-0.24}$	$3.31^{+0.18}_{-0.24}$	14.55	-
PDS 297	09:42:40.3	-56:15:34	0.610 ± 0.031	1590^{+140}_{-110}	10750^{+250}_{-250}	$2.12^{+0.14}_{-0.16}$	$1.21^{+0.12}_{-0.19}$	12.02	-
PDS 324	10:57:24.2	-62:53:13	0.307 ± 0.023	2880^{+330}_{-240}	24500^{+4500}_{-5000}	$3.43^{+0.35}_{-0.42}$	$3.57^{+0.26}_{-0.29}$	14.45	-
PDS 33	08:48:45.7	-40:48:21	1.046 ± 0.027	951^{+43}_{-37}	9750^{+250}_{-250}	$1.193^{+0.075}_{-0.054}$	$0.520^{+0.040}_{-0.000}$	12.34	-
PDS 34	08:49:58.5	-45:53:06	0.457 ± 0.017	2130^{+140}_{-110}	19500^{+5000}_{-3000}	$2.87^{+0.39}_{-0.31}$	$2.96^{+0.29}_{-0.23}$	14.04	-
PDS 344	11:40:32.8	-64:32:06	0.395 ± 0.016	2440^{+170}_{-140}	15250^{+500}_{-500}	$2.24^{+0.09}_{-0.10}$	$0.860^{+0.010}_{-0.020}$	13.15	-
PDS 361S	13:03:21.5	-62:13:26	0.266 ± 0.030	2950^{+430}_{-300}	18500^{+1000}_{-1000}	$2.87^{+0.37}_{-0.35}$	$1.20^{+0.50}_{-0.50}$	12.85	Yes[27]
PDS 37	10:10:00.3	-57:02:07	0.416 ± 0.056	1930^{+360}_{-230}	17500^{+3500}_{-3500}	$4.00^{+0.37}_{-0.39}$	$5.81^{+0.13}_{-0.13}$	13.54	-
PDS 389	15:14:47.0	-62:17:00	1.238 ± 0.033	804^{+38}_{-32}	8520^{+380}_{-180}	$1.95^{+0.13}_{-0.15}$	$4.84^{+0.17}_{-0.28}$	14.18	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
PDS 415N	16:18:37.2	-24:05:18	6.93 ± 0.13	$144.2^{+4.7}_{-4.2}$	6250^{+250}_{-250}	$0.44^{+0.16}_{-0.12}$	$1.48^{+0.35}_{-0.25}$	10.88	-
PDS 431	16:54:59.2	-43:21:50	0.531 ± 0.027	1810^{+160}_{-120}	10500^{+500}_{-500}	$1.88^{+0.12}_{-0.11}$	$1.760^{+0.030}_{-0.030}$	13.42	-
PDS 469	17:50:58.1	-14:16:12	0.864 ± 0.083	1080^{+190}_{-120}	9800^{+900}_{-300}	$1.60^{+0.35}_{-0.17}$	$1.73^{+0.34}_{-0.12}$	12.80	-
PDS 477	18:00:30.3	-16:47:26	-0.19 ± 0.10	2470^{+640}_{-410}	24500^{+4500}_{-5000}	$3.61^{+0.46}_{-0.50}$	$4.28^{+0.26}_{-0.29}$	14.38	-
PDS 520	18:30:06.2	+00:42:33	2.60 ± 0.19	380^{+56}_{-37}	6750^{+250}_{-100}	$0.59^{+0.23}_{-0.16}$	$4.07^{+0.29}_{-0.20}$	15.17	-
PDS 543	18:48:00.7	+02:54:17	0.639 ± 0.063	1410^{+240}_{-160}	29000^{+3900}_{-4500}	$5.21^{+0.32}_{-0.36}$	$7.12^{+0.18}_{-0.26}$	12.40	-
PDS 69	13:57:43.9	-39:58:47	1.551 ± 0.044	643^{+33}_{-28}	15000^{+2000}_{-2000}	$2.70^{+0.20}_{-0.21}$	$1.600^{+0.070}_{-0.070}$	9.80	Yes[23]
PX Vul	19:26:40.3	+23:53:51	1.581 ± 0.068	627^{+50}_{-39}	6750^{+310}_{-150}	$1.36^{+0.18}_{-0.13}$	$1.21^{+0.29}_{-0.20}$	11.51	-
RR Tau	05:39:30.5	+26:22:27	1.284 ± 0.046	773^{+49}_{-41}	10000^{+200}_{-200}	$2.01^{+0.19}_{-0.12}$	$1.55^{+0.28}_{-0.12}$	10.92	Yes[2]
RY Ori	05:32:09.9	-02:49:47	2.713 ± 0.039	$368.5^{+9.0}_{-8.3}$	6250^{+150}_{-80}	$0.86^{+0.10}_{-0.09}$	$0.96^{+0.20}_{-0.19}$	11.36	-
SAO 185668	17:43:55.6	-22:05:45	0.604 ± 0.061	1480^{+250}_{-160}	16500^{+3000}_{-800}	$3.80^{+0.40}_{-0.22}$	$2.01^{+0.23}_{-0.15}$	9.53	-
SAO 220669	08:55:45.9	-44:25:14	1.066 ± 0.031	932^{+47}_{-40}	16000^{+1500}_{-1500}	$3.58^{+0.16}_{-0.18}$	$1.89^{+0.07}_{-0.10}$	8.87	-
SV Cep	22:21:33.2	+73:40:27	2.904 ± 0.020	$344.3^{+4.0}_{-3.8}$	10250^{+200}_{-200}	$1.073^{+0.080}_{-0.080}$	$0.96^{+0.12}_{-0.12}$	10.97	Yes[2]
T Ori	05:35:50.5	-05:28:35	2.452 ± 0.044	408^{+13}_{-11}	9000^{+500}_{-500}	$1.56^{+0.10}_{-0.07}$	$1.500^{+0.080}_{-0.050}$	10.43	Yes[2]
TY CrA	19:01:40.8	-36:52:34	7.33 ± 0.15	$136.5^{+4.7}_{-4.2}$	10700^{+800}_{-900}	$1.41^{+0.14}_{-0.23}$	$1.98^{+0.12}_{-0.34}$	9.26	Yes[21]
UX Ori	05:04:30.0	-03:47:14	3.077 ± 0.051	$324.9^{+9.3}_{-8.4}$	8500^{+250}_{-250}	$0.962^{+0.064}_{-0.044}$	$0.480^{+0.070}_{-0.030}$	10.34	Yes[2]
V1012 Ori	05:11:36.5	-02:22:48	2.586 ± 0.051	386^{+13}_{-12}	8500^{+250}_{-250}	$0.773^{+0.048}_{-0.050}$	$1.320^{+0.020}_{-0.040}$	12.04	-
V1295 Aql	20:03:02.5	+05:44:17	1.122 ± 0.067	870^{+100}_{-70}	9500^{+200}_{-200}	$2.90^{+0.16}_{-0.20}$	$0.40^{+0.12}_{-0.26}$	7.73	-
V1478 Cyg	20:32:45.5	+40:39:37	0.62 ± 0.10	1300^{+320}_{-190}	19800^{+9200}_{-9200}	$5.17^{+0.81}_{-0.89}$	$8.91^{+0.60}_{-0.34}$	13.15	Yes[33]
V1493 Cyg	20:52:04.6	+44:37:30	0.828 ± 0.088	1100^{+210}_{-130}	8900^{+600}_{-380}	$2.15^{+0.34}_{-0.20}$	$5.75^{+0.35}_{-0.17}$	15.34	-
V1685 Cyg	20:20:28.2	+41:21:51	1.092 ± 0.031	910^{+46}_{-39}	19500^{+5000}_{-3000}	$3.80^{+0.42}_{-0.29}$	$3.75^{+0.39}_{-0.20}$	10.63	Yes[4]
V1686 Cyg	20:20:29.3	+41:21:28	0.78 ± 0.12	1080^{+280}_{-160}	6010^{+160}_{-110}	$1.53^{+0.27}_{-0.22}$	$1.85^{+0.19}_{-0.24}$	12.93	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
V1787 Ori	05:38:09.3	-06:49:17	2.555 ± 0.053	391^{+14}_{-13}	8150^{+180}_{-160}	$1.15^{+0.11}_{-0.09}$	$4.08^{+0.19}_{-0.16}$	13.84	-
V1818 Ori	05:53:42.6	-10:24:01	1.412 ± 0.088	695^{+82}_{-59}	13000^{+1000}_{-1500}	$2.96^{+0.24}_{-0.29}$	$3.72^{+0.18}_{-0.24}$	11.07	-
V1977 Cyg	20:47:37.5	+43:47:25	1.158 ± 0.030	860^{+39}_{-34}	11000^{+200}_{-200}	$2.48^{+0.11}_{-0.10}$	$2.26^{+0.12}_{-0.12}$	10.90	Yes[21]
V2019 Cyg	20:48:04.8	+43:47:26	1.182 ± 0.029	843^{+36}_{-31}	10700^{+800}_{-900}	$2.33^{+0.15}_{-0.24}$	$2.50^{+0.12}_{-0.34}$	11.42	-
V346 Ori	05:24:42.8	+01:43:48	2.728 ± 0.043	366^{+10}_{-9}	7750^{+250}_{-250}	$0.918^{+0.045}_{-0.022}$	$0.000^{+0.050}_{-0.000}$	10.21	-
V350 Ori	05:40:11.8	-09:42:11	2.55 ± 0.12	390^{+35}_{-27}	9000^{+250}_{-250}	$1.04^{+0.10}_{-0.09}$	$0.690^{+0.020}_{-0.030}$	10.82	Yes[2]
V351 Ori	05:44:18.8	+00:08:40	2.925 ± 0.045	$341.8^{+9.1}_{-8.3}$	7830^{+160}_{-220}	$1.40^{+0.12}_{-0.03}$	$0.03^{+0.25}_{-0.03}$	8.90	-
V361 Cep	21:42:50.2	+66:06:35	1.116 ± 0.025	893^{+35}_{-31}	15750^{+750}_{-750}	$3.06^{+0.15}_{-0.14}$	$1.90^{+0.15}_{-0.17}$	10.03	-
V373 Cep	21:43:06.8	+66:06:54	1.081 ± 0.022	922^{+33}_{-29}	13000^{+1000}_{-1500}	$2.29^{+0.18}_{-0.24}$	$3.07^{+0.18}_{-0.24}$	12.73	Yes[3]
V374 Cep	23:05:07.5	+62:15:36	1.141 ± 0.030	872^{+40}_{-35}	15000^{+800}_{-1000}	$3.36^{+0.15}_{-0.17}$	$3.00^{+0.17}_{-0.16}$	10.21	Yes[2]
V380 Ori	05:36:25.4	-06:42:58	2.04 ± 0.16	482^{+76}_{-49}	9750^{+750}_{-750}	$2.00^{+0.21}_{-0.18}$	$2.210^{+0.050}_{-0.070}$	10.53	Yes[1]
V388 Vel	08:42:17.3	-40:44:10	0.196 ± 0.061	2470^{+540}_{-340}	9500^{+300}_{-600}	$2.45^{+0.24}_{-0.32}$	$4.00^{+0.12}_{-0.35}$	14.69	-
V431 Sct	18:29:25.7	-06:04:37	0.634 ± 0.068	1400^{+250}_{-160}	22500^{+4500}_{-4500}	$3.86^{+0.46}_{-0.38}$	$3.78^{+0.33}_{-0.16}$	11.75	Yes[21]
V594 Cas	00:43:18.3	+61:54:40	1.755 ± 0.028	569^{+16}_{-14}	11500^{+1500}_{-800}	$2.13^{+0.24}_{-0.13}$	$1.90^{+0.24}_{-0.12}$	10.61	-
V594 Cyg	21:20:23.4	+43:18:10	0.670 ± 0.091	1280^{+280}_{-170}	11500^{+1500}_{-800}	$3.51^{+0.39}_{-0.23}$	$5.97^{+0.24}_{-0.12}$	12.99	-
V599 Ori	05:38:58.6	-07:16:46	2.436 ± 0.042	410^{+12}_{-11}	8000^{+250}_{-250}	$1.444^{+0.057}_{-0.057}$	$4.650^{+0.060}_{-0.070}$	13.76	-
V669 Cep	22:26:38.7	+61:13:32	0.990 ± 0.066	980^{+120}_{-90}	15000^{+800}_{-1000}	$2.60^{+0.21}_{-0.21}$	$3.05^{+0.17}_{-0.16}$	12.41	Yes[24]
V718 Sco	16:13:11.6	-22:29:07	6.556 ± 0.081	$152.5^{+3.2}_{-3.0}$	8000^{+250}_{-250}	$0.899^{+0.046}_{-0.039}$	$0.740^{+0.060}_{-0.050}$	9.10	Yes[25]
V921 Sco	16:59:06.8	-42:42:08	0.587 ± 0.054	1550^{+240}_{-160}	29000^{+3900}_{-4500}	$4.76^{+0.33}_{-0.34}$	$4.88^{+0.18}_{-0.26}$	11.41	Yes[26]
VV Ser	18:28:47.9	+00:08:40	2.381 ± 0.046	420^{+14}_{-13}	13800^{+200}_{-200}	$1.95^{+0.10}_{-0.08}$	$2.91^{+0.12}_{-0.12}$	11.82	-
VX Cas	00:31:30.7	+61:58:51	1.861 ± 0.037	537^{+18}_{-16}	10000^{+200}_{-200}	$1.21^{+0.16}_{-0.10}$	$0.81^{+0.28}_{-0.12}$	11.39	Yes[2]
WRAY 15-1435	16:13:06.7	-50:23:20	0.467 ± 0.048	1860^{+300}_{-200}	24500^{+4500}_{-5000}	$3.68^{+0.39}_{-0.44}$	$3.56^{+0.26}_{-0.29}$	12.86	-

Table A.1 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
WW Vul	19:25:58.8	+21:12:31	1.984 ± 0.037	504^{+16}_{-15}	8970^{+200}_{-200}	$1.42^{+0.13}_{-0.09}$	$0.95^{+0.22}_{-0.12}$	10.64	-
XY Per A	03:49:36.3	+38:58:55	2.169 ± 0.089	459^{+35}_{-27}	9750^{+200}_{-200}	$2.00^{+0.13}_{-0.12}$	$1.47^{+0.12}_{-0.12}$	9.67	Yes[1]

Notes: Atmospheric parameters T_{eff}, A_V and V taken from the following sources in order of choice: Fairlamb et al., 2015, Montesinos et al., 2009, Hernández et al., 2004, Mendigutía et al., 2012, Carmona et al., 2010, Chen et al., 2016, Alecian et al., 2013, Sartori et al., 2010, Manoj et al., 2006, Hernández et al., 2005, Vieira et al., 2003, APASS Data Release 9 (Henden et al., 2018) and the SIMBAD database. If not available they were derived as described in Sect. 2.1.2. See Sect. 2.2 for derivation of *L*. The references for binarity are: (1) Baines et al. (2006), (2) Wheelwright et al. (2010), (3) Leinert et al. (1997), (4) Maheswar et al. (2002), (5) Wheelwright et al. (2011), (6) Alecian et al. (2013), (7) Hamaguchi et al. (2008), (8) Dunhill et al. (2015), (9) Coulson and Walther (1995), (10) Liu et al. (2000), (11) Biller et al. (2012), (12) Schütz et al. (2011), (13) Boersma et al. (2009), (14) Malkov et al. (2006), (15) Arellano Ferro and Giridhar (2003), (16) Kubát et al. (2010), (17) Morrell and Levato (1991), (18) Lazareff et al. (2017), (19) Mayer et al. (2016), (20) Folsom et al. (2008), (21) Corporon and Lagrange (1999), (22) Doering and Meixner (2009), (23) Chelli et al. (1995), (24) Miroshnichenko et al. (2002), (25) Friedemann et al. (1996), (26) Kraus et al. (2012), (27) Torres et al. (2000), (28) Aspin (1998), (29) Connelley et al. (2008), (30) Millour et al. (2009), (31) Frasca et al. (2016), (32) Marston and McCollum (2008), (33) Zhang et al. (2017).

Table A.2 Other parameters of each known Herbig Ae/Be star belonging to the high-quality sample of 218 sources.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
AB Aur	$0.32^{+0.25}_{-0.13}$	$0.27^{+0.25}_{-0.12}$	-55.7 ± 4.5[6]	P[6]	0.21	$2.15^{+0.36}_{-0.21}$	$4.0^{+1.4}_{-1.5}$
AK Sco	$0.212^{+0.035}_{-0.053}$	$0.196^{+0.016}_{-0.024}$	-5.9 ± 1.1[1]	d[28]	12.47	$1.401^{+0.070}_{-0.070}$	$8.4^{+1.7}_{-0.4}$
AS 310	-	-	-8.89 ± 0.38[5]	d[26]	-0.21	$11.9^{+4.8}_{-3.4}$	$0.06^{+0.54}_{-0.04}$
AS 470	$0.031^{+0.054}_{-0.031}$	$(0.95^{+0.73}_{-0.82})\cdot10^{-2}$	-49.2 ± 4.0[23]	-	1.22	$7.0^{+2.8}_{-1.8}$	$0.10^{+0.20}_{-0.07}$
AS 477	$0.19^{+0.07}_{-0.11}$	$0.162^{+0.056}_{-0.092}$	-38.8 ± 1.5[22]	P[22]	0.24	$3.3^{+1.1}_{-0.4}$	$1.25^{+0.64}_{-0.73}$
BD+30 549	$0.030^{+0.013}_{-0.015}$	$(1.63^{+0.52}_{-0.67})\cdot10^{-2}$	-	-	0.22	$2.28^{+0.37}_{-0.19}$	5^{+15}_{-2}
BD+41 3731	-	$(0.24^{+0.61}_{-0.24})\cdot10^{-3}$	-	s[8]	0.01	$5.0^{+1.2}_{-0.7}$	$0.5^{+3.0}_{-0.3}$
BF Ori	$0.113^{+0.011}_{-0.012}$	$0.139^{+0.012}_{-0.013}$	-14.7 ± 1.5[1]	d[3]	9.25∗	$1.807^{+0.090}_{-0.090}$	$6.38^{+0.32}_{-0.46}$
BH Cep	$0.17^{+0.14}_{-0.08}$	$0.170^{+0.077}_{-0.042}$	-6.20 ± 0.19[3]	d[3]	11.09∗	$1.37^{+0.15}_{-0.10}$	$10.6^{+3.0}_{-3.1}$
BO Cep	$0.121^{+0.043}_{-0.055}$	$0.072^{+0.012}_{-0.015}$	-7.50 ± 0.22[3]	d[3]	0.39∗	$1.215^{+0.061}_{-0.061}$	$17.1^{+0.9}_{-2.4}$
CO Ori	$0.31^{+0.12}_{-0.10}$	$0.173^{+0.049}_{-0.041}$	-21.10 ± 0.63[3]	s[3]	-*	$2.60^{+0.20}_{-0.24}$	$1.76^{+0.62}_{-0.36}$
CPM 25	$0.059^{+0.053}_{-0.033}$	$0.15^{+0.13}_{-0.09}$	-200 ± 10[27]	d[21]	-	$5.2^{+2.2}_{-1.2}$	$0.7^{+5.7}_{-0.4}$
CQ Tau	$0.20^{+0.14}_{-0.13}$	$0.37^{+0.14}_{-0.14}$	-4.80 ± 0.14[3]	d[3]	27.33∗	$1.47^{+0.19}_{-0.11}$	$8.9^{+2.8}_{-2.5}$
DG Cir	$0.56^{+0.92}_{-0.28}$	$0.7^{+1.1}_{-0.3}$	-61.1 ± 1.6[1]	P[21]	18.42	$2.30^{+0.60}_{-0.65}$	4^{+16}_{-3}
GSC 1876-0892	$0.022^{+0.021}_{-0.013}$	$0.049^{+0.043}_{-0.027}$	-5.21 ± 0.25[27]	d[21]	0.98	$9.4^{+5.2}_{-2.4}$	$0.09^{+0.13}_{-0.06}$
GSC 3975-0579	$0.190^{+0.056}_{-0.079}$	$0.173^{+0.044}_{-0.063}$	-16.09 ± 0.34[11]	-	0.04	$2.06^{+0.37}_{-0.12}$	$4.4^{+0.7}_{-1.6}$
GSC 6546-3156	$0.050^{+0.016}_{-0.028}$	$0.27^{+0.05}_{-0.11}$	-18.78 ± 0.45[15]	s[21]	-	$2.04^{+0.29}_{-0.14}$	6^{+14}_{-2}
GSC 8143-1225	$0.28^{+0.11}_{-0.11}$	$0.157^{+0.038}_{-0.043}$	-	-	49.58	$1.275^{+0.064}_{-0.071}$	$17.7^{+2.3}_{-3.0}$
GSC 8581-2002	$(4.45^{+0.27}_{-0.57})\cdot10^{-2}$	$(6.64^{+0.41}_{-0.72})\cdot10^{-3}$	-2.6 ± 1.4[1]	-	-0.12	$1.880^{+0.094}_{-0.094}$	8^{+12}_{-1}
GSC 8645-1401	$0.20^{+0.12}_{-0.08}$	$0.197^{+0.071}_{-0.047}$	-14.42 ± 0.45[15]	P[21]	9.76	$4.04^{+0.66}_{-0.56}$	$0.56^{+0.33}_{-0.23}$
GSC 8994-3902	-	-	-	-	0.04	$8.4^{+3.9}_{-2.0}$	$0.12^{+0.44}_{-0.08}$

Table A.2 continued.

Name	Near IR excess (1.24 − 3.4 µm)	Mid IR excess (3.4 − 22 µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
HBC 217	$0.146^{+0.036}_{-0.068}$	$0.162^{+0.016}_{-0.035}$	-14.55 ± 0.52[5]	-	0.63	$1.51^{+0.16}_{-0.08}$	$7.3^{+0.7}_{-1.8}$
HBC 222	$0.165^{+0.055}_{-0.071}$	$0.120^{+0.019}_{-0.026}$	-5.451 ± 0.060[5]	-	1.45	$1.53^{+0.14}_{-0.08}$	$7.2^{+0.9}_{-1.7}$
HBC 334	$0.020^{+0.009}_{-0.010}$	$0.078^{+0.027}_{-0.036}$	-5.03 ± 0.01[5]	-	0.18	$3.71^{+0.49}_{-0.19}$	$2.1^{+4.3}_{-1.1}$
HBC 442	$0.100^{+0.040}_{-0.062}$	$0.168^{+0.019}_{-0.033}$	-4.773 ± 0.065[5]	P[5]	2.93	$1.80^{+0.25}_{-0.09}$	$4.5^{+0.2}_{-1.5}$
HBC 7	$(0.20^{+0.31}_{-0.20}) \cdot 10^{-2}$	$(0.51^{+0.50}_{-0.30}) \cdot 10^{-3}$	-41.8 ± 2.1[5]	-	-0.06	$9.5^{+4.1}_{-2.1}$	$0.09^{+0.11}_{-0.06}$
HBC 705	$(0.36^{+0.46}_{-0.27}) \cdot 10^{-2}$	$(0.65^{+0.62}_{-0.38}) \cdot 10^{-3}$	-27.1 ± 1.3[5]	-	1.15	$8.5^{+3.6}_{-1.8}$	$0.12^{+0.21}_{-0.08}$
HD 100453	$0.177^{+0.031}_{-0.042}$	$0.309^{+0.037}_{-0.045}$	-5.0 ± 1.3[1]	s[31]	-0.17	$1.251^{+0.063}_{-0.063}$	$6.53^{+0.45}_{-0.49}$
HD 100546	$0.072^{+0.011}_{-0.015}$	$0.29^{+0.10}_{-0.08}$	-39.5 ± 1.6[1]	s[6]	$-0.06*$	$2.05^{+0.10}_{-0.12}$	$5.5^{+1.4}_{-0.8}$
HD 101412	$0.103^{+0.009}_{-0.012}$	$0.184^{+0.015}_{-0.019}$	-15.4 ± 1.7[1]	d[21]	-0.04	$2.10^{+0.11}_{-0.11}$	$4.37^{+0.22}_{-0.32}$
HD 104237	$0.154^{+0.022}_{-0.032}$	$0.151^{+0.032}_{-0.032}$	-27.5 ± 1.7[1]	P[4]	0.50	$1.849^{+0.092}_{-0.092}$	$5.48^{+0.27}_{-0.40}$
HD 114981	$(0.29^{+0.04}_{-0.10}) \cdot 10^{-2}$	$(3.22^{+0.30}_{-0.69}) \cdot 10^{-4}$	-13.24 ± 0.72[1]	d[21]	0.35	$6.09^{+0.59}_{-0.34}$	$0.277^{+0.053}_{-0.068}$
HD 130437	$(0.53^{+0.78}_{-0.28}) \cdot 10^{-2}$	$(0.09^{+0.12}_{-0.04}) \cdot 10^{-2}$	-	-	0.43	$13.4^{+4.6}_{-3.8}$	$0.046^{+0.077}_{-0.026}$
HD 132947	-	-	-1.6 ± 1.2[1]	-	-0.19	$2.22^{+0.11}_{-0.11}$	$4.05^{+0.32}_{-0.20}$
HD 135344	-	-	-10.03 ± 0.65[6]	-	-	$1.736^{+0.087}_{-0.087}$	$5.75^{+0.29}_{-0.67}$
HD 135344B	$0.248^{+0.050}_{-0.044}$	$0.171^{+0.030}_{-0.024}$	-10.26 ± 0.94[1]	s[6]	-	$1.432^{+0.072}_{-0.072}$	$8.93^{+0.45}_{-0.91}$
HD 139614	$0.106^{+0.015}_{-0.024}$	$0.211^{+0.005}_{-0.015}$	-12.8 ± 1.5[1]	s[21]	-0.19	$1.481^{+0.074}_{-0.074}$	$14.5^{+1.4}_{-3.6}$
HD 141569	$(1.92^{+0.23}_{-0.34}) \cdot 10^{-2}$	$(8.75^{+0.67}_{-0.74}) \cdot 10^{-3}$	-10.4 ± 1.7[1]	d[3]	-0.18	$1.860^{+0.093}_{-0.093}$	9^{+11}_{-1}
HD 141926	$(1.97^{+0.49}_{-0.35}) \cdot 10^{-3}$	$(0.41^{+0.10}_{-0.07}) \cdot 10^{-3}$	-46.88 ± 0.55[1]	s[21]	0.65	$19.5^{+2.4}_{-2.2}$	$(2.37^{+0.76}_{-0.53}) \cdot 10^{-2}$
HD 142527	$0.59^{+0.07}_{-0.10}$	$0.53^{+0.10}_{-0.10}$	-13.1 ± 1.2[1]	s[8]	$0.30*$	$1.61^{+0.12}_{-0.08}$	$6.6^{+0.3}_{-1.5}$
HD 142666	$0.196^{+0.049}_{-0.043}$	$0.256^{+0.046}_{-0.038}$	-6.6 ± 1.4[1]	P[31]	$3.01*$	$1.493^{+0.075}_{-0.075}$	$9.33^{+0.77}_{-0.47}$
HD 143006	$0.24^{+0.15}_{-0.10}$	$0.260^{+0.089}_{-0.050}$	-10.21 ± 0.15[12]	s[12]	0.38	$1.56^{+0.10}_{-0.14}$	$3.7^{+1.7}_{-0.8}$
HD 144432	$0.225^{+0.026}_{-0.044}$	$0.221^{+0.015}_{-0.027}$	-11.80 ± 0.35[1]	d[3]	0.08	$1.386^{+0.069}_{-0.069}$	$4.98^{+0.25}_{-0.55}$
HD 149914	$0.013^{+0.009}_{-0.010}$	$(1.26^{+0.80}_{-0.80}) \cdot 10^{-3}$	-	-	-0.21	$2.97^{+0.39}_{-0.29}$	$1.74^{+0.64}_{-0.54}$

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
HD 150193	$0.162^{+0.024}_{-0.018}$	$0.186^{+0.040}_{-0.030}$	-21.7 ± 1.7[1]	d[3]	0.32	$1.891^{+0.095}_{-0.095}$	$5.48^{+0.44}_{-0.27}$
HD 155448	-	-	-	-	-0.09	$4.8^{+0.9}_{-1.0}$	$0.44^{+0.45}_{-0.20}$
HD 158643	$0.016^{+0.010}_{-0.016}$	$0.023^{+0.007}_{-0.012}$	-3.30 ± 0.10[3]	d[3]	-0.13	$3.35^{+0.79}_{-0.22}$	$1.22^{+0.29}_{-0.57}$
HD 163296	$0.175^{+0.022}_{-0.029}$	$0.199^{+0.051}_{-0.047}$	-22.80 ± 0.68[3]	d[3]	0.20	$1.833^{+0.092}_{-0.092}$	$7.6^{+1.1}_{-1.2}$
HD 169142	$0.028^{+0.032}_{-0.009}$	$0.029^{+0.020}_{-0.007}$	-22.04 ± 0.15[12]	s[12]	0.00	$2.00^{+0.13}_{-0.13}$	9^{+11}_{-4}
HD 17081	$(0.20^{+0.43}_{-0.20}) \cdot 10^{-2}$	$(0.21^{+0.22}_{-0.21}) \cdot 10^{-3}$	-	-	0.27	$3.94^{+0.81}_{-0.47}$	$0.89^{+0.40}_{-0.41}$
HD 174571	$(0.00^{+0.16}_{-0.00}) \cdot 10^{-2}$	-	1.79 ± 0.20[32]	d[21]	0.52	$12.9^{+6.4}_{-3.2}$	$0.041^{+0.057}_{-0.027}$
HD 176386	$0.032^{+0.034}_{-0.010}$	-	-8.235 ± 0.050[13]	s[13]	-0.16	$2.30^{+0.14}_{-0.30}$	4^{+16}_{-1}
HD 179218	$0.038^{+0.030}_{-0.014}$	$0.161^{+0.067}_{-0.032}$	-13.60 ± 0.41[3]	s[3]	-0.13	$2.98^{+0.18}_{-0.30}$	$1.66^{+0.54}_{-0.26}$
HD 199603	-	$(0.04^{+0.13}_{-0.04}) \cdot 10^{-2}$	-	-	0.53	$2.03^{+0.10}_{-0.10}$	$4.05^{+0.20}_{-0.20}$
HD 200775	$0.041^{+0.023}_{-0.022}$	$0.039^{+0.028}_{-0.022}$	-63.8 ± 5.9[6]	d[6]	0.07	$5.3^{+1.3}_{-0.5}$	$0.41^{+0.15}_{-0.20}$
HD 235495	$0.33^{+0.06}_{-0.14}$	$0.32^{+0.06}_{-0.13}$	-	-	0.31	$2.05^{+0.30}_{-0.10}$	$4.7^{+0.8}_{-1.5}$
HD 244314	$0.173^{+0.019}_{-0.012}$	$0.175^{+0.016}_{-0.011}$	-29.6 ± 1.7[1]	s[21]	0.16	$1.691^{+0.093}_{-0.085}$	$7.43^{+0.37}_{-0.54}$
HD 244604	$0.169^{+0.020}_{-0.011}$	$0.145^{+0.016}_{-0.010}$	-12.5 ± 1.6[1]	P[7]	0.07	$1.98^{+0.10}_{-0.10}$	$4.89^{+0.24}_{-0.52}$
HD 245185	$0.133^{+0.018}_{-0.024}$	$0.333^{+0.039}_{-0.049}$	-28.0 ± 1.5[1]	s[26]	0.38	$1.92^{+0.18}_{-0.10}$	8^{+12}_{-3}
HD 249879	$0.065^{+0.023}_{-0.029}$	$0.193^{+0.060}_{-0.077}$	-47.7 ± 2.0[15]	s[21]	0.00	$2.25^{+0.49}_{-0.16}$	5^{+15}_{-2}
HD 250550	$0.286^{+0.042}_{-0.048}$	$0.357^{+0.055}_{-0.061}$	-58.9 ± 1.1[1]	P[22]	0.63	$2.60^{+0.30}_{-0.14}$	$2.56^{+0.43}_{-0.67}$
HD 259431	$0.07^{+0.11}_{-0.03}$	$0.11^{+0.19}_{-0.06}$	-76.5 ± 2.1[7]	d[7]	0.35	$5.2^{+1.8}_{-1.3}$	$0.42^{+0.53}_{-0.28}$
HD 287823	$0.128^{+0.008}_{-0.017}$	$0.158^{+0.007}_{-0.013}$	-5.4 ± 1.6[1]	d[31]	-0.17	$1.704^{+0.085}_{-0.085}$	$7.43^{+0.37}_{-0.37}$
HD 288012	-	-	-	-	-0.15	$2.22^{+0.39}_{-0.13}$	$3.8^{+0.6}_{-1.4}$
HD 290380	$0.24^{+0.04}_{-0.10}$	$0.208^{+0.020}_{-0.054}$	-11.59 ± 0.70[14]	s[14]	0.20	$1.53^{+0.16}_{-0.08}$	$7.2^{+0.8}_{-1.7}$
HD 290409	$0.072^{+0.006}_{-0.011}$	$0.114^{+0.012}_{-0.016}$	-14.6 ± 1.6[1]	d[21]	0.24	$1.90^{+0.18}_{-0.09}$	7^{+13}_{-2}
HD 290500	$0.115^{+0.010}_{-0.018}$	$0.121^{+0.013}_{-0.017}$	-14.6 ± 1.3[1]	d[21]	2.05	$1.383^{+0.082}_{-0.069}$	$10.4^{+9.3}_{-3.3}$

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
HD 290764	$0.245^{+0.065}_{-0.058}$	$0.221^{+0.040}_{-0.033}$	-15.7 ± 1.4[1]	s[21]	0.11	$1.69^{+0.13}_{-0.08}$	$6.9^{+0.5}_{-1.4}$
HD 290770	$0.152^{+0.013}_{-0.020}$	$0.136^{+0.011}_{-0.016}$	-37.1 ± 1.4[1]	P[31]	-0.04	$2.22^{+0.11}_{-0.11}$	$4.59^{+0.49}_{-0.54}$
HD 305298	$(0.87^{+0.15}_{-0.10})\cdot10^{-3}$	$(4.62^{+0.47}_{-0.35})\cdot10^{-3}$	-3.24 ± 0.52[1]	s[31]	-0.09	$17.7^{+2.1}_{-2.0}$	$0.04^{+0.31}_{-0.01}$
HD 313571	$(1.24^{+0.61}_{-0.66})\cdot10^{-2}$	$(0.31^{+0.12}_{-0.15})\cdot10^{-2}$	-38.8 ± 1.7[2]	s[2]	0.23	$7.3^{+2.9}_{-1.2}$	$0.17^{+0.14}_{-0.11}$
HD 31648	$0.166^{+0.027}_{-0.066}$	$0.167^{+0.032}_{-0.053}$	-19.40 ± 0.58[3]	d[3]	0.01	$1.78^{+0.13}_{-0.09}$	$6.2^{+0.3}_{-1.1}$
HD 319896	$(0.96^{+0.55}_{-0.37})\cdot10^{-2}$	$(0.80^{+0.26}_{-0.21})\cdot10^{-2}$	-31.6 ± 1.4[15]	s[21]	0.00	$5.9^{+1.2}_{-0.8}$	$0.30^{+0.18}_{-0.13}$
HD 323771	$0.087^{+0.039}_{-0.024}$	$0.085^{+0.036}_{-0.022}$	-59.4 ± 2.7[15]	P[21]	0.91	$3.84^{+0.36}_{-0.38}$	$1.08^{+0.55}_{-0.26}$
HD 34282	$0.183^{+0.014}_{-0.017}$	$0.136^{+0.009}_{-0.012}$	-12.1 ± 1.7[1]	d[3]	0.71	$1.450^{+0.072}_{-0.072}$	$6.5^{+2.4}_{-0.6}$
HD 344261	$0.024^{+0.056}_{-0.024}$	-	-	-	-0.24	$1.343^{+0.067}_{-0.067}$	$12.1^{+3.8}_{-2.0}$
HD 34700	-	-	-2.400 ± 0.072[3]	s[3]	-0.09	$2.66^{+0.32}_{-0.13}$	$1.40^{+0.23}_{-0.44}$
HD 35187	$0.10^{+0.10}_{-0.06}$	$0.060^{+0.048}_{-0.031}$	-13.28 ± 0.20[22]	s[22]	-	$2.10^{+0.25}_{-0.25}$	5^{+15}_{-2}
HD 35929	$0.074^{+0.019}_{-0.028}$	$(4.27^{+0.35}_{-0.58})\cdot10^{-2}$	-7.2 ± 1.1[1]	s[7]	-0.26	$2.92^{+0.15}_{-0.15}$	$1.46^{+0.07}_{-0.17}$
HD 36112	$0.239^{+0.079}_{-0.090}$	$0.258^{+0.077}_{-0.076}$	-16.51 ± 0.63[7]	s[7]	-0.04	$1.56^{+0.11}_{-0.08}$	$8.3^{+0.4}_{-1.4}$
HD 36408	$(0.05^{+0.73}_{-0.05})\cdot10^{-2}$	-	-2.67 ± 0.20[15]	s[8]	-0.23	$6.2^{+1.2}_{-1.0}$	$0.22^{+0.17}_{-0.09}$
HD 36917	$0.047^{+0.051}_{-0.022}$	$0.050^{+0.046}_{-0.020}$	-9.17 ± 0.10[22]	s[22]	0.30	$3.71^{+0.94}_{-0.75}$	$0.99^{+0.90}_{-0.50}$
HD 36982	-	-	-	-	-0.13	$5.20^{+0.42}_{-0.29}$	$0.73^{+0.47}_{-0.17}$
HD 37258	$0.142^{+0.027}_{-0.025}$	$0.147^{+0.025}_{-0.023}$	-15.3 ± 1.6[1]	d[22]	2.22∗	$1.88^{+0.14}_{-0.11}$	8^{+12}_{-2}
HD 37357	$0.086^{+0.007}_{-0.012}$	$0.085^{+0.006}_{-0.010}$	-9.9 ± 1.5[1]	P[21]	0.56	$3.0^{+1.0}_{-0.4}$	$1.69^{+0.87}_{-0.93}$
HD 37371	-	-	-12.37 ± 0.22[11]	-	-0.19	$3.85^{+0.63}_{-0.67}$	$0.86^{+0.65}_{-0.34}$
HD 37490	$(0.09^{+0.34}_{-0.09})\cdot10^{-2}$	$(0.16^{+0.12}_{-0.09})\cdot10^{-2}$	-8.33 ± 0.18[9]	d[8]	1.55	$8.6^{+3.9}_{-1.6}$	$0.10^{+0.11}_{-0.07}$
HD 37806	$0.187^{+0.070}_{-0.067}$	$0.200^{+0.087}_{-0.076}$	-25.07 ± 0.85[9]	d[22]	0.49	$3.11^{+0.55}_{-0.33}$	$1.56^{+0.64}_{-0.60}$
HD 38087	$0.022^{+0.015}_{-0.011}$	$(0.27^{+0.16}_{-0.13})\cdot10^{-2}$	-	-	-	$3.21^{+0.79}_{-0.38}$	$1.8^{+9.2}_{-0.6}$
HD 38120	$0.089^{+0.052}_{-0.052}$	$0.22^{+0.10}_{-0.10}$	-45.1 ± 1.8[15]	s[21]	0.02	$2.37^{+0.43}_{-0.24}$	3^{+14}_{-1}

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
HD 39014	$0.009^{+0.032}_{-0.009}$	$(0.29^{+0.28}_{-0.29})\cdot10^{-2}$	-	-	0.78	$2.00^{+0.23}_{-0.10}$	$4.4^{+0.2}_{-1.2}$
HD 41511	$0.45^{+0.28}_{-0.24}$	$0.16^{+0.12}_{-0.09}$	-	P[8]	0.33	$5.8^{+1.3}_{-0.8}$	$0.22^{+0.14}_{-0.10}$
HD 45677	$0.146^{+0.056}_{-0.070}$	$0.51^{+0.28}_{-0.27}$	-61.8 ± 5.7[6]	d[6]	0.46	$4.7^{+1.2}_{-0.4}$	$0.6^{+3.8}_{-0.3}$
HD 46060	-	$(0.22^{+0.26}_{-0.14})\cdot10^{-3}$	-	-	-0.21	$9.6^{+3.4}_{-2.4}$	$0.09^{+0.12}_{-0.05}$
HD 50083	$(0.42^{+0.32}_{-0.31})\cdot10^{-2}$	$(1.53^{+0.70}_{-0.78})\cdot10^{-3}$	-47.8 ± 1.3[7]	s[7]	0.05	$11.4^{+3.7}_{-1.8}$	$0.047^{+0.037}_{-0.022}$
HD 50138	$0.238^{+0.062}_{-0.067}$	$0.34^{+0.19}_{-0.14}$	-81.7 ± 7.1[6]	d[6]	0.62	$4.17^{+0.46}_{-0.32}$	$0.63^{+0.19}_{-0.18}$
HD 56895B	-	-	-	-	-0.16	$1.53^{+0.11}_{-0.08}$	$8.3^{+0.4}_{-1.4}$
HD 58647	$0.046^{+0.019}_{-0.018}$	$0.054^{+0.025}_{-0.019}$	-11.40 ± 0.30[3]	d[3]	-0.17	$3.87^{+0.33}_{-0.19}$	$0.84^{+0.12}_{-0.18}$
HD 59319	$(0.33^{+0.09}_{-0.23})\cdot10^{-2}$	$(0.83^{+0.08}_{-0.16})\cdot10^{-3}$	-1.3 ± 6.2[1]	-	-0.12	$3.81^{+0.31}_{-0.26}$	$0.96^{+0.24}_{-0.20}$
HD 68695	$0.129^{+0.009}_{-0.017}$	$0.138^{+0.009}_{-0.015}$	-16.0 ± 1.7[1]	s[21]	0.24	$1.833^{+0.092}_{-0.092}$	$7.6^{+1.1}_{-1.2}$
HD 76534	$(2.46^{+0.39}_{-0.43})\cdot10^{-3}$	$(7.41^{+0.73}_{-0.79})\cdot10^{-4}$	-16.84 ± 0.67[1]	d[8]	0.65	$7.46^{+0.51}_{-0.37}$	$0.171^{+0.023}_{-0.028}$
HD 85567	$0.079^{+0.016}_{-0.015}$	$0.089^{+0.025}_{-0.020}$	-42.8 ± 2.1[1]	s[31]	0.17	$6.32^{+0.53}_{-0.39}$	$0.217^{+0.045}_{-0.051}$
HD 87403	$(1.70^{+0.14}_{-0.43})\cdot10^{-2}$	$(2.12^{+0.16}_{-0.36})\cdot10^{-3}$	-3.0 ± 1.2[1]	d[31]	-0.19	$5.51^{+0.65}_{-0.53}$	$0.28^{+0.11}_{-0.08}$
HD 87643	$0.11^{+0.22}_{-0.08}$	$0.19^{+0.38}_{-0.13}$	-145 ± 14[6]	d[30]	5.31	18^{+11}_{-7}	$0.020^{+0.052}_{-0.010}$
HD 94509	$(1.23^{+0.11}_{-0.31})\cdot10^{-2}$	$(3.26^{+0.56}_{-0.69})\cdot10^{-3}$	-23.17 ± 0.82[1]	d[31]	-0.10	$5.7^{+1.1}_{-0.8}$	$0.28^{+0.17}_{-0.12}$
HD 95881	$0.228^{+0.028}_{-0.036}$	$0.284^{+0.044}_{-0.048}$	-21.9 ± 1.0[1]	d[6]	0.10	$5.50^{+0.50}_{-0.27}$	$0.280^{+0.050}_{-0.071}$
HD 96042	$(1.44^{+0.32}_{-0.31})\cdot10^{-3}$	$(1.23^{+0.21}_{-0.21})\cdot10^{-4}$	-0.75 ± 0.89[1]	d[21]	0.05	$20.7^{+3.9}_{-2.9}$	$(1.90^{+0.79}_{-0.54})\cdot10^{-2}$
HD 9672	$(0.27^{+0.42}_{-0.27})\cdot10^{-2}$	$(0.73^{+0.53}_{-0.73})\cdot10^{-3}$	-	-	-0.14	$1.810^{+0.090}_{-0.090}$	$6.89^{+0.34}_{-0.51}$
HD 97048	$0.109^{+0.017}_{-0.018}$	$0.139^{+0.022}_{-0.022}$	-38.0 ± 1.4[1]	d[31]	-0.10	$2.25^{+0.11}_{-0.13}$	$4.4^{+1.1}_{-0.3}$
HD 98922	$0.189^{+0.015}_{-0.016}$	$0.219^{+0.019}_{-0.019}$	-24.7 ± 1.1[1]	P[4]	$-0.02*$	$6.17^{+0.37}_{-0.31}$	$0.204^{+0.010}_{-0.038}$
HR 5999	$0.180^{+0.017}_{-0.019}$	$0.163^{+0.017}_{-0.018}$	-15.64 ± 0.28[17]	d[8]	$23.51*$	$2.43^{+0.12}_{-0.12}$	$2.73^{+0.26}_{-0.35}$
HT CMa	$0.419^{+0.061}_{-0.064}$	$0.596^{+0.085}_{-0.088}$	-32.8 ± 1.2[1]	d[22]	0.58	$2.12^{+0.19}_{-0.12}$	5^{+15}_{-1}
HU CMa	$(6.23^{+0.52}_{-0.54})\cdot10^{-2}$	$(12.87^{+0.90}_{-0.92})\cdot10^{-2}$	-61.09 ± 0.99[1]	d[22]	0.96	$3.02^{+0.15}_{-0.15}$	$2.04^{+0.34}_{-0.15}$

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
Hen 3-1121	$0.01^{+0.16}_{-0.01}$	$0.02^{+0.25}_{-0.02}$	-1.590 ± 0.050^{2}	s[2]	1.83	10^{+19}_{-4}	$0.1^{+1.3}_{-0.1}$
Hen 3-1121S	-	$(0.96^{+0.82}_{-0.35})\cdot10^{-3}$	-1.29 ± 0.00^{2}	s[2]	-	$12.1^{+5.2}_{-3.2}$	$0.08^{+0.53}_{-0.04}$
Hen 3-1191	$0.10^{+0.10}_{-0.04}$	$0.30^{+0.42}_{-0.15}$	-	-	-	$8.1^{+2.1}_{-0.4}$	$0.23^{+0.37}_{-0.11}$
Hen 3-823	$(0.50^{+0.32}_{-0.31})\cdot10^{-2}$	$(1.07^{+0.45}_{-0.53})\cdot10^{-3}$	-29.8 ± 1.2^{2}	d[2]	0.46	$4.8^{+1.6}_{-0.5}$	$0.6^{+1.5}_{-0.3}$
Hen 3-847	$0.126^{+0.029}_{-0.022}$	$1.72^{+0.67}_{-0.47}$	-108.62 ± 0.87^{1}	s[21]	1.24	$3.00^{+0.60}_{-0.15}$	$2.4^{+8.5}_{-1.1}$
Hen 3-938	$0.018^{+0.012}_{-0.005}$	$0.021^{+0.018}_{-0.008}$	-93.0 ± 4.5^{27}	d[21]	2.03	$25.8^{+6.6}_{-6.0}$	$0.017^{+0.012}_{-0.006}$
IL Cep	-	-	-22.83 ± 0.54^{7}	s[7]	$-0.15*$	$9.8^{+2.7}_{-1.3}$	$0.070^{+0.044}_{-0.033}$
IP Per	$0.25^{+0.15}_{-0.08}$	$0.155^{+0.073}_{-0.035}$	-30.6 ± 1.1^{5}	d[26]	-	$1.56^{+0.11}_{-0.12}$	$12.0^{+8.0}_{-3.3}$
KK Oph	$1.24^{+0.34}_{-0.25}$	$1.58^{+0.52}_{-0.37}$	-41.5 ± 1.7^{1}	s[3]	-*	$1.513^{+0.076}_{-0.076}$	$18.5^{+1.5}_{-1.4}$
LKHa 260	$0.050^{+0.025}_{-0.016}$	$0.080^{+0.037}_{-0.024}$	-	-	50.67	$3.03^{+0.53}_{-0.31}$	$2.2^{+8.7}_{-0.8}$
LKHa 338	$0.72^{+0.50}_{-0.18}$	$3.0^{+1.9}_{-0.7}$	-59.1 ± 2.6^{5}	s[25]	0.54	$1.885^{+0.094}_{-0.094}$	9^{+11}_{-2}
LkHa 208	$0.17^{+0.11}_{-0.09}$	$1.02^{+0.43}_{-0.30}$	-14.16 ± 0.24^{5}	d[26]	-0.18	$1.56^{+0.47}_{-0.14}$	9^{+11}_{-5}
LkHa 215	$0.067^{+0.034}_{-0.021}$	$0.071^{+0.032}_{-0.021}$	-31.2 ± 1.3^{5}	d[7]	1.87	$3.82^{+0.59}_{-0.37}$	$1.03^{+0.26}_{-0.38}$
LkHa 257	$0.026^{+0.012}_{-0.008}$	$(1.79^{+0.71}_{-0.44})\cdot10^{-2}$	-	-	0.37	$3.08^{+0.15}_{-0.15}$	$3.6^{+1.1}_{-1.1}$
LkHa 259	$0.32^{+0.13}_{-0.06}$	$0.84^{+0.23}_{-0.12}$	-31.3 ± 1.2^{5}	-	3.19	$1.70^{+0.10}_{-0.13}$	$6.4^{+1.6}_{-0.9}$
LkHa 324	$(0.26^{+0.76}_{-0.26})\cdot10^{-2}$	$(1.99^{+0.64}_{-0.81})\cdot10^{-2}$	-21.97 ± 0.76^{5}	d[26]	-0.13	$2.82^{+0.61}_{-0.20}$	$2.12^{+0.44}_{-0.92}$
LkHa 339	$(1.63^{+0.15}_{-0.16})\cdot10^{-2}$	$(10.28^{+0.77}_{-0.71})\cdot10^{-2}$	-18.4 ± 1.9^{1}	s[31]	0.33	$2.59^{+0.13}_{-0.13}$	$2.54^{+0.23}_{-0.16}$
MQ Cas	-	-	-	-	-*	$1.80^{+0.30}_{-0.09}$	$11.3^{+8.7}_{-6.2}$
MWC 1021	$0.006^{+0.029}_{-0.006}$	$0.004^{+0.021}_{-0.003}$	-	-	-	32^{+2}_{-14}	$0.010^{+0.019}_{-0.000}$
MWC 1080	$0.021^{+0.018}_{-0.008}$	$0.019^{+0.016}_{-0.007}$	-113.2 ± 3.4^{7}	P[7]	4.02	$16.1^{+6.3}_{-4.2}$	$0.04^{+0.45}_{-0.02}$
MWC 137	$(0.84^{+0.80}_{-0.36})\cdot10^{-2}$	$0.013^{+0.013}_{-0.006}$	-371 ± 37^{6}	s[6]	0.96	23^{+11}_{-7}	$0.018^{+0.019}_{-0.008}$
MWC 297	$(1.17^{+0.60}_{-0.40})\cdot10^{-2}$	$0.015^{+0.010}_{-0.006}$	-594.5 ± 1.0^{1}	s[7]	1.44	$16.9^{+1.9}_{-1.2}$	$(2.75^{+0.62}_{-0.61})\cdot10^{-2}$
MWC 342	$0.021^{+0.023}_{-0.011}$	$0.027^{+0.035}_{-0.015}$	-291 ± 80^{10}	d[10]	3.84	$22.4^{+9.0}_{-6.0}$	$0.016^{+0.020}_{-0.006}$

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
MWC 593	-	-	-39.6 ± 1.8^{2}	s[2]	0.62	$8.0^{+1.8}_{-1.3}$	$0.123^{+0.086}_{-0.056}$
MWC 655	$0.011^{+0.014}_{-0.005}$	$(0.26^{+0.31}_{-0.12})\cdot10^{-2}$	-15.49 ± 0.72^{32}	-	2.00	$11.5^{+4.2}_{-3.0}$	$0.07^{+0.54}_{-0.04}$
MWC 657	$0.017^{+0.085}_{-0.014}$	$0.03^{+0.14}_{-0.02}$	-186 ± 13^{36}	d[36]	-	18^{+15}_{-9}	$0.02^{+0.50}_{-0.01}$
MWC 878	$0.012^{+0.018}_{-0.006}$	$0.041^{+0.064}_{-0.021}$	-55.2 ± 2.7^{2}	s[2]	0.09	$13.5^{+5.7}_{-3.9}$	$0.05^{+0.47}_{-0.03}$
MWC 953	$(0.27^{+0.38}_{-0.27})\cdot10^{-2}$	$(0.63^{+0.61}_{-0.37})\cdot10^{-3}$	-32.2 ± 1.6^{2}	d[2]	3.85	$12.7^{+6.6}_{-3.3}$	$0.042^{+0.063}_{-0.028}$
NSV 2968	$(0.83^{+0.78}_{-0.32})\cdot10^{-2}$	$(0.79^{+0.67}_{-0.28})\cdot10^{-2}$	-41.2 ± 2.0^{15}	P[21]	59.72	$9.0^{+2.0}_{-1.0}$	$0.20^{+0.40}_{-0.10}$
NV Ori	$0.126^{+0.056}_{-0.082}$	$0.173^{+0.030}_{-0.049}$	-4.00 ± 0.12^{3}	d[3]	1.28	$1.84^{+0.25}_{-0.11}$	$4.9^{+1.0}_{-1.5}$
PDS 002	$0.055^{+0.017}_{-0.049}$	$0.112^{+0.004}_{-0.029}$	-8.45 ± 0.20^{13}	s[21]	-0.12	$1.384^{+0.092}_{-0.069}$	$10.9^{+0.5}_{-2.2}$
PDS 004	$0.070^{+0.019}_{-0.035}$	$0.102^{+0.020}_{-0.040}$	-18.78 ± 0.45^{15}	s[21]	0.18	$1.90^{+0.32}_{-0.09}$	7^{+13}_{-3}
PDS 021	$0.051^{+0.019}_{-0.014}$	$0.087^{+0.031}_{-0.024}$	-96.02 ± 0.83^{1}	P[21]	1.01	$3.94^{+0.51}_{-0.45}$	$1.0^{+3.1}_{-0.3}$
PDS 022	$0.038^{+0.013}_{-0.024}$	$0.182^{+0.034}_{-0.071}$	-34.8 ± 1.2^{15}	s[21]	-	$2.46^{+0.73}_{-0.24}$	$2.9^{+0.9}_{-1.5}$
PDS 025	$0.60^{+0.12}_{-0.12}$	$0.59^{+0.11}_{-0.11}$	-22.16 ± 0.65^{15}	d[21]	17.20	$1.228^{+0.061}_{-0.061}$	$11.3^{+8.4}_{-2.4}$
PDS 123	$0.04^{+0.19}_{-0.03}$	$0.03^{+0.16}_{-0.03}$	-	P[25]	10.10	$5.0^{+3.7}_{-2.3}$	1^{+10}_{-1}
PDS 124	$(9.20^{+0.63}_{-0.92})\cdot10^{-2}$	$0.214^{+0.014}_{-0.019}$	-28.0 ± 2.0^{1}	P[31]	0.51	$2.07^{+0.10}_{-0.12}$	6^{+14}_{-1}
PDS 126	$0.141^{+0.065}_{-0.056}$	$0.127^{+0.037}_{-0.029}$	-14.26 ± 0.25^{15}	d[25]	0.05	$1.94^{+0.23}_{-0.17}$	$4.7^{+1.2}_{-1.2}$
PDS 129	$0.111^{+0.086}_{-0.056}$	$0.164^{+0.054}_{-0.034}$	-	-	0.25	$1.57^{+0.21}_{-0.16}$	$7.2^{+2.1}_{-2.2}$
PDS 130	$0.157^{+0.011}_{-0.012}$	$0.254^{+0.017}_{-0.018}$	-42.6 ± 1.3^{1}	P[21]	0.18	$2.33^{+0.12}_{-0.12}$	$3.48^{+0.27}_{-0.26}$
PDS 133	$0.104^{+0.060}_{-0.036}$	$0.151^{+0.086}_{-0.051}$	-110.9 ± 4.0^{1}	P[4]	13.08*	$2.93^{+0.45}_{-0.44}$	3^{+14}_{-1}
PDS 134	-	-	-18.20 ± 0.73^{1}	d[21]	-0.12	$4.28^{+0.52}_{-0.38}$	$0.73^{+0.22}_{-0.21}$
PDS 138	-	-	-	d[21]	0.41	$28.3^{+7.4}_{-8.2}$	$0.013^{+0.012}_{-0.003}$
PDS 174	$0.025^{+0.012}_{-0.008}$	$0.090^{+0.041}_{-0.026}$	-60.7 ± 1.5^{1}	d[21]	-	$2.71^{+0.36}_{-0.23}$	2^{+18}_{-1}
PDS 211	$0.054^{+0.047}_{-0.015}$	$0.15^{+0.10}_{-0.04}$	-43.1 ± 1.8^{15}	s[21]	1.04	$2.41^{+0.29}_{-0.28}$	3^{+14}_{-1}
PDS 24	$0.143^{+0.022}_{-0.021}$	$0.371^{+0.055}_{-0.050}$	-38.5 ± 2.2^{1}	d[21]	2.35	$1.95^{+0.10}_{-0.10}$	10^{+10}_{-4}

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
PDS 241	$(2.04^{+0.42}_{-0.53})\cdot10^{-3}$	$0.107^{+0.017}_{-0.024}$	-12.57 ± 0.51^{1}	s^{21}	0.12	$11.1^{+2.3}_{-1.3}$	$0.078^{+0.036}_{-0.028}$
PDS 27	$0.044^{+0.046}_{-0.020}$	$0.12^{+0.14}_{-0.06}$	-77.60 ± 0.85^{1}	P^{4}	15.40	$12.2^{+5.5}_{-3.4}$	$0.042^{+0.072}_{-0.027}$
PDS 277	$0.128^{+0.021}_{-0.046}$	$0.199^{+0.005}_{-0.030}$	-6.53 ± 0.15^{15}	s^{21}	-0.14	$1.543^{+0.077}_{-0.077}$	$7.76^{+0.39}_{-0.87}$
PDS 286	$(1.57^{+0.50}_{-0.42})\cdot10^{-3}$	$(0.48^{+0.15}_{-0.14})\cdot10^{-4}$	-30.77 ± 0.57^{1}	s^{21}	1.15	$31.2^{+4.5}_{-5.5}$	$(1.12^{+0.58}_{-0.12})\cdot10^{-2}$
PDS 290	-	$(0.27^{+0.79}_{-0.27})\cdot10^{-3}$	-12.66 ± 0.35^{15}	-	-	$2.35^{+0.30}_{-0.12}$	5^{+15}_{-1}
PDS 297	$(0.46^{+0.80}_{-0.46})\cdot10^{-2}$	$(0.67^{+0.66}_{-0.44})\cdot10^{-3}$	-3.9 ± 1.2^{1}	d^{31}	-0.09	$2.99^{+0.27}_{-0.31}$	$1.75^{+0.63}_{-0.35}$
PDS 324	-	-	-5.19 ± 0.20^{15}	-	0.49	$7.7^{+2.3}_{-2.2}$	$0.3^{+1.5}_{-0.1}$
PDS 33	$0.107^{+0.007}_{-0.012}$	$0.242^{+0.015}_{-0.023}$	-18.9 ± 1.6^{1}	s^{31}	-0.11	$1.850^{+0.093}_{-0.093}$	$10.7^{+9.3}_{-3.9}$
PDS 34	$0.039^{+0.035}_{-0.022}$	$0.052^{+0.044}_{-0.028}$	-50.2 ± 2.5^{15}	d^{21}	3.28	$5.3^{+1.6}_{-1.2}$	$0.7^{+4.8}_{-0.4}$
PDS 344	$(1.56^{+0.26}_{-0.17})\cdot10^{-2}$	$(4.14^{+0.57}_{-0.35})\cdot10^{-2}$	-30.03 ± 0.90^{1}	d^{21}	-0.07	$3.48^{+0.17}_{-0.23}$	$1.8^{+8.4}_{-0.2}$
PDS 361S	$0.017^{+0.019}_{-0.010}$	-	-9.32 ± 0.70^{1}	d^{21}	0.78	$5.0^{+1.0}_{-0.7}$	$0.6^{+3.8}_{-0.3}$
PDS 37	$0.019^{+0.020}_{-0.009}$	$0.058^{+0.059}_{-0.025}$	-123.76 ± 0.61^{1}	P^{4}	4.84	$10.9^{+4.5}_{-3.0}$	$0.06^{+0.10}_{-0.03}$
PDS 389	$0.086^{+0.049}_{-0.027}$	$0.057^{+0.020}_{-0.012}$	-18.94 ± 0.45^{15}	P^{21}	-0.00	$2.91^{+0.36}_{-0.33}$	$1.68^{+0.70}_{-0.48}$
PDS 415N	$0.056^{+0.091}_{-0.056}$	$0.116^{+0.045}_{-0.042}$	-1.9 ± 1.4^{1}	d^{31}	5.79	$1.21^{+0.16}_{-0.09}$	$13.1^{+5.4}_{-4.5}$
PDS 431	$(1.62^{+0.29}_{-0.32})\cdot10^{-2}$	-	-9.2 ± 1.2^{1}	d^{21}	-0.20	$2.52^{+0.27}_{-0.15}$	$2.77^{+0.45}_{-0.73}$
PDS 469	$0.045^{+0.014}_{-0.026}$	$0.087^{+0.017}_{-0.034}$	-7.78 ± 0.10^{15}	s^{21}	0.71	$2.15^{+0.55}_{-0.15}$	$4.0^{+1.4}_{-1.8}$
PDS 477	$0.036^{+0.047}_{-0.017}$	$0.048^{+0.060}_{-0.022}$	-121.2 ± 6.0^{15}	d^{21}	1.31	$8.3^{+3.3}_{-2.4}$	$0.2^{+1.6}_{-0.1}$
PDS 520	$1.35^{+0.39}_{-0.40}$	$1.31^{+0.33}_{-0.35}$	-36.5 ± 1.6^{15}	d^{21}	-	$1.26^{+0.15}_{-0.06}$	$14.1^{+5.9}_{-4.0}$
PDS 543	-	$(0.50^{+0.45}_{-0.19})\cdot10^{-3}$	-2.190 ± 0.050^{15}	s^{21}	-0.13	$30.7^{+5.0}_{-9.1}$	$0.011^{+0.012}_{-0.001}$
PDS 69	-	-	-76.4 ± 1.5^{1}	d^{21}	-0.11	$4.18^{+0.73}_{-0.51}$	$0.8^{+5.6}_{-0.3}$
PX Vul	$0.29^{+0.11}_{-0.11}$	$0.197^{+0.049}_{-0.054}$	-14.40 ± 0.43^{3}	d^{3}	3.67∗	$2.10^{+0.32}_{-0.22}$	$3.5^{+1.2}_{-1.1}$
RR Tau	$0.155^{+0.040}_{-0.054}$	$0.169^{+0.040}_{-0.051}$	-25.60 ± 0.77^{3}	d^{3}	-*	$2.82^{+0.46}_{-0.19}$	$1.98^{+0.40}_{-0.69}$
RY Ori	$0.151^{+0.080}_{-0.071}$	$0.157^{+0.037}_{-0.033}$	-15.80 ± 0.47^{3}	d^{3}	-*	$1.54^{+0.15}_{-0.08}$	$7.2^{+0.9}_{-1.7}$

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
SAO 185668	-	$(0.32^{+0.12}_{-0.15})\cdot10^{-2}$	-	-	-0.24	$9.4^{+3.6}_{-1.6}$	$0.081^{+0.073}_{-0.047}$
SAO 220669	$(0.49^{+0.28}_{-0.15})\cdot10^{-2}$	$(1.19^{+0.51}_{-0.32})\cdot10^{-3}$	-1.12 ± 0.73^{1}	-	-0.07	$7.9^{+1.2}_{-1.1}$	$0.127^{+0.082}_{-0.049}$
SV Cep	$0.207^{+0.052}_{-0.042}$	$0.348^{+0.072}_{-0.059}$	-11.70 ± 0.35^{3}	d^{3}	4.49∗	$1.550^{+0.077}_{-0.077}$	6^{+13}_{-1}
T Ori	$0.397^{+0.075}_{-0.080}$	$0.287^{+0.069}_{-0.066}$	-17.2 ± 1.4^{1}	d^{3}	-*	$2.11^{+0.14}_{-0.11}$	$4.15^{+0.56}_{-0.67}$
TY CrA	$0.042^{+0.040}_{-0.012}$	$0.54^{+0.53}_{-0.20}$	-	-	1.39	$2.06^{+0.22}_{-0.19}$	6^{+14}_{-2}
UX Ori	$0.367^{+0.031}_{-0.046}$	$0.399^{+0.035}_{-0.048}$	-12.3 ± 1.5^{1}	d^{31}	16.25∗	$1.612^{+0.091}_{-0.081}$	$11.4^{+8.6}_{-2.7}$
V1012 Ori	$0.160^{+0.014}_{-0.012}$	$0.159^{+0.011}_{-0.009}$	-11.6 ± 1.7^{1}	P^{31}	10.12∗	$1.300^{+0.065}_{-0.065}$	$8.5^{+1.1}_{-0.9}$
V1295 Aql	$0.086^{+0.050}_{-0.022}$	$0.119^{+0.062}_{-0.030}$	-25.60 ± 0.77^{3}	P^{3}	-0.09	$5.89^{+0.80}_{-0.76}$	$0.22^{+0.11}_{-0.07}$
V1478 Cyg	$0.03^{+0.21}_{-0.03}$	$0.02^{+0.16}_{-0.02}$	-	-	-	28^{+4}_{-15}	$0.010^{+0.043}_{-0.000}$
V1493 Cyg	$0.219^{+0.065}_{-0.091}$	$0.216^{+0.058}_{-0.080}$	-18.69 ± 0.48^{5}	-	7.82	$3.3^{+1.1}_{-0.5}$	$1.16^{+0.73}_{-0.67}$
V1685 Cyg	-	-	-147.2 ± 4.4^{7}	d^{7}	1.02	$9.1^{+3.9}_{-1.8}$	$0.10^{+0.11}_{-0.07}$
V1686 Cyg	-	-	-22.70 ± 0.68^{3}	d^{3}	-*	$2.85^{+0.72}_{-0.55}$	$1.2^{+1.1}_{-0.6}$
V1787 Ori	$0.239^{+0.061}_{-0.059}$	$0.309^{+0.063}_{-0.059}$	-21.16 ± 0.60^{27}	d^{21}	0.47	$1.659^{+0.094}_{-0.083}$	$7.4^{+0.6}_{-1.1}$
V1818 Ori	$0.054^{+0.050}_{-0.022}$	$0.16^{+0.15}_{-0.06}$	-45.7 ± 2.0^{27}	d^{21}	11.23	$5.3^{+1.3}_{-1.1}$	$0.37^{+0.39}_{-0.19}$
V1977 Cyg	$0.151^{+0.037}_{-0.030}$	$0.145^{+0.040}_{-0.031}$	-32.70 ± 0.98^{3}	d^{3}	3.91∗	$3.89^{+0.35}_{-0.26}$	$0.84^{+0.19}_{-0.19}$
V2019 Cyg	$0.040^{+0.038}_{-0.012}$	-	-39.6 ± 1.6^{17}	s^{26}	1.37	$3.50^{+0.48}_{-0.64}$	$1.13^{+0.91}_{-0.37}$
V346 Ori	$0.108^{+0.011}_{-0.021}$	$(12.82^{+0.29}_{-0.94})\cdot10^{-2}$	-4.9 ± 1.4^{1}	P^{31}	0.07	$1.572^{+0.079}_{-0.079}$	$9.33^{+0.47}_{-0.47}$
V350 Ori	$0.124^{+0.012}_{-0.010}$	$0.142^{+0.012}_{-0.010}$	-12.6 ± 1.6^{1}	d^{3}	22.78∗	$1.706^{+0.094}_{-0.085}$	$12.2^{+7.8}_{-4.7}$
V351 Ori	$0.181^{+0.023}_{-0.070}$	$0.147^{+0.013}_{-0.038}$	-10.165 ± 0.050^{11}	d^{8}	2.29	$1.98^{+0.19}_{-0.10}$	$4.5^{+0.2}_{-1.0}$
V361 Cep	$(0.29^{+0.31}_{-0.23})\cdot10^{-2}$	$(0.45^{+0.15}_{-0.12})\cdot10^{-2}$	-32.6 ± 1.4^{5}	d^{8}	-0.21	$5.31^{+0.69}_{-0.48}$	$0.41^{+0.15}_{-0.13}$
V373 Cep	$0.27^{+0.20}_{-0.09}$	$1.12^{+0.78}_{-0.35}$	-69.9 ± 2.1^{3}	d^{3}	7.23∗	$3.18^{+0.51}_{-0.39}$	$1.63^{+0.75}_{-0.60}$
V374 Cep	$(1.12^{+0.67}_{-0.44})\cdot10^{-2}$	$(1.69^{+0.80}_{-0.52})\cdot10^{-3}$	-63.4 ± 1.7^{7}	d^{7}	2.88	$6.8^{+1.0}_{-0.9}$	$0.188^{+0.095}_{-0.065}$
V380 Ori	$0.266^{+0.078}_{-0.058}$	$0.32^{+0.12}_{-0.09}$	-95.5 ± 1.4^{1}	s^{4}	1.28	$2.82^{+0.59}_{-0.38}$	$2.0^{+1.0}_{-0.8}$

Table A.2 continued.

Name	Near IR excess ($1.24-3.4$ µm)	Mid IR excess ($3.4-22$ µm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
V388 Vel	$0.23^{+0.15}_{-0.05}$	$1.09^{+0.61}_{-0.18}$	-	-	-	$4.1^{+1.0}_{-0.9}$	$0.68^{+0.72}_{-0.35}$
V431 Sct	$0.036^{+0.048}_{-0.022}$	$0.22^{+0.34}_{-0.14}$	-126 ± 31[18]	d[18]	2.10	$9.4^{+4.5}_{-2.3}$	$0.10^{+0.50}_{-0.07}$
V594 Cas	$0.29^{+0.11}_{-0.13}$	$0.29^{+0.13}_{-0.13}$	-80.7 ± 4.0[22]	P[22]	2.06	$2.94^{+0.59}_{-0.23}$	$1.89^{+0.49}_{-0.78}$
V594 Cyg	$0.82^{+0.36}_{-0.37}$	$0.065^{+0.039}_{-0.033}$	-	-	11.43	$8.7^{+3.1}_{-1.8}$	$0.075^{+0.091}_{-0.046}$
V599 Ori	$0.167^{+0.028}_{-0.026}$	$0.102^{+0.011}_{-0.010}$	-11.7 ± 1.5[1]	d[21]	3.69	$2.03^{+0.10}_{-0.10}$	$4.29^{+0.42}_{-0.54}$
V669 Cep	$0.066^{+0.032}_{-0.020}$	$0.30^{+0.13}_{-0.08}$	-132 ± 60[19]	d[19]	0.67	$4.00^{+0.49}_{-0.48}$	$0.96^{+0.44}_{-0.30}$
V718 Sco	$0.136^{+0.024}_{-0.026}$	$0.180^{+0.019}_{-0.019}$	-6.2 ± 1.5[1]	d[21]	2.91∗	$1.605^{+0.080}_{-0.080}$	$9.8^{+2.8}_{-0.5}$
V921 Sco	$0.027^{+0.024}_{-0.011}$	$0.09^{+0.12}_{-0.05}$	-195 ± 19[20]	d[20]	3.49	$20.0^{+7.0}_{-5.0}$	$0.023^{+0.026}_{-0.012}$
VV Ser	$0.177^{+0.041}_{-0.039}$	$0.158^{+0.040}_{-0.037}$	-49.7 ± 1.5[3]	d[3]	-*	$2.89^{+0.14}_{-0.14}$	$2.8^{+8.1}_{-0.2}$
VX Cas	$0.216^{+0.049}_{-0.069}$	$0.227^{+0.045}_{-0.063}$	-22.10 ± 0.66[3]	d[3]	11.70∗	$1.88^{+0.18}_{-0.09}$	9^{+11}_{-4}
WRAY 15-1435	$(0.57^{+0.81}_{-0.29})\cdot10^{-2}$	$(0.65^{+0.80}_{-0.30})\cdot10^{-2}$	-21.2 ± 1.0[2]	s[2]	1.54	$8.7^{+2.9}_{-2.5}$	$0.2^{+1.2}_{-0.1}$
WW Vul	$0.267^{+0.060}_{-0.074}$	$0.225^{+0.044}_{-0.054}$	-19.10 ± 0.57[3]	d[3]	3.47∗	$1.95^{+0.11}_{-0.10}$	$5.08^{+0.84}_{-0.71}$
XY Per A	$0.216^{+0.057}_{-0.045}$	$0.165^{+0.051}_{-0.039}$	-9.80 ± 0.29[3]	d[3]	3.01	$2.82^{+0.29}_{-0.20}$	$1.95^{+0.43}_{-0.44}$

Notes: The Hα line profile classification is as follows: single-peaked (s), double-peaked (d) and showing a P-Cygni profile (P), both regular or inverse. An asterisk together with the variability indicator indicates that the source had been catalogued as UXOR type in the literature. See Sect. 2.2 for derivation of mass and age. References for EW values and line shapes: (1) Fairlamb et al. (2017), (2) Carmona et al. (2010), (3) Mendigutía et al. (2011b), (4) Ababakr et al. (2016), (5) Hernández et al. (2004), (6) Baines et al. (2006), (7) Wheelwright et al. (2010), (8) van den Ancker et al. (1998), (9) Oudmaijer and Drew (1999), (10) Kučerová et al. (2013), (11) Hernández et al. (2005), (12) Dunkin et al. (1997), (13) Pogodin et al. (2012), (14) Miroshnichenko et al. (1999), (15) Sartori et al. (2010), (16) Polster et al. (2012), (17) Manoj et al. (2006), (18) Miroshnichenko et al. (2004), (19) Miroshnichenko et al. (2002), (20) Borges Fernandes et al. (2007), (21) Vieira et al. (2003), (22) Boehm and Catala (1995), (23) Nakano et al. (2012), (24) Spezzi et al. (2008), (25) Hou et al. (2016), (26) Grinin and Rostopchina (1996), (27) Vieira et al. (2011), (28) Acke et al. (2005), (29) Herbig and Bell (1988), (30) Oudmaijer et al. (1998), (31) X-Shooter spectra, 2015, priv. comm, from ESO observing program 084.C-0952A, (32) Ababakr et al. (2017), (33) Zuckerman et al. (2008), (34) Frasca et al. (2016), (35) Miroshnichenko et al. (1998), (36) Miroshnichenko et al. (2000).

Table A.3 IR excess at each bandpass (defined as $F_{observed}/F_{CK}$) for each known Herbig Ae/Be star belonging to the high-quality sample of 218 sources.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
AB Aur	2.79	6.01	13.23	28.77	-	240.49	2244.34	Yes
AK Sco	1.34	1.88	3.14	7.78	11.61	49.83	400.83	-
AS 310	0.68	0.71	0.81	-	-	-	-	Yes
AS 470	1.21	1.28	1.71	2.15	2.77	3.21	5.36	-
AS 477	2.00	-	-	27.30	59.04	104.62	475.86	Yes
BD+30 549	1.81	2.08	2.35	2.79	3.39	51.04	378.04	-
BD+41 3731	0.95	0.93	0.92	0.94	1.03	2.54	35.38	-
BF Ori	1.50	2.39	4.56	10.69	21.87	147.30	617.52	Yes
BH Cep	1.23	1.79	3.24	7.10	10.97	52.96	585.75	Yes
BO Cep	1.39	1.76	2.27	4.22	5.86	15.29	319.78	-
CO Ori	1.94	2.52	4.28	6.55	13.63	36.94	172.89	-
CPM 25	2.82	7.14	19.80	73.01	215.11	2106.36	13608.54	-
CQ Tau	1.06	1.92	4.31	8.63	19.75	150.12	1840.10	Yes
DG Cir	8.79	15.43	30.63	66.60	174.10	1164.79	11702.50	Yes
GSC 1876-0892	1.69	3.44	8.57	22.43	59.77	650.69	6589.47	-
GSC 3975-0579	1.54	2.90	7.03	21.27	36.92	79.38	933.24	-
GSC 6546-3156	1.57	1.91	2.37	6.11	10.31	717.95	5359.18	-
GSC 8143-1225	1.75	2.55	4.77	9.99	15.83	26.31	361.41	-
GSC 8581-2002	1.96	1.97	2.12	1.98	2.03	7.17	76.96	-
GSC 8645-1401	1.38	2.20	4.18	9.16	15.58	69.63	719.30	-
GSC 8994-3902	1.14	1.18	1.28	-	-	-	-	Yes
HBC 217	1.32	1.63	2.42	5.27	8.16	33.06	1021.50	-
HBC 222	1.30	1.64	2.74	6.10	9.02	18.50	323.71	-
HBC 334	2.04	2.72	3.52	10.86	20.90	788.83	11504.97	Yes
HBC 442	1.29	1.40	1.91	3.40	5.54	69.45	550.58	-
HBC 7	1.22	1.26	1.55	1.61	2.34	3.56	8.26	-
HBC 705	1.45	1.55	1.94	1.79	2.64	4.36	16.23	Yes
HD 100453	1.34	2.00	4.26	9.24	24.07	134.69	1857.99	-
HD 100546	1.34	2.09	3.64	10.62	21.19	490.87	7717.64	Yes
HD 101412	1.60	2.34	4.88	15.60	44.31	209.20	1006.20	Yes
HD 104237	1.50	2.40	4.60	8.98	21.58	84.07	367.56	Yes

Table A.3 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
HD 114981	1.23	1.24	1.34	1.41	1.39	1.32	4.53	-
HD 130437	2.19	2.56	3.24	4.23	5.85	8.52	32.43	Yes
HD 132947	1.30	1.42	1.54	-	-	-	-	Yes
HD 135344	0.56	0.47	0.49	0.49	0.52	0.56	46.31	-
HD 135344B	1.38	2.05	4.04	8.55	17.29	16.84	420.92	-
HD 139614	1.39	1.76	3.16	6.59	12.20	150.84	2119.75	-
HD 141569	1.29	1.32	1.45	2.77	2.69	6.48	102.87	Yes
HD 141926	1.59	1.80	2.31	3.01	4.38	5.53	18.99	-
HD 142527	2.45	4.16	8.24	15.15	36.02	158.75	1216.10	-
HD 142666	1.64	2.50	4.63	9.18	20.00	170.41	908.47	Yes
HD 143006	1.22	1.50	2.56	7.36	14.36	24.41	365.00	-
HD 144432	1.86	2.85	5.34	8.29	17.05	146.13	837.23	Yes
HD 149914	1.30	1.33	1.42	1.36	1.50	1.06	5.65	-
HD 150193	1.90	3.19	6.08	12.60	28.58	194.31	980.17	Yes
HD 155448	-	0.94	0.94	-	-	-	-	-
HD 158643	0.97	1.17	1.80	3.31	-	26.38	72.09	-
HD 163296	1.81	3.36	7.07	17.49	37.20	202.40	853.49	Yes
HD 169142	1.28	1.72	2.71	2.86	4.77	48.62	1212.98	-
HD 17081	1.08	1.10	1.11	1.06	1.27	0.69	1.56	-
HD 174571	0.99	1.02	1.08	0.67	0.72	0.76	4.03	-
HD 176386	1.73	1.77	2.04	3.89	3.54	-	1703.78	-
HD 179218	1.04	1.37	2.54	6.64	16.19	289.90	1956.36	Yes
HD 199603	0.98	0.95	1.03	0.97	1.14	0.61	0.76	-
HD 200775	2.25	3.98	8.36	25.21	-	149.52	2914.79	Yes
HD 235495	3.04	6.46	14.38	36.37	74.80	332.15	1821.12	-
HD 244314	1.70	2.81	5.47	13.82	22.29	143.73	778.93	-
HD 244604	1.78	3.20	7.17	11.50	23.27	137.15	839.31	Yes
HD 245185	1.91	3.16	6.58	17.93	36.27	705.27	4958.59	Yes
HD 249879	1.31	1.92	5.02	23.39	59.15	474.67	2392.06	-
HD 250550	3.11	7.68	18.39	44.57	97.12	614.45	6073.51	Yes
HD 259431	2.08	4.10	9.61	27.60	94.90	305.32	2145.16	Yes
HD 287823	1.29	1.99	4.48	12.55	24.20	76.66	1144.68	-
HD 288012	0.63	0.59	0.64	-	-	-	-	-
HD 290380	1.41	2.12	3.72	7.95	12.64	51.05	701.84	-

Table A.3 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
HD 290409	1.67	2.42	4.04	4.53	6.53	219.21	3060.97	-
HD 290500	2.10	2.82	4.14	12.80	20.09	100.80	2143.13	-
HD 290764	1.71	2.95	6.34	15.76	28.69	70.06	1819.14	-
HD 290770	2.15	4.07	9.14	19.26	34.92	200.05	1406.43	-
HD 305298	1.46	1.65	1.70	2.63	4.76	271.40	12010.35	-
HD 313571	1.86	2.18	2.65	4.29	4.69	10.16	152.52	-
HD 31648	1.60	2.68	5.49	9.48	22.53	120.18	611.63	Yes
HD 319896	1.43	1.61	1.98	5.75	9.81	22.37	414.03	-
HD 323771	2.54	5.90	13.58	38.02	76.64	312.52	2185.92	-
HD 34282	1.82	3.78	8.31	17.85	28.81	104.86	1196.27	Yes
HD 344261	1.15	1.24	1.24	1.23	1.20	0.81	-	-
HD 34700	1.00	1.02	1.24	-	-	-	-	-
HD 35187	1.90	2.74	4.65	7.72	13.40	59.77	569.45	-
HD 35929	1.35	1.47	1.97	3.06	5.72	12.01	29.45	Yes
HD 36112	1.63	2.71	5.61	14.70	31.22	99.40	992.76	Yes
HD 36408	1.06	1.01	1.02	0.79	0.81	0.74	8.02	-
HD 36917	1.77	2.20	2.98	8.99	20.56	84.71	264.73	-
HD 36982	1.80	2.02	2.38	-	-	-	-	-
HD 37258	1.97	3.27	6.62	15.84	28.99	189.17	970.64	Yes
HD 37357	1.54	2.36	4.22	7.84	14.86	102.39	696.70	Yes
HD 37371	0.35	0.36	0.35	0.32	0.40	0.77	7.26	-
HD 37490	0.88	1.02	1.19	2.35	4.17	3.60	7.30	Yes
HD 37806	2.20	4.90	10.98	22.93	60.67	293.60	1185.58	Yes
HD 38087	2.01	2.45	2.77	2.94	2.84	2.04	275.15	-
HD 38120	1.80	3.11	6.10	10.25	20.61	658.25	5227.02	-
HD 39014	0.98	1.15	1.17	1.26	1.75	0.82	1.00	-
HD 41511	4.36	8.56	12.99	-	-	89.68	289.86	-
HD 45677	3.19	7.44	32.20	-	-	4700.04	28602.62	-
HD 46060	0.79	0.75	0.74	0.80	0.77	3.47	171.16	-
HD 50083	1.23	1.34	1.70	2.17	3.60	4.74	22.96	-
HD 50138	2.13	4.29	10.45	26.11	-	444.08	2375.59	-
HD 56895B	1.12	1.07	1.19	-	-	-	-	-
HD 58647	1.29	1.73	3.27	8.58	27.12	39.00	105.20	-
HD 59319	1.15	1.11	1.16	1.24	1.21	0.88	129.98	-

Table A.3 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
HD 68695	1.53	2.61	5.20	15.95	26.36	96.34	1314.61	-
HD 76534	1.31	1.31	1.40	2.21	2.55	3.90	29.05	Yes
HD 85567	1.87	3.74	8.55	22.94	58.45	209.67	836.94	Yes
HD 87403	1.35	1.36	1.49	1.67	1.76	1.15	10.21	-
HD 87643	3.76	13.14	41.40	-	-	2278.72	12716.58	-
HD 94509	1.28	1.40	1.53	2.17	2.55	3.48	17.17	Yes
HD 95881	2.28	4.48	10.93	33.15	86.18	295.79	1007.97	Yes
HD 96042	1.47	1.45	1.71	1.48	1.49	1.66	45.46	-
HD 9672	1.04	1.00	1.12	1.05	1.22	0.79	3.97	-
HD 97048	1.84	3.01	5.88	19.83	20.04	239.06	3219.89	Yes
HD 98922	2.15	4.46	11.09	-	-	352.13	1148.15	Yes
HR 5999	1.59	2.82	6.25	-	-	112.50	414.85	Yes
HT CMa	4.29	9.27	20.90	64.36	148.82	983.40	5396.58	Yes
HU CMa	1.58	3.04	7.04	20.04	42.48	535.48	3456.52	Yes
Hen 3-1121	1.73	4.13	10.01	26.17	62.07	699.79	6435.45	-
Hen 3-1121S	0.87	0.82	0.85	0.87	1.24	29.55	1962.39	-
Hen 3-1191	4.48	21.19	98.70	510.27	1995.88	10315.68	37787.17	Yes
Hen 3-823	1.35	1.45	1.73	2.15	2.63	3.62	9.41	-
Hen 3-847	1.38	2.77	12.52	127.89	761.32	9538.41	58854.42	Yes
Hen 3-938	2.93	8.65	27.33	69.17	213.51	823.51	3658.69	-
IL Cep	0.85	0.84	0.89	0.92	1.12	1.12	-	Yes
IP Per	1.80	3.18	6.85	15.40	27.39	36.94	645.90	Yes
KK Oph	3.04	12.33	41.94	102.96	243.95	1149.69	4781.29	Yes
LKHa 260	1.49	3.11	-	22.31	53.84	388.18	312.56	-
LKHa 338	4.59	11.05	38.14	193.55	635.30	6577.01	38467.51	-
LkHa 208	1.73	2.31	4.08	12.66	43.63	1204.97	7023.91	Yes
LkHa 215	2.25	4.00	7.74	27.52	53.10	183.08	1801.94	Yes
LkHa 257	1.55	2.63	4.83	8.08	13.28	58.14	1381.77	Yes
LkHa 259	2.90	3.65	5.33	10.63	24.83	580.94	10317.88	Yes
LkHa 324	-	1.06	1.13	1.61	2.73	73.32	653.45	Yes
LkHa 339	0.75	1.07	2.15	5.96	13.88	243.18	2014.27	Yes
MQ Cas	5.39	19.10	58.45	-	-	-	-	-
MWC 1021	1.54	1.76	2.77	-	-	37.54	74.27	-
MWC 1080	3.17	8.57	20.05	-	-	579.77	2359.22	Yes

Table A.3 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
MWC 137	1.94	3.25	8.25	30.58	86.52	328.65	3052.44	Yes
MWC 297	1.55	3.71	8.52	-	-	280.89	4022.76	Yes
MWC 342	2.99	6.55	15.48	-	-	717.92	3665.94	-
MWC 593	1.61	1.83	2.24	-	-	-	-	-
MWC 655	3.26	4.01	6.05	5.07	6.92	41.93	1084.04	-
MWC 657	1.41	2.98	6.76	18.91	66.61	242.44	800.51	-
MWC 878	1.61	2.72	7.30	43.34	157.49	976.37	2629.98	-
MWC 953	1.26	1.42	1.65	1.99	2.20	3.13	102.09	-
NSV 2968	2.31	4.35	10.22	12.67	41.90	297.56	2246.08	-
NV Ori	1.19	1.64	2.92	5.05	8.88	88.23	553.15	-
PDS 002	1.08	1.24	1.76	3.39	4.95	51.00	718.19	-
PDS 004	1.43	2.17	3.79	7.57	15.55	160.54	1084.87	-
PDS 021	1.93	4.07	8.97	18.07	41.18	595.37	4130.11	-
PDS 022	1.21	1.48	2.47	6.19	13.77	340.58	4104.11	-
PDS 025	3.07	6.73	15.36	37.43	63.06	373.77	3752.05	-
PDS 123	2.58	6.31	14.05	32.32	65.17	326.20	2513.33	-
PDS 124	1.69	2.51	4.95	15.08	30.21	456.06	2857.71	-
PDS 126	1.51	2.16	4.09	8.08	14.23	79.11	390.36	-
PDS 129	1.38	1.58	2.17	3.86	6.50	78.22	630.91	-
PDS 130	2.13	4.13	8.84	22.69	45.11	500.19	3967.51	-
PDS 133	2.13	6.27	16.14	23.84	77.26	645.47	6061.49	-
PDS 134	1.79	1.90	2.09	-	-	-	-	-
PDS 138	1.20	1.20	1.24	-	-	-	-	-
PDS 174	2.24	3.43	5.38	12.27	26.19	537.78	29319.27	-
PDS 211	1.35	2.16	4.29	7.32	18.45	433.58	3154.52	-
PDS 24	2.15	3.93	8.37	18.41	38.89	852.57	9838.27	-
PDS 241	1.30	1.51	1.86	7.04	19.17	5641.92	83493.60	-
PDS 27	1.78	4.54	12.10	31.59	134.49	1091.96	7522.03	-
PDS 277	1.23	1.54	2.80	6.42	10.08	70.34	1354.70	-
PDS 286	1.73	1.85	2.39	0.99	1.33	1.82	21.31	-
PDS 290	0.64	0.67	0.76	0.63	0.61	1.88	111.26	-
PDS 297	1.12	1.12	1.15	1.13	1.16	1.11	18.69	-
PDS 324	1.34	1.53	1.79	-	-	-	-	-
PDS 33	1.80	2.83	4.94	11.66	21.41	458.11	4167.59	-

Table A.3 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
PDS 34	2.61	5.93	14.05	31.75	63.65	605.75	7602.75	-
PDS 344	1.78	2.17	2.65	4.71	8.69	390.97	3434.61	-
PDS 361S	2.87	3.39	3.74	5.80	6.63	4.67	-	-
PDS 37	0.71	2.38	6.48	18.45	55.76	526.99	4328.48	-
PDS 389	1.55	2.02	3.23	4.82	8.86	38.90	328.32	-
PDS 415N	0.92	1.14	1.62	4.08	9.07	21.21	332.90	-
PDS 431	1.41	1.41	1.55	1.50	1.54	31.35	-	-
PDS 469	1.44	1.85	2.61	4.02	6.10	214.67	1185.71	-
PDS 477	3.47	9.52	24.09	61.85	129.93	1201.43	9937.68	-
PDS 520	4.96	8.63	18.99	41.85	94.93	429.65	3731.82	-
PDS 543	0.71	0.62	0.65	0.76	1.94	19.23	572.04	-
PDS 69	1.15	2.19	4.42	-	-	-	-	-
PX Vul	1.75	2.68	4.57	10.61	18.18	46.64	341.19	-
RR Tau	1.23	3.42	8.44	24.63	48.23	156.34	996.01	Yes
RY Ori	1.35	1.63	2.70	4.49	7.27	58.82	358.89	-
SAO 185668	0.87	0.85	0.84	0.82	1.01	25.56	881.06	-
SAO 220669	1.43	1.43	1.46	1.65	1.56	3.78	251.64	-
SV Cep	2.62	5.05	10.71	23.65	49.17	743.44	4567.70	Yes
T Ori	2.65	5.88	14.38	35.54	52.41	192.59	1528.40	Yes
TY CrA	1.59	2.16	2.66	5.32	14.92	3250.79	6107.38	Yes
UX Ori	2.89	4.96	10.85	23.75	46.70	338.18	2472.27	Yes
V1012 Ori	1.42	2.75	5.85	10.93	20.05	94.49	1839.69	Yes
V1295 Aql	1.30	2.07	4.38	11.48	29.27	118.54	396.22	Yes
V1478 Cyg	2.70	4.71	13.20	-	-	202.01	860.06	-
V1493 Cyg	2.39	3.92	7.04	16.28	35.08	197.28	900.73	Yes
V1685 Cyg	1.67	3.51	7.74	-	-	-	3124.43	Yes
V1686 Cyg	3.11	9.36	28.85	-	-	-	-	Yes
V1787 Ori	2.03	3.14	6.49	15.63	33.22	236.07	1369.96	Yes
V1818 Ori	0.36	1.94	7.77	28.86	86.74	453.29	2698.23	-
V1977 Cyg	2.09	4.14	10.33	25.91	66.33	153.92	794.03	Yes
V2019 Cyg	1.36	1.88	3.05	6.09	10.78	-	6223.77	Yes
V346 Ori	1.14	1.69	3.15	9.45	15.76	44.81	776.76	Yes
V350 Ori	1.28	2.45	5.47	11.91	20.34	133.53	1164.99	Yes
V351 Ori	1.73	2.44	4.68	11.46	22.68	44.22	691.22	-

Table A.3 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
V361 Cep	1.08	1.20	1.42	2.00	2.93	6.86	1741.54	Yes
V373 Cep	3.30	8.69	21.53	139.91	732.16	2133.92	42216.18	Yes
V374 Cep	1.57	1.83	2.42	1.98	2.76	4.93	27.19	Yes
V380 Ori	2.31	5.36	12.67	27.79	74.38	387.31	1548.31	Yes
V388 Vel	1.34	3.55	9.33	40.19	94.40	1665.31	21691.47	Yes
V431 Sct	1.34	2.96	15.27	129.70	-	4848.99	34190.80	Yes
V594 Cas	3.39	8.94	21.62	49.46	113.26	522.05	2764.98	Yes
V594 Cyg	26.23	29.73	43.29	25.55	34.83	54.10	169.97	-
V599 Ori	1.70	2.62	5.27	6.94	12.85	41.44	959.30	Yes
V669 Cep	1.83	3.48	10.37	44.40	174.02	2080.63	9413.59	-
V718 Sco	1.72	2.24	3.81	7.66	16.83	155.15	707.61	-
V921 Sco	3.58	8.44	24.92	-	-	3378.91	57901.14	Yes
VV Ser	3.57	8.83	21.97	46.08	121.07	428.03	1766.30	Yes
VX Cas	2.00	4.62	11.25	27.13	45.17	276.95	2735.55	Yes
WRAY 15-1435	2.24	2.88	3.34	4.39	7.36	192.05	3602.51	-
WW Vul	2.19	4.60	10.53	17.32	32.02	231.68	1366.23	Yes
XY Per A	2.61	4.51	9.33	21.53	47.11	131.33	670.35	Yes

Notes: The 'The' column value is 'Yes' if source is present in Table 1 of The et al. (1994).

A.2 Low-quality sample parameters

Table A.4 Main parameters of each known Herbig Ae/Be star belonging to the low-quality sample of 34 sources.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
BP Psc	23:22:24.7	-02:13:42	2.79 ± 0.39	350^{+110}_{-50}	5350^{+80}_{-70}	$0.73^{+0.34}_{-0.23}$	$0.83^{+0.25}_{-0.24}$	11.53	-
DK Cha	12:53:17.1	-77:07:11	4.10 ± 0.37	243^{+47}_{-28}	7250^{+130}_{-130}	$0.47^{+0.20}_{-0.16}$	$8.12^{+0.11}_{-0.14}$	18.54	-
GSC 5360-1033	05:57:49.5	-14:05:34	1.649 ± 0.034	605^{+22}_{-19}	15000^{+800}_{-1000}	$1.01^{+0.27}_{-0.30}$	$1.60^{+0.50}_{-0.50}$	13.91	Yes[29]
GSC 5988-2257	07:41:41.1	-20:00:13	-4.66 ± 0.86	980^{+520}_{-270}	16500^{+3000}_{-800}	$1.52^{+0.63}_{-0.39}$	$3.18^{+0.23}_{-0.15}$	15.52	-
GSC 6542-2339	07:24:37.0	-24:34:47	1.12 ± 0.11	850^{+160}_{-100}	32900^{+2000}_{-3900}	$3.03^{+0.27}_{-0.29}$	$5.24^{+0.14}_{-0.18}$	15.12	-
HBC 1	00:07:02.6	+65:38:38	0.16 ± 0.52	760^{+440}_{-190}	8150^{+180}_{-160}	$-0.65^{+0.47}_{-0.32}$	$1.05^{+0.19}_{-0.16}$	16.76	-
HBC 324	00:07:30.7	+65:39:53	-0.27 ± 0.47	900^{+470}_{-220}	7830^{+160}_{-220}	$0.80^{+0.47}_{-0.35}$	$2.51^{+0.25}_{-0.25}$	14.94	-
HBC 694	20:24:29.5	+42:14:02	0.90 ± 0.43	670^{+370}_{-150}	8150^{+180}_{-160}	$-0.50^{+0.58}_{-0.43}$	$1.93^{+0.50}_{-0.50}$	17.00	-
HBC 717	20:52:06.0	+44:17:16	0.49 ± 0.12	1390^{+390}_{-220}	6400^{+150}_{-150}	$1.88^{+0.32}_{-0.23}$	$2.82^{+0.27}_{-0.20}$	13.55	-
HD 245906	05:39:30.5	+26:19:55	0.67 ± 0.48	690^{+400}_{-170}	7990^{+160}_{-160}	$1.66^{+0.46}_{-0.34}$	$0.85^{+0.16}_{-0.25}$	10.55	Yes[2]
HD 53367	07:04:25.5	-10:27:16	7.77 ± 0.79	130^{+30}_{-17}	29500^{+1000}_{-1000}	$3.13^{+0.23}_{-0.17}$	$2.051^{+0.043}_{-0.050}$	7.36	Yes[4]
HD 72106B	08:29:34.9	-38:36:21	0.03 ± 0.83	600^{+430}_{-170}	8750^{+250}_{-250}	$1.85^{+0.53}_{-0.38}$	$0.51^{+0.13}_{-0.21}$	9.50	Yes[20]
Hen 2-80	12:22:23.2	-63:17:17	0.71 ± 0.38	750^{+390}_{-170}	14000^{+1000}_{-1000}	$2.12^{+0.49}_{-0.37}$	$2.97^{+0.16}_{-0.18}$	12.79	-
MWC 314	19:21:34.0	+14:52:57	0.191 ± 0.042	2980^{+550}_{-370}	16500^{+3000}_{-800}	$5.29^{+0.52}_{-0.37}$	$4.50^{+0.50}_{-0.50}$	9.80	Yes[31]
MWC 623	19:56:31.5	+31:06:20	0.173 ± 0.036	3280^{+570}_{-390}	15800^{+1000}_{-1000}	$4.58^{+0.28}_{-0.25}$	$3.77^{+0.19}_{-0.17}$	10.92	Yes[21]
MWC 930	18:26:25.2	-07:13:18	-0.162 ± 0.094	2590^{+650}_{-420}	11900^{+1700}_{-1400}	$5.85^{+0.39}_{-0.39}$	$8.72^{+0.18}_{-0.31}$	11.51	-
NX Pup	07:19:28.3	-44:35:11	-9.84 ± 0.65	1670^{+630}_{-380}	7000^{+250}_{-250}	$2.46^{+0.30}_{-0.22}$	$0.000^{+0.070}_{-0.000}$	9.63	Yes[4]
PDS 144S	15:49:15.3	-26:00:55	6.69 ± 0.12	$149.6^{+4.6}_{-4.2}$	7750^{+250}_{-250}	$-0.673^{+0.057}_{-0.057}$	$0.570^{+0.070}_{-0.080}$	12.79	-
PDS 229N	06:55:40.0	-03:09:50	-0.52 ± 0.53	880^{+470}_{-220}	12500^{+250}_{-250}	$1.70^{+0.44}_{-0.32}$	$2.13^{+0.12}_{-0.12}$	13.07	Yes[22]
PDS 322	10:52:08.7	-56:12:07	-1.30 ± 0.27	1730^{+610}_{-360}	19500^{+5000}_{-3000}	$2.88^{+0.60}_{-0.47}$	$1.39^{+0.29}_{-0.23}$	11.97	-

Table A.4 continued.

Name	RA (h:m:s)	DEC (deg:m:s)	Parallax (mas)	Distance (pc)	T_{eff} (K)	Log(L) ($L_\odot$)	A_V (mag)	V (mag)	Binary
PDS 364	13:20:03.6	-62:23:54	-0.39 ± 0.12	2430^{+660}_{-420}	12500^{+1000}_{-1000}	$2.32^{+0.31}_{-0.26}$	$1.870^{+0.050}_{-0.030}$	13.46	-
PDS 371	13:47:31.4	-36:39:50	9.87 ± 0.21	$101.4^{+3.8}_{-3.4}$	32900^{+2000}_{-3900}	$0.98^{+0.15}_{-0.21}$	$5.16^{+0.14}_{-0.18}$	15.55	-
PDS 453	17:20:56.1	-26:03:31	5.70 ± 0.57	176^{+40}_{-22}	7000^{+120}_{-250}	$-0.43^{+0.27}_{-0.23}$	$1.44^{+0.25}_{-0.29}$	13.41	-
PDS 530	18:41:34.4	+08:08:21	0.27 ± 0.17	1390^{+470}_{-260}	8150^{+180}_{-160}	$1.35^{+0.33}_{-0.24}$	$1.53^{+0.19}_{-0.16}$	13.54	-
PDS 551	18:55:23.0	+04:04:35	1.99 ± 0.12	496^{+58}_{-42}	29000^{+3900}_{-4500}	$0.93^{+0.41}_{-0.43}$	$2.90^{+0.50}_{-0.50}$	16.60	-
PDS 581	19:36:18.9	+29:32:50	0.96 ± 0.38	690^{+350}_{-150}	24500^{+4500}_{-5000}	$2.89^{+0.62}_{-0.56}$	$2.63^{+0.26}_{-0.29}$	11.75	-
PV Cep	20:45:54.0	+67:57:39	2.910 ± 0.059	343^{+12}_{-11}	8150^{+180}_{-160}	$0.00^{+0.11}_{-0.09}$	$5.12^{+0.19}_{-0.16}$	17.46	-
R CrA	19:01:53.7	-36:57:09	10.54 ± 0.70	95^{+13}_{-9}	8150^{+180}_{-160}	$-0.06^{+0.19}_{-0.15}$	$2.13^{+0.19}_{-0.16}$	11.85	-
UY Ori	05:32:00.3	-04:55:54	2.811 ± 0.082	355^{+18}_{-16}	9750^{+250}_{-250}	$0.394^{+0.072}_{-0.059}$	$1.110^{+0.020}_{-0.000}$	12.79	-
V590 Mon	06:40:44.6	+09:48:02	1.14 ± 0.13	820^{+170}_{-100}	12500^{+1000}_{-1000}	$1.38^{+0.26}_{-0.21}$	$1.030^{+0.040}_{-0.050}$	12.60	Yes[2]
V645 Cyg	21:39:58.3	+50:14:21	0.53 ± 0.40	790^{+410}_{-180}	36900^{+2000}_{-2000}	$3.60^{+0.48}_{-0.34}$	$4.51^{+0.12}_{-0.12}$	13.10	-
V892 Tau	04:18:40.6	+28:19:15	8.52 ± 0.12	$117.5^{+2.7}_{-2.5}$	11500^{+1500}_{-800}	$0.13^{+0.33}_{-0.28}$	$4.87^{+0.50}_{-0.50}$	15.17	Yes[5]
VY Mon	06:31:06.9	+10:26:05	-1.94 ± 0.38	1470^{+580}_{-330}	12000^{+4000}_{-4000}	$3.56^{+0.64}_{-0.68}$	$5.68^{+0.17}_{-0.45}$	12.97	-
Z CMa	07:03:43.2	-11:33:06	4.30 ± 0.89	230^{+150}_{-50}	8500^{+500}_{-500}	$2.25^{+0.51}_{-0.29}$	$3.37^{+0.12}_{-0.16}$	9.25	Yes[1]

Notes: Atmospheric parameters T_{eff}, A_V and V taken from the following sources in order of choice: Fairlamb et al., 2015, Montesinos et al., 2009, Hernández et al., 2004, Mendigutía et al., 2012, Carmona et al., 2010, Chen et al., 2016, Alecian et al., 2013, Sartori et al., 2010, Manoj et al., 2006, Hernández et al., 2005, Vieira et al., 2003, APASS Data Release 9 (Henden et al., 2018) and the SIMBAD database. If not available they were derived as described in Sect. 2.1.2. See Sect. 2.2 for derivation of *L*. The references for binarity are: (1) Baines et al. (2006), (2) Wheelwright et al. (2010), (3) Leinert et al. (1997), (4) Maheswar et al. (2002), (5) Wheelwright et al. (2011), (6) Alecian et al. (2013), (7) Hamaguchi et al. (2008), (8) Dunhill et al. (2015), (9) Coulson and Walther (1995), (10) Liu et al. (2000), (11) Biller et al. (2012), (12) Schütz et al. (2011), (13) Boersma et al. (2009), (14) Malkov et al. (2006), (15) Arellano Ferro and Giridhar (2003), (16) Kubát et al. (2010), (17) Morrell and Levato (1991), (18) Lazareff et al. (2017), (19) Mayer et al. (2016), (20) Folsom et al. (2008), (21) Corporon and Lagrange (1999), (22) Doering and Meixner (2009), (23) Chelli et al. (1995), (24) Miroshnichenko et al. (2002), (25) Friedemann et al. (1996), (26) Kraus et al. (2012), (27) Torres et al. (2000), (28) Aspin (1998), (29) Connelley et al. (2008), (30) Millour et al. (2009), (31) Frasca et al. (2016), (32) Marston and McCollum (2008), (33) Zhang et al. (2017).

Table A.5 Other parameters of each known Herbig Ae/Be star belonging to the low-quality sample of 34 sources.

Name	Near IR excess ($1.24-3.4$ μm)	Mid IR excess ($3.4-22$ μm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
BP Psc	$0.46^{+0.24}_{-0.19}$	$0.81^{+0.27}_{-0.21}$	-14.83 ± 0.35[33]	-	-	$1.90^{+0.50}_{-0.26}$	$1.7^{+1.6}_{-1.0}$
DK Cha	$8.0^{+2.2}_{-1.5}$	$12.6^{+4.5}_{-3.0}$	-95.3 ± 4.4[24]	-	-	$1.369^{+0.068}_{-0.068}$	$17.2^{+2.8}_{-3.6}$
GSC 5360-1033	-	-	-9.36 ± 0.20[15]	d[21]	-	-	-
GSC 5988-2257	$0.077^{+0.028}_{-0.037}$	$0.147^{+0.049}_{-0.068}$	-19.83 ± 0.75[27]	d[21]	-	-	-
GSC 6542-2339	-	-	-28.0 ± 1.2[15]	d[21]	-	-	-
HBC 1	83^{+19}_{-17}	237^{+58}_{-51}	-40.2 ± 3.1[29]	-	-	-	-
HBC 324	$0.124^{+0.062}_{-0.053}$	$1.11^{+0.31}_{-0.24}$	-26.16 ± 0.84[5]	-	-	$1.50^{+0.29}_{-0.07}$	$13.2^{+6.8}_{-7.3}$
HBC 694	$0.99^{+0.68}_{-0.41}$	$2.9^{+1.9}_{-1.1}$	-	-	-	-	-
HBC 717	$0.32^{+0.12}_{-0.12}$	$0.44^{+0.11}_{-0.11}$	-23.59 ± 0.95[5]	-	-	$3.4^{+1.0}_{-0.5}$	$0.88^{+0.52}_{-0.50}$
HD 245906	$0.145^{+0.066}_{-0.037}$	$0.091^{+0.027}_{-0.015}$	-12.80 ± 0.18[5]	P[26]	-	$2.4^{+1.1}_{-0.5}$	$2.8^{+2.7}_{-1.8}$
HD 53367	$(0.79^{+0.27}_{-0.21})\cdot10^{-3}$	$(0.42^{+0.11}_{-0.08})\cdot10^{-3}$	-14.00 ± 0.70[1]	d[21]	-	-	-
HD 72106B	$0.077^{+0.033}_{-0.020}$	$0.114^{+0.031}_{-0.018}$	-5.8 ± 1.6[1]	s[21]	-	$2.7^{+1.5}_{-0.7}$	$2.1^{+2.6}_{-1.5}$
Hen 2-80	$0.146^{+0.070}_{-0.045}$	$0.51^{+0.24}_{-0.15}$	-155.5 ± 7.5[2]	d[2]	-	$3.07^{+0.90}_{-0.39}$	$2.2^{+8.7}_{-1.3}$
MWC 314	$0.017^{+0.021}_{-0.012}$	$(0.27^{+0.32}_{-0.18})\cdot10^{-2}$	-130 ± 15[34]	d[35]	-	$24.6^{+1.2}_{-1.2}$	$(10.00^{+0.80}_{-0.50})\cdot10^{-3}$
MWC 623	$0.084^{+0.042}_{-0.028}$	$0.057^{+0.043}_{-0.024}$	-134 ± 14[16]	s[16]	-	$18.2^{+4.7}_{-3.1}$	$0.015^{+0.010}_{-0.005}$
MWC 930	-	-	-	-	-	-	-
NX Pup	$0.78^{+0.11}_{-0.15}$	$1.25^{+0.26}_{-0.27}$	-54.0 ± 3.0[1]	d[8]	-*	$5.0^{+1.4}_{-0.7}$	$0.28^{+0.20}_{-0.17}$
PDS 144S	$2.88^{+0.34}_{-0.29}$	$6.0^{+1.1}_{-0.9}$	-29.2 ± 1.5[1]	s[31]	-	-	-
PDS 229N	$0.044^{+0.012}_{-0.010}$	$0.057^{+0.011}_{-0.009}$	-2.2 ± 1.3[1]	P[21]	-	$2.51^{+0.47}_{-0.13}$	5^{+13}_{-3}
PDS 322	$(0.23^{+0.34}_{-0.23})\cdot10^{-2}$	$0.029^{+0.025}_{-0.016}$	3.792 ± 0.020[15]	-	-	$5.4^{+2.6}_{-1.5}$	$1.1^{+5.3}_{-0.9}$
PDS 364	$0.104^{+0.026}_{-0.023}$	$0.227^{+0.057}_{-0.049}$	-88.0 ± 1.2[1]	d[21]	-	$3.30^{+0.90}_{-0.50}$	$1.45^{+0.75}_{-0.75}$
PDS 371	$(0.53^{+0.33}_{-0.16})\cdot10^{-2}$	$(0.66^{+0.36}_{-0.17})\cdot10^{-2}$	-43.0 ± 2.0[27]	s[21]	-	-	-

Table A.5 continued.

Name	Near IR excess (1.24 – 3.4 μm)	Mid IR excess (3.4 – 22 μm)	Hα EW (Å)	Hα line shape	V_i	Mass ($M_\odot$)	Age (Myr)
PDS 453	$0.50^{+0.22}_{-0.15}$	$0.73^{+0.25}_{-0.17}$	-1.42 ± 0.20^{15}	d[21]	-*	-	-
PDS 530	$0.47^{+0.10}_{-0.10}$	$1.37^{+0.25}_{-0.25}$	-37.2 ± 1.4^{15}	s[21]	-	$1.89^{+0.52}_{-0.23}$	$5.2^{+2.2}_{-2.4}$
PDS 551	$0.34^{+0.45}_{-0.18}$	$0.8^{+1.0}_{-0.4}$	-51.2 ± 2.5^{15}	d[21]	-	-	-
PDS 581	$0.061^{+0.087}_{-0.030}$	$0.25^{+0.43}_{-0.14}$	-201 ± 10^{15}	s[21]	-	$5.4^{+2.9}_{-0.3}$	$0.6^{+1.1}_{-0.4}$
PV Cep	$4.7^{+1.2}_{-1.0}$	$24.7^{+6.6}_{-5.7}$	-59.1 ± 2.5^{28}	P[28]	-	-	-
R CrA	$16.6^{+7.1}_{-5.2}$	23^{+12}_{-8}	-94.7 ± 4.3^{17}	d[26]	-	-	-
UY Ori	$0.231^{+0.018}_{-0.021}$	$0.930^{+0.058}_{-0.072}$	-10.3 ± 1.6^{1}	P[31]	-	-	-
V590 Mon	$0.131^{+0.037}_{-0.028}$	$0.53^{+0.15}_{-0.11}$	-69.7 ± 1.1^{1}	d[31]	-	$2.30^{+0.13}_{-0.11}$	6^{+14}_{-1}
V645 Cyg	$(0.97^{+0.39}_{-0.28}) \cdot 10^{-2}$	$0.119^{+0.078}_{-0.046}$	-125 ± 12^{17}	d[28]	-	-	-
V892 Tau	$1.7^{+1.6}_{-0.9}$	$3.9^{+3.6}_{-2.1}$	-24.47 ± 0.89^{5}	-	-	-	-
VY Mon	$0.17^{+0.37}_{-0.10}$	$0.32^{+0.79}_{-0.20}$	-26.3 ± 2.0^{1}	P[31]	-*	$8.8^{+7.5}_{-4.0}$	$0.08^{+0.49}_{-0.07}$
Z CMa	$0.25^{+0.11}_{-0.07}$	$0.43^{+0.23}_{-0.15}$	-25.0 ± 7.0^{1}	P[8]	-	$3.8^{+2.0}_{-0.8}$	$0.80^{+0.83}_{-0.59}$

Notes: The Hα line profile classification is as follows: single-peaked (s), double-peaked (d) and showing a P-Cygni profile (P), both regular or inverse. See Sect. 2.2 for derivation of mass and age. An asterisk in the variability indicator column indicates that the source had been catalogued as UXOR type in the literature. Variability indicator values (V_i) for these objects could not be derived as they are not astrometrically well behaved. Similarly, many of these sources fall outside the pre-main sequence tracks and isochrones in the HR diagram and no masses or ages could be derived for them. I decided to present the masses and ages in the cases they were computable but these values have to be taken with caution. References for EW values and line shapes: (1) Fairlamb et al. (2017), (2) Carmona et al. (2010), (3) Mendigutía et al. (2011b), (4) Ababakr et al. (2016), (5) Hernández et al. (2004), (6) Baines et al. (2006), (7) Wheelwright et al. (2010), (8) van den Ancker et al. (1998), (9) Oudmaijer and Drew (1999), (10) Kučerová et al. (2013), (11) Hernández et al. (2005), (12) Dunkin et al. (1997), (13) Pogodin et al. (2012), (14) Miroshnichenko et al. (1999), (15) Sartori et al. (2010), (16) Polster et al. (2012), (17) Manoj et al. (2006), (18) Miroshnichenko et al. (2004), (19) Miroshnichenko et al. (2002), (20) Borges Fernandes et al. (2007), (21) Vieira et al. (2003), (22) Boehm and Catala (1995), (23) Nakano et al. (2012), (24) Spezzi et al. (2008), (25) Hou et al. (2016), (26) Grinin and Rostopchina (1996), (27) Vieira et al. (2011), (28) Acke et al. (2005), (29) Herbig and Bell (1988), (30) Oudmaijer et al. (1998), (31) X-Shooter spectra, 2015, priv. comm, from ESO observing program 084.C-0952A, (32) Ababakr et al. (2017), (33) Zuckerman et al. (2008), (34) Frasca et al. (2016), (35) Miroshnichenko et al. (1998), (36) Miroshnichenko et al. (2000).

Table A.6 IR excess at each bandpass (defined as $F_{observed}/F_{CK}$) for each known Herbig Ae/Be star belonging to the low-quality sample of 34 sources.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
BP Psc	1.49	2.28	4.20	10.16	22.35	168.78	2200.40	-
DK Cha	14.73	49.13	170.17	325.14	-	7497.91	50750.83	Yes
GSC 5360-1033	2.65	6.78	12.87	-	-	-	-	-
GSC 5988-2257	2.62	6.21	16.62	47.97	117.47	1245.27	6594.26	-
GSC 6542-2339	2.19	3.64	5.68	-	-	-	-	-
HBC 1	88.74	484.43	2157.04	9200.72	27745.90	156330.65	1383272.17	Yes
HBC 324	0.93	1.57	3.60	15.52	54.31	943.29	13361.89	Yes
HBC 694	3.40	8.51	21.52	103.52	255.39	1995.47	31718.13	Yes
HBC 717	1.49	2.35	4.58	12.21	23.09	149.10	1086.81	-
HD 245906	1.65	2.56	4.15	6.60	9.67	58.51	465.93	-
HD 53367	1.27	1.32	1.39	3.04	4.84	5.37	116.70	Yes
HD 72106B	1.79	1.92	2.73	5.48	10.88	138.92	976.86	-
Hen 2-80	2.60	6.89	17.96	64.65	211.29	2910.69	15743.40	-
MWC 314	2.32	2.65	3.51	2.91	6.01	7.36	23.11	-
MWC 623	4.04	7.46	12.17	28.53	78.51	193.41	397.67	-
MWC 930	0.35	0.33	0.35	0.48	0.86	0.74	2.16	-
NX Pup	1.51	4.32	13.40	50.79	137.41	359.52	1521.35	Yes
PDS 144S	3.81	14.93	58.24	317.57	681.57	2883.63	21943.52	-
PDS 229N	1.36	-	-	11.85	22.39	173.19	938.57	-
PDS 322	1.08	1.24	1.58	3.81	8.96	181.89	22064.34	-
PDS 364	1.95	4.35	10.24	25.65	54.14	737.90	10364.41	-
PDS 371	1.03	3.51	9.47	22.73	55.17	247.26	2992.30	-
PDS 453	2.35	3.96	7.79	22.88	44.93	279.44	4685.85	-
PDS 530	0.86	3.03	13.43	62.07	146.37	960.03	7986.35	-
PDS 551	32.07	111.81	327.37	1230.71	4108.60	29748.04	220447.30	-
PDS 581	1.71	9.18	41.39	207.15	1047.80	4498.40	47015.82	-
PV Cep	2.83	24.69	147.34	483.75	2383.40	21362.00	157219.18	Yes
R CrA	18.56	89.19	567.83	-	-	12353.87	184275.03	Yes
UY Ori	1.70	3.77	11.11	36.23	84.99	1932.72	12539.63	-
V590 Mon	1.88	4.42	12.28	44.17	104.49	2075.86	21654.33	Yes
V645 Cyg	0.74	2.67	19.54	-	-	12523.05	124226.58	Yes

Table A.6 continued.

Name	J excess 1.24 µm	H excess 1.66 µm	K_s excess 2.16 µm	W1 excess 3.4 µm	W2 excess 4.6 µm	W3 excess 12 µm	W4 excess 22 µm	The
V892 Tau	18.38	57.37	142.10	149.55	392.45	14054.94	136211.22	Yes
VY Mon	2.70	5.67	13.53	38.47	-	897.71	7399.24	Yes
Z CMa	1.21	2.85	9.21	-	-	373.86	2204.64	Yes

Notes: The 'The' column value is 'Yes' if source is present in Table 1 of The et al. (1994).

Appendix B

Machine learning algorithms

In this doctoral thesis I have used artificial neural networks (ANN) and two unsupervised algorithms: principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP). In this appendix I briefly introduce these algorithms in more depth.

B.1 Artificial neural networks

The main machine learning (ML) algorithm used in this dissertation is the artificial neural network. ANNs can be used both for regression or classification problems. At heart an ANN is a function of the type $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. $\mathbb{R}^n$ is the space of features and $\mathbb{R}^m$ is the prediction space. As any other supervised algorithm (see Chapter 1), ANNs are trained to approximate unknown functions with hypothesis functions by mapping between multidimensional inputs and the labels describing those inputs. The mapping is performed by repeatedly applying blocks of operations, called layers, to an input vector. Each layer consists of a linear transformation of the form:

$$\boldsymbol{z}_{l+1} = W_{l+1}\boldsymbol{x}_i + \boldsymbol{b}_{l+1}. \tag{B.1}$$

This is followed by a non-linear activation function,

$$\boldsymbol{x}_{l+1} = \sigma_{l+1}(\boldsymbol{z}_{l+1}). \tag{B.2}$$

W_l and $\boldsymbol{b}_l$ are respectively called the weight matrix and bias vector of the l'th layer, and in essence contain the algorithm parameters (θ_m). $\boldsymbol{x}_0$ is the input vector of features of a given object, $\boldsymbol{x}_l$ is called the activation of the layer l, and σ is the activation function, which is applied element-wise to its input. The size of each layer (i.e. the number of columns of the W matrix) is called the number of neurons. Thus, the activation function is the function that defines the output of the neurons within a given layer. The last layer of the ANN is called the L'th layer and the output of the ANN can be read as the L'th layers' activation. This is the hypothesis function of the ANN ($\boldsymbol{h}_\theta(x_i) := \boldsymbol{x}_L$, see Eq. B.2 and Chapter 1). $\boldsymbol{h}_\theta(x_i)$ is calculated by successively applying Eq. B.1 and Eq. B.2 to the ANNs' input $\boldsymbol{x}_0$ of the object i. The number of iterations depends on the number of layers of the ANN (in particular, $L-1$ iterations; note that W_l and $\boldsymbol{b}_l$ are different for each layer, see Eq. B.1). In classification problems each element of the $\boldsymbol{h}_\theta(x_i)$ vector corresponds to a category. By using certain activation functions in the last layer these elements can be understood as a normalised probability, and the sum of the elements of the $\boldsymbol{h}_\theta(x_i)$ vector equals one.

ANNs can be described in their entirety by their activation functions σ_l and the parameters describing the linear transformations $\theta = \{W_l, \boldsymbol{b}_l\}_{l=0}^{L}$. These latter ones are determined by the training. σ_l and the dimensionality of the layers $l = 1$ through $l = L-1$, which are called *hidden layers*, are considered hyper-parameters and are not changed during training. These hyper-parameters define the 'architecture' of the network.

Training the network means to tune the θ parameters in order to minimise the amount of wrong predictions. It is thus necessary to define a loss function (see Chapter 1). Without entering into details, an example of a simple ANN loss function is:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\log\left(\left(h_\theta(x^{(i)})\right)_k\right) + \left(1-y_k^{(i)}\right)\log\left(1-\left(h_\theta(x^{(i)})\right)_k\right)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(W_{j,i}^{(l)})^2. \quad \text{(B.3)}$$

The fundamental principles of this loss function are very similar to those of the loss function of Eq. 1.4 in Chapter 1. In the loss function of Eq. B.3 some hyper-parameters are present such as the number of neurons per layer (s_l), the number of layers (L), and the regularisation parameter λ (K is the number of categories). Regularisation is basically a term added to the loss function that sets a limit to the sensitivity of the θ parameters and imposes a penalty to complex configurations. This helps to prevent overfitting as the algorithm cannot get unreasonably complex through iterations with the optimiser. If $\lambda = 0$ the last term of Eq. B.3 is null and the ANN has no regularisation.

ANNs normally work by a two steps learning process. Forward propagation and back-propagation. First, θ is selected at random. Secondly, every object of the training set is passed forward through the layers until $\boldsymbol{h}_\theta(x_i)$ is obtained. This $\boldsymbol{h}_\theta(x_i)$ is compared with the known labelled truth $\boldsymbol{y}_i$ (i.e. the cross-entropy between the two vectors is computed). Thirdly, the cross-entropy is minimised by finding the cross-entropy's dependency of the parameters in θ. Lastly, the network is walked back by inverting the functions (Eq. B.1 and Eq. B.2) and introducing small corrections to the θ parameters so they adapt better to the labelled truth. After enough forward- and back-propagation iterations the θ parameters are fixed to the specific problem and the ANN is trained (i.e. the network has *learnt*). During training the θ parameters can be updated for every training object or in batches of objects.

One important technique that can be used to prevent overfitting is the so-called 'dropout'. Dropout consists on turning off a percentage of random neurons in every

training iteration, so not all neurons are used per iteration. Similar to regularisation, this causes the θ parameters to be less sensitive to the training data. Dropout also prevents certain neurons to pick on outliers or spurious training data and import that information when generalising. Intuitively, it can be seen as the network thinking as a whole instead of individual neurons carrying a lot of decision weight.

B.2 Unsupervised learning: PCA and UMAP

Principal component analysis is a dimensionality reduction algorithm. These algorithms reduce the number of dimensions of a data set while removing the smallest possible amount of information. They are mainly used for feature extraction or representation purposes. PCA works by finding the directions in the parameter space that maximise the variance (called principal components), and hence the information retained within them. There are as many principal components as dimensions of the parameter space, and it is up to the user to choose how many to use. These principal components are a linear combination of the original dimensions, and they are linearly uncorrelated between them. Normally, the selection of the number of PCA components is expressed in terms of the retained variance. In other words, the amount of information that was retained by moving to a lower dimension.

UMAP is based on the same conceptual idea as PCA, but it is a much more powerful non-linear algorithm. The dimensionality reduction is based on the topological properties of the data set, which the algorithm tries to keep invariant at lower dimensions (details can be found in McInnes et al., 2018). In contrast to PCA, UMAP demands the user to select several hyper-parameters, and hence its output might be biased by those decisions.

Appendix C

Spectroscopically observed candidates

This appendix contains Tables C.1, C.2, C.3, C.4, C.5, and C.6. These tables tabulate the main observed and derived parameters of the spectroscopically observed sources presented in Chapter 4. Tables C.1, C.2, and C.3 contain the data for the 145 observed pre-main sequence candidates (Sect. C.1) and Tables C.4, C.5, and C.6 contain the data for the 14 observed classical Be candidates (Sect. C.2):

1. Tables C.1 and C.4 tabulate coordinates, catalogue probability and warning flags, SIMBAD object type and spectral type (if available), our derivation of the spectral type, T_{eff}, and comments about the detected possible contaminants.

2. Tables C.2 and C.5 tabulate coordinates, instrument used to observe each source, Hα line profile and equivalent width (both observed and corrected from line absorption), Hβ emission information, and comments about the detected possible contaminants.

3. Tables C.3 and C.6 tabulate Gaia DR2 parallax, distance and G, G_{BP}, and G_{RP} photometry; together with my derivation of A_G, absolute magnitude, mass, luminosity, and age.

C.1 Observed PMS candidates

Table C.1 Derived spectral types and effective temperatures for the 145 observed PMS candidates.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
429587679827887360	00:15:19.7	+61:14:58	0.70	W	Star	-	22000 ± 1000	-
431934385541454080	00:20:45.3	+64:30:22	0.67	W	-	K/M	3865 ± 1500	evolved
428909457258627200	00:22:42.1	+61:12:23	0.97	H/V	Star	B9.5	10400 ± 500	-
430481140413947520	00:23:13.7	+61:59:32	0.69	H/V/W	-	-	25000 ± 1500	-
430854252809011200	00:36:51.8	+63:29:30	0.99	H/V/W	FUOr/F5III	A7	7800 ± 300	-
430707601141581952	00:37:22.5	+62:44:27	0.99	H/V	Em*	-	5000 ± 500	-
425559107889037568	00:38:50.1	+59:54:10	1.0	H/V	LPV*	G0	5920 ± 100	-
523968761523650304	00:49:47.5	+63:38:10	0.63	H	Star	-	9000 ± 500	-
426213008067891968	01:06:36.2	+59:46:09	0.73	H	Star	B9.5	10400 ± 500	-
509013483546651008	01:35:45.1	+58:14:21	0.85	H	Star	-	7500 ± 500	-
508164248253696384	01:49:38.8	+60:37:10	0.75	-	-	A2	8840 ± 360	-
506799479443438080	02:13:51.0	+58:22:52	0.77	H/V	DwarfNova	-	-	emiss, cont
514573748208687616	02:14:48.4	+62:26:22	0.93	W	Em*	B	21000 ± 11000	emiss, B[e]
465485914143803136	02:34:10.3	+61:24:40	0.92	H/W	Be*/B8Ve	-	8500 ± 500	-
464144544317484544	02:46:44.0	+59:21:17	0.54	-	Star	B9.5	10400 ± 500	-
464675226177231744	02:53:59.0	+60:39:59	0.72	H	YSO/B8.5	B7	14000 ± 1000	-
466147613983225472	02:59:05.1	+60:54:04	0.77	H	Be*/B4Ve	B8	12500 ± 1500	-
462017332914172800	03:14:22.9	+59:14:57	0.56	H	Star	-	8000 ± 500	-
442287245287225856	03:28:32.6	+51:13:54	0.64	H	-	-	9000 ± 500	-
462339210641162496	03:31:56.9	+60:07:38	0.53	H/W	-	-	7000 ± 500	-
250764453016220800	04:05:49.4	+51:28:35	0.94	H/V/W	Em*	-	10000 ± 1000	-
272076905372370816	04:29:17.0	+52:34:26	0.54	W	-	B1	26000 ± 2500	-
258477905041318144	04:30:16.2	+48:52:10	0.65	H	Star	-	9500 ± 750	-

Table C.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
2051184640104856 32	05:01:20.3	+43:32:51	0.85	PN	Em*/B2:e	-	22500 ± 2000	-
2013744250574245 12	05:08:13.0	+42:14:56	0.99	H/V	-	-	7500 ± 500	-
1814582150252921 60	05:14:26.9	+32:48:03	0.77	H/W	Em*/Be	-	11000 ± 1000	-
3349361015210583168	06:01:04.6	+16:54:41	0.97	H/V	V*	-	6000 ± 500	-
3348898739291597568	06:01:53.8	+16:22:37	0.99	H/V	-	K2-3	4900 ± 300	-
3123854268434443648	06:30:27.4	+01:44:06	0.79	V	-	-	6000 ± 500	-
3131337377071123200	06:32:24.6	+04:59:38	0.87	H	Star/A2	-	8500 ± 500	-
3049153761980978560	06:57:11.4	-10:57:29	0.93	H	Candidate YSO	-	9500 ± 500	-
3044960396431590784	07:05:47.2	-12:23:49	0.88	H/V	Candidate YSO	-	7500 ± 500	-
3045350894857953664	07:09:27.9	-12:12:16	0.83	H/V	Candidate YSO	-	7750 ± 500	-
3045461567573658368	07:14:31.2	-11:29:27	0.73	H	-	-	6750 ± 500	-
5546155247158673792	08:03:47.2	-33:14:44	0.84	H/W	-	-	11000 ± 1000	-
5544618168572613504	08:08:45.4	-36:08:40	0.85	V	Star/K7V	-	4500 ± 500	-
5543776728634411008	08:23:01.9	-33:50:42	0.63	H/W	-	-	6500 ± 500	-
5524619735681532160	08:51:11.4	-41:38:50	1.0	H/V	Candidate YSO	-	10000 ± 500	-
5328760495308550400	08:54:12.4	-47:37:21	0.86	H	Em*	-	4250 ± 500	-
5312087505304128000	09:26:45.8	-53:32:47	0.99	H/V	-	-	5000 ± 500	-
5258744904832986880	10:15:19.0	-57:30:57	0.74	H/V/W	-	-	10000 ± 500	abs
5350683490177538944	10:37:00.2	-58:35:34	0.98	H/V/W	Star	-	9500 ± 500	-
5254137160888722816	10:39:15.0	-60:50:17	0.97	V/W	Em*	B1	26000 ± 2500	-
5338434587073245056	10:52:59.8	-59:12:35	0.82	H/W	EmObj	-	9500 ± 500	-
5337981412138546432	10:56:32.8	-61:14:09	0.99	H/V	Star	-	7500 ± 500	-
5338316767522194688	10:56:50.8	-60:14:23	0.52	H	Star	B3	17000 ± 1000	-
5337719389118815488	11:10:13.2	-60:14:03	1.0	H/V/W	Orion V*	-	9000 ± 1000	-
5337668772877422464	11:10:39.2	-60:23:53	0.93	H/V	Candidate pMS*	-	7500 ± 500	-
5337250722275063936	11:10:46.6	-61:06:26	0.86	H/W	-	-	11000 ± 1000	-
5333514409948908288	11:31:13.4	-63:50:12	0.91	H/W	Em*/Be	B2	20600 ± 2500	-

Table C.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
5333596770242078976	11:36:19.6	-63:15:06	0.91	H	Em*	-	12000 ± 1000	-
5333977613578341248	11:37:30.2	-63:00:51	0.82	H	-	-	9000 ± 1000	abs
5333545642950621696	11:37:47.0	-63:42:56	0.92	H/PN/W	Star	-	-	emiss, B[e]
5333575329766038144	11:39:56.8	-63:29:27	0.89	H/W	Em*	-	10000 ± 1000	-
5334867393402370816	11:45:01.6	-62:11:49	0.83	H/V	Candidate AGB*	-	5500 ± 500	-
6054820460425919616	12:23:13.3	-62:20:06	0.97	PN/W	Em*/Be	-	15000 ± 5000	emiss, B[e]
5858981977697486592	13:06:34.8	-65:04:09	0.8	H	-	-	7750 ± 500	-
5868283055734381184	13:18:12.5	-62:43:40	0.83	H/W	Em*	-	10000 ± 1000	-
5868479043676144384	13:19:09.9	-62:25:09	0.83	H	Candidate YSO	-	11000 ± 2000	-
5868455988315314560	13:30:00.3	-61:37:34	0.98	H/V	Candidate YSO	-	5500 ± 500	-
5866357054949304832	13:59:29.9	-61:23:08	0.52	W	Em*	B5	15700 ± 1000	-
5866510505529424896	14:14:18.1	-61:41:54	0.93	H	Candidate AGB*	-	10000 ± 500	-
5854532425868957184	14:18:51.4	-61:24:38	0.52	H/W	Candidate YSO	-	10000 ± 1000	-
5884829984147822976	15:48:33.1	-54:29:02	0.63	V/W	-	-	5500 ± 500	-
5935057668182928256	16:16:05.3	-50:46:04	0.64	H/V/W	-	B2-B4	19000 ± 4000	-
5941445658948980864	16:23:39.8	-49:12:10	0.94	H/V	-	-	6000 ± 250	abs
5940955031917319040	16:41:18.9	-48:44:31	0.98	H/V/W	deltaCep/A0	-	9500 ± 500	-
5969536355649568640	16:52:53.8	-41:22:09	0.94	H/V	-	-	10500 ± 500	-
5966697730168623488	16:57:26.7	-40:25:56	0.62	H/W	Em*	-	10000 ± 1000	-
5972078735771910784	17:11:08.2	-40:31:57	0.99	V/W	Em*	-	4500 ± 500	-
4054609866596686336	17:33:20.8	-32:34:26	1.0	H/V/W	Erupt*RCrB	-	8000 ± 500	-
4054618520934162944	17:34:37.9	-32:36:19	0.91	H/V	EB*/A5IIIp	-	7750 ± 500	-
4094703381988286592	18:13:13.6	-19:24:09	0.8	V/W	Em*	B1.5-B3	20900 ± 4500	-
4097464668003385216	18:14:06.5	-17:02:59	0.97	H/V/W	X	-	5750 ± 500	-
4145745975926657664	18:15:28.3	-15:51:03	0.99	H/V/W	Candidate YSO	A9	7440 ± 200	-
4090478474203479296	18:15:59.5	-22:41:38	0.99	V/W	PulsV*WVir	G0	5920 ± 100	-
4097686082159430784	18:16:49.7	-16:03:16	0.63	H/V/W	-	-	5500 ± 500	-

Table C.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
4097384579759711744	18:17:12.0	-16:56:02	0.67	H/W	Candidate AGB*	-	6000 ± 500	abs
4153689436387288448	18:18:30.8	-11:48:34	0.82	H/V/W	-	B1.5	24800 ± 3000	abs
4152422554127130240	18:20:58.1	-13:40:32	0.81	V/W	Star	B	21000 ± 11000	-
4152405172406857088	18:22:00.1	-13:48:56	0.88	V/W	-	B	21000 ± 11000	-
4098138462472816384	18:22:44.2	-15:33:09	0.6	H	Em*	B	21000 ± 11000	-
4272195138879459200	18:34:12.7	-00:26:22	0.87	H/PN/V	Em*/F6	B	21000 ± 11000	emiss, B[e]
4155634296310906112	18:37:49.0	-09:03:34	0.97	V	-	B9.5	10400 ± 500	-
4256499816550576256	18:40:50.3	-05:39:41	0.63	W	-	B	21000 ± 11000	-
4256500851683628544	18:40:55.8	-05:34:38	0.97	H/V/W	Em*	B	21000 ± 11000	-
4253225574005033088	18:45:03.6	-06:06:15	0.66	V/W	EB*WUMa	-	4500 ± 500	cont
4259271891523989376	18:50:31.3	-01:24:10	0.52	-	Candidate AGB*	B7	14000 ± 1000	-
4265423830878562944	18:53:48.2	-00:37:30	0.95	V/W	Candidate AGB*	A7	7800 ± 300	-
4267078153499687296	18:56:47.7	+01:48:51	1.0	H/V/W	Candidate LP*	G0	5920 ± 100	-
2019217237173395968	19:25:14.2	+22:40:38	0.82	H/W	-	-	6000 ± 500	-
4315228344710978816	19:36:22.2	+11:52:44	0.96	H/V/W	-	A7	7800 ± 300	-
4318785810234714752	19:39:09.8	+15:56:48	0.96	PN/W	PN	-	-	emiss, cont
2020080113265080832	19:42:28.1	+23:06:51	0.98	H/V/W	Candidate AGB*	F5	6510 ± 200	-
2020089630912890368	19:43:04.4	+23:18:49	0.55	H	Em*	B7	14000 ± 1000	-
2032083584842678144	19:44:29.7	+30:14:50	0.97	H/V	LPV*	G0	5920 ± 100	-
2027980718477508480	19:47:16.5	+27:19:56	0.98	H/V/W	Irregular V*	A2	8840 ± 360	-
1828276425855506304	19:47:39.9	+23:26:39	0.52	H	C*/C-rich	-	5000 ± 500	cont
2031912537697905536	19:48:38.6	+30:02:41	0.99	H/V/W	LPV*	G0	5920 ± 100	-
2035402872974695936	19:50:08.2	+34:09:58	0.99	V	PulsV*WVir	F2	6810 ± 200	cont
2071705173505640448	19:54:28.1	+36:26:08	0.79	W	V*	M1	3680 ± 150	evolved
1823174039068377216	19:55:54.3	+19:32:35	0.94	H/V	V*	-	7000 ± 500	
1834095999134055680	19:57:55.9	+23:28:26	0.54	W	Star	B5	15700 ± 1000	abs
2059037012843484032	20:08:44.8	+35:40:31	0.9	H/V	V*	-	7000 ± 500	-

Table C.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
1836558703328498944	20:10:27.2	+27:05:28	0.99	H/V	-	F0	7200 ± 200	-
2062526622213061760	20:14:09.0	+40:29:50	0.93	H/V	Em*/B/Ae?	-	8500 ± 500	-
2061354783350172672	20:17:33.5	+38:57:20	0.76	V/W	Em*	B	21000 ± 11000	-
2068318948239115008	20:21:36.1	+40:47:56	0.89	H	Star/A0	-	7500 ± 500	-
2067485110409323648	20:25:38.8	+40:19:01	0.86	H/V/W	-	-	5500 ± 500	-
2066412811690449792	20:38:45.9	+42:07:05	0.96	H/PN/V/W	Star/B9.0	B	21000 ± 11000	emiss, B[e]
2066439165609688192	20:39:12.7	+42:20:55	0.98	H/V/W	YSO/B7.5	A0	9700 ± 600	-
2067030702869079424	20:47:23.6	+43:44:40	0.99	H/V	Ae*/A0/5	B7	14000 ± 1000	-
2066784304885540992	20:51:50.2	+43:19:30	0.6	V/W	EB*	B	21000 ± 11000	abs
2166981120209304576	20:59:05.9	+48:05:37	0.99	H/V	Orion V*	A	8600 ± 1500	-
1969788039090648832	21:10:36.9	+42:20:57	0.93	H	-	B9.5	10400 ± 500	-
2164505844663760768	21:11:19.1	+47:38:48	1.0	H/V	Em*/A	B	21000 ± 11000	-
2164554394961279360	21:20:03.2	+46:57:47	0.8	H	Star	-	7000 ± 500	-
1971703422709930496	21:20:36.3	+45:14:41	0.98	H/V/W	-	-	5750 ± 500	-
1972408072220900864	21:22:05.9	+46:51:41	0.89	H/V	-	-	6000 ± 500	-
2178435351312223488	21:35:19.2	+57:36:38	0.97	H/V	Ae*/A8	B9.5	10400 ± 500	-
1978002902777895552	21:40:03.1	+48:02:10	0.71	H/W	Star	-	7500 ± 500	-
2173413297587679744	21:43:57.0	+53:03:14	0.6	H/W	-	-	6750 ± 500	-
1981413759654971008	21:53:27.5	+52:59:58	0.98	PN/V	Em*/B[e]	B	21000 ± 11000	emiss, B[e]
1980664098879084288	21:54:21.1	+51:27:54	0.94	PN/W	Em*/Be	B	21000 ± 11000	emiss, B[e]
2198056166359436160	21:56:51.2	+55:22:46	0.61	W	Star	B0	31500 ± 3500	-
2203217037728935296	21:59:06.1	+60:40:53	0.83	H	Candidate LP*	-	10000 ± 1000	-
2005453516247868416	22:09:29.7	+54:21:27	0.51	H/W	-	-	10000 ± 1000	-
2199663691020796928	22:10:29.3	+58:49:00	0.68	H/V/W	-	-	6500 ± 500	-
2004435162317441280	22:11:13.1	+52:53:34	0.83	H/W	Star	B9.5	10400 ± 500	-
2204517656901678848	22:11:42.8	+60:36:43	0.99	H/V/W	-	G0	5920 ± 100	-
2200017424528999936	22:23:29.7	+57:37:49	0.94	PN/W	Em*	B	21000 ± 11000	-

Table C.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability PMS	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
2208194286334441344	22:41:17.0	+63:02:38	0.99	H/ID/V/W	LPV*	A4	8270 ± 300	-
2007423050801656960	22:47:40.1	+58:09:04	0.66	H/W	Be*/B1-3Ve	B5	15700 ± 1000	-
2007419820986293504	22:47:45.6	+58:06:49	0.52	W	Star/B1V	B1.5	24800 ± 3000	abs
2206875422143048960	22:50:57.1	+61:23:38	1.0	H/V	Star	A9	7440 ± 200	-
2013474586693428096	22:54:03.8	+58:54:02	1.0	H/V	Em*/B2e	-	6500 ± 500	-
2014636396825359488	22:54:15.8	+60:47:49	0.83	V/W	Be*/B2.5e	B	21000 ± 11000	-
2013507915639151488	22:58:59.1	+59:18:02	0.56	H/W	Em*	B2	20600 ± 2500	-
2207681742123365120	23:02:28.5	+63:16:31	0.92	H/V	Candidate Mi*	-	8500 ± 500	-
2208975695512792064	23:10:11.3	+65:55:59	1.0	H/V	-	-	6500 ± 1000	-
2014090042628166656	23:12:26.4	+60:58:13	0.6	H/W	Em*	B	21000 ± 11000	abs
2013834955933319680	23:16:07.6	+60:01:44	0.74	H/W	Em*	B	21000 ± 11000	-
2015682857022899840	23:27:57.8	+61:46:08	0.95	H/V	Em*	A0	9700 ± 600	-
2012261211191786624	23:40:28.0	+61:02:00	0.84	H/V/W	-	B1.5	24800 ± 3000	abs
2016307791941936896	23:42:26.7	+63:37:39	0.99	H/V/W	Mira	B	21000 ± 11000	emiss, B[e]

Notes: The first five columns were taken from the catalogue of Table 3.3. The 'HAeBe', 'ID AllW/ID' and 'ID IPH/VPH', 'Var', and 'W3W4' warning flags of Chapter 3 have been abbreviated to H, ID, V, and W respectively. For several objects the temperature was derived directly from model fitting the spectra, and thus no spectral type was derived for those sources. 'SIMBAD' column lists the object types and spectral types tabulated at the SIMBAD database at the time of writing. 'Comments' column indicates the possible detected contaminants (see Sect. 4.5): 'Emiss' marks sources with a strong emission line spectrum. 'Abs' marks sources with the Hα line in absorption.'B[e]' marks B[e] (FS CMa) stars. 'Evolved' marks evolved stars. 'Cont' indicate other types of contaminants (see main text for further discussion).

Table C.2 Hydrogen line properties of the 145 observed PMS candidates.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comments
429587679827887360	00:15:19.7	+61:14:58	CAFOS	-	-	-	-	-
431934385541454080	00:20:45.3	+64:30:22	CAFOS	-	-	-	-	evolved
428909457258627200	00:22:42.1	+61:12:23	IDS	-	-13.63	-22.60	d	-
430481140413947520	00:23:13.7	+61:59:32	CAFOS	Yes	-	-	-	-
430854252809011200	00:36:51.8	+63:29:30	IDS	-	2.29	-7.65	s	-
430707601141581952	00:37:22.5	+62:44:27	CAFOS	-	-	-	-	-
425559107889037568	00:38:50.1	+59:54:10	IDS	-	0.36	-2.56	P	-
523968761523650304	00:49:47.5	+63:38:10	CAFOS	-	-	-	-	-
426213008067891968	01:06:36.2	+59:46:09	IDS	-	3.6	-5.37	d	-
509013483546651008	01:35:45.1	+58:14:21	CAFOS	-	-	-	-	-
508164248253696384	01:49:38.8	+60:37:10	IDS	Yes	-69.68	-79.68	d	-
506799479443438080	02:13:51.0	+58:22:52	CAFOS	Yes	-	-	-	emiss, cont
514573748208687616	02:14:48.4	+62:26:22	IDS	Yes	-19.06	-	s	emiss, B[e]
465485914143803136	02:34:10.3	+61:24:40	CAFOS	Yes	-	-	-	-
464144544317484544	02:46:44.0	+59:21:17	IDS	-	-7.05	-16.02	s	-
464675226177231744	02:53:59.0	+60:39:59	IDS	-	-0.81	-6.45	s	-
466147613983225472	02:59:05.1	+60:54:04	IDS	Yes	-54.41	-60.85	s	-
462017332914172800	03:14:22.9	+59:14:57	CAFOS	-	-	-	-	-
442287245287225856	03:28:32.6	+51:13:54	CAFOS	-	-	-	-	-
462339210641162496	03:31:56.9	+60:07:38	CAFOS	-	-	-	-	-
250764453016220800	04:05:49.4	+51:28:35	CAFOS	Yes	-	-	-	-
272076905372370816	04:29:17.0	+52:34:26	CAFOS	-	-	-	-	-
258477905041318144	04:30:16.2	+48:52:10	CAFOS	Yes	-	-	-	-
205118464010485632	05:01:20.3	+43:32:51	CAFOS	Yes	-	-	-	-
201374425057424512	05:08:13.0	+42:14:56	CAFOS	-	-	-	-	-

Table C.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comments
181458215025292160	05:14:26.9	+32:48:03	CAFOS	Yes	-	-	-	-
3349361015210583168	06:01:04.6	+16:54:41	EFOSC2	-	-0.2	-3.86	d	-
3348898739291597568	06:01:53.8	+16:22:37	EFOSC2	Yes	-	-	-	-
3123854268434443648	06:30:27.4	+01:44:06	EFOSC2	Yes	-18.5	-22.16	d	-
3131337377071123200	06:32:24.6	+04:59:38	EFOSC2	-	7.2	-2.69	d	-
3049153761980978560	06:57:11.4	-10:57:29	EFOSC2	-	-6.1	-16.79	s	-
3044960396431590784	07:05:47.2	-12:23:49	EFOSC2	-	4.7	-4.22	s	-
3045350894857953664	07:09:27.9	-12:12:16	EFOSC2	-	6.4	-3.6	s	-
3045461567573658368	07:14:31.2	-11:29:27	EFOSC2	-	0.1	-4.51	d	-
5546155247158673792	08:03:47.2	-33:14:44	EFOSC2	-	-31.0	-37.48	s	-
5544618168572613504	08:08:45.4	-36:08:40	EFOSC2	-	-11.6	-12.09	d	-
5543776728634411008	08:23:01.9	-33:50:42	EFOSC2	-	3.6	-1.49	s	-
5524619735681532160	08:51:11.4	-41:38:50	EFOSC2	-	-3.3	-13.0	d	-
5328760495308550400	08:54:12.4	-47:37:21	EFOSC2	Yes	-39.0	-39.54	s	-
5312087505304128000	09:26:45.8	-53:32:47	EFOSC2	-	-17.0	-18.01	s	-
5258744904832986880	10:15:19.0	-57:30:57	EFOSC2	-	10.4	0.7	abs	abs
5350683490177538944	10:37:00.2	-58:35:34	EFOSC2	-	-32.0	-42.69	d	-
5254137160888722816	10:39:15.0	-60:50:17	EFOSC2	Yes	-60.0	-61.04	s	-
5338434587073245056	10:52:59.8	-59:12:35	EFOSC2	-	-53.0	-63.69	s	-
5337981412138546432	10:56:32.8	-61:14:09	EFOSC2	-	2.45	-6.47	s	-
5338316767522194688	10:56:50.8	-60:14:23	EFOSC2	Yes	-67.0	-71.16	s	-
5337719389118815488	11:10:13.2	-60:14:03	EFOSC2	-	-23.5	-34.03	d	-
5337668772877422464	11:10:39.2	-60:23:53	EFOSC2	-	-2.85	-11.77	d	-
5337250722275063936	11:10:46.6	-61:06:26	EFOSC2	-	-35.0	-41.48	s	-
5333514409948908288	11:31:13.4	-63:50:12	EFOSC2	Yes	-63.0	-63.5	P	-
5333596770242078976	11:36:19.6	-63:15:06	EFOSC2	-	-20.1	-27.25	d	-
5333977613578341248	11:37:30.2	-63:00:51	EFOSC2	-	11.2	0.67	abs	abs

Table C.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comments
5333545642950621696	11:37:47.0	-63:42:56	EFOSC2	Yes	-231.0	-	d	emiss, B[e]
5333575329766038144	11:39:56.8	-63:29:27	EFOSC2	Yes	-48.0	-57.7	s	-
5334867393402370816	11:45:01.6	-62:11:49	EFOSC2	-	-2.9	-4.14	P	-
6054820460425919616	12:23:13.3	-62:20:06	EFOSC2	Yes	-175.0	-	-	emiss, B[e]
5858981977697486592	13:06:34.8	-65:04:09	EFOSC2	-	0.0	-10.0	s	-
5868283055734381184	13:18:12.5	-62:43:40	EFOSC2	-	-57.0	-66.7	P	-
5868479043676144384	13:19:09.9	-62:25:09	EFOSC2	Yes	-27.0	-33.48	P	-
5868455988315314560	13:30:00.3	-61:37:34	EFOSC2	-	2.0	0.76	P	-
5866357054949304832	13:59:29.9	-61:23:08	EFOSC2	Yes	-59.0	-63.73	s	-
5866510505529424896	14:14:18.1	-61:41:54	EFOSC2	Yes	-13.0	-22.7	s	-
5854532425868957184	14:18:51.4	-61:24:38	EFOSC2	-	-1.6	-11.3	s	-
5884829984147822976	15:48:33.1	-54:29:02	EFOSC2	-	-0.5	-1.74	s	-
5935057668182928256	16:16:05.3	-50:46:04	EFOSC2	-	-5.5	-9.66	d	-
5941445658948980864	16:23:39.8	-49:12:10	EFOSC2	-	2.5	-1.16	abs	abs
5940955031917319040	16:41:18.9	-48:44:31	EFOSC2	-	-3.6	-14.29	d	-
5969536355649568640	16:52:53.8	-41:22:09	EFOSC2	-	6.9	-1.88	d	-
5966697730168623488	16:57:26.7	-40:25:56	EFOSC2	-	-70.0	-79.7	P	-
5972078735771910784	17:11:08.2	-40:31:57	EFOSC2	Yes	-16.5	-16.99	s	-
4054609866596686336	17:33:20.8	-32:34:26	EFOSC2	-	-8.7	-18.55	d	-
4054618520934162944	17:34:37.9	-32:36:19	EFOSC2	-	2.9	-7.1	d	-
4094703381988286592	18:13:13.6	-19:24:09	IDS	Yes	-25.25	-26.17	s	-
4097464668003385216	18:14:06.5	-17:02:59	EFOSC2	-	-7.9	-10.52	d	-
4145745975926657664	18:15:28.3	-15:51:03	IDS	-	-5.66	-13.87	d	-
4090478474203479296	18:15:59.5	-22:41:38	IDS	-	-1.72	-4.64	s	-
4097686082159430784	18:16:49.7	-16:03:16	EFOSC2	-	-6.4	-7.64	d	-
4097384579759711744	18:17:12.0	-16:56:02	EFOSC2	-	3.3	-0.36	abs	abs
4153689436387288448	18:18:30.8	-11:48:34	IDS	-	5.44	4.52	abs	abs

Table C.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comments
4152422554127130240	18:20:58.1	-13:40:32	IDS	Yes	-19.28	-19.82	s	-
4152405172406857088	18:22:00.1	-13:48:56	IDS	Yes	-76.21	-76.75	s	-
4098138462472816384	18:22:44.2	-15:33:09	IDS	Yes	-79.78	-80.32	d	-
4272195138879459200	18:34:12.7	-00:26:22	IDS	Yes	-377.7	-	s	emiss, B[e]
4155634296310906112	18:37:49.0	-09:03:34	IDS	-	-12.55	-21.52	d	-
4256499816550576256	18:40:50.3	-05:39:41	IDS	-	-8.31	-8.85	d	-
4256500851683628544	18:40:55.8	-05:34:38	IDS	Yes	-55.42	-55.96	s	-
4253225574005033088	18:45:03.6	-06:06:15	EFOSC2	-	-	-	-	cont
4259271891523989376	18:50:31.3	-01:24:10	IDS	Yes	-16.31	-21.95	s	-
4265423830878562944	18:53:48.2	-00:37:30	IDS	-	-8.88	-18.82	s	-
4267078153499687296	18:56:47.7	+01:48:51	IDS	-	0.26	-2.66	d	-
2019217237173395968	19:25:14.2	+22:40:38	CAFOS	Yes	-	-	-	-
4315228344710978816	19:36:22.2	+11:52:44	IDS	-	1.61	-8.33	d	-
4318785810234714752	19:39:09.8	+15:56:48	IDS	Yes	-	-	-	emiss, cont
2020080113265080832	19:42:28.1	+23:06:51	IDS	-	0.33	-4.74	s	-
2020089630912890368	19:43:04.4	+23:18:49	IDS	Yes	-34.92	-40.56	s	-
2032083584842678144	19:44:29.7	+30:14:50	IDS	-	-1.86	-4.78	d	-
2027980718477508480	19:47:16.5	+27:19:56	IDS	-	2.06	-7.94	d	-
1828276425855506304	19:47:39.9	+23:26:39	CAFOS	-	-	-	-	cont
2031912537697905536	19:48:38.6	+30:02:41	IDS	-	-1.96	-4.88	s	-
2035402872974695936	19:50:08.2	+34:09:58	IDS	-	-1.28	-4.19	d	cont
2071705173505640448	19:54:28.1	+36:26:08	CAFOS	-	-	-	-	evolved
1823174039068377216	19:55:54.3	+19:32:35	CAFOS	-	-	-	-	-
1834095999134055680	19:57:55.9	+23:28:26	IDS	-	5.68	0.95	abs	abs
2059037012843484032	20:08:44.8	+35:40:31	CAFOS	-	-	-	-	-
1836558703328498944	20:10:27.2	+27:05:28	IDS	-	-0.32	-8.24	d	-
2062526622213061760	20:14:09.0	+40:29:50	CAFOS	-	-	-	-	-

Table C.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comments
2061354783350172672	20:17:33.5	+38:57:20	IDS	Yes	-34.19	-34.73	s	-
2068318948239115008	20:21:36.1	+40:47:56	CAFOS	-	-	-	-	-
2067485110409323648	20:25:38.8	+40:19:01	CAFOS	-	-	-	-	-
2066412811690449792	20:38:45.9	+42:07:05	IDS	Yes	-267.4	-	s	emiss, B[e]
2066439165609688192	20:39:12.7	+42:20:55	IDS	Yes	-24.77	-35.31	d	-
2067030702869079424	20:47:23.6	+43:44:40	IDS	-	-4.77	-10.41	d	-
2066784304885540992	20:51:50.2	+43:19:30	IDS	-	1.38	0.84	abs	abs
2166981120209304576	20:59:05.9	+48:05:37	IDS	-	-8.17	-17.69	d	-
1969788039090648832	21:10:36.9	+42:20:57	IDS	-	-2.88	-11.85	d	-
2164505844663760768	21:11:19.1	+47:38:48	IDS	Yes	-53.98	-54.52	d	-
2164554394961279360	21:20:03.2	+46:57:47	CAFOS	-	-	-	-	-
1971703422709930496	21:20:36.3	+45:14:41	CAFOS	-	-	-	-	-
1972408072220900864	21:22:05.9	+46:51:41	CAFOS	-	-	-	-	-
2178435351312223488	21:35:19.2	+57:36:38	IDS	-	0.85	-8.12	d	-
1978002902777895552	21:40:03.1	+48:02:10	CAFOS	Yes	-	-	-	-
2173413297587679744	21:43:57.0	+53:03:14	CAFOS	-	-	-	-	-
1981413759654971008	21:53:27.5	+52:59:58	IDS	Yes	-1257.0	-	d	emiss, B[e]
1980664098879084288	21:54:21.1	+51:27:54	IDS	Yes	-157.7	-	s	emiss, B[e]
2198056166359436160	21:56:51.2	+55:22:46	CAFOS	-	-	-	-	-
2203217037728935296	21:59:06.1	+60:40:53	CAFOS	Yes	-	-	-	-
2005453516247868416	22:09:29.7	+54:21:27	CAFOS	-	-	-	-	-
2199663691020796928	22:10:29.3	+58:49:00	CAFOS	-	-	-	-	-
2004435162317441280	22:11:13.1	+52:53:34	IDS	Yes	-0.58	-9.55	s	-
2204517656901678848	22:11:42.8	+60:36:43	IDS	-	-4.49	-7.41	d	-
2200017424528999936	22:23:29.7	+57:37:49	IDS	Yes	-168.5	-169.04	d	-
2208194286334441344	22:41:17.0	+63:02:38	IDS	-	-3.04	-13.0	d	-
2007423050801656960	22:47:40.1	+58:09:04	CAFOS	-	-	-	-	-

Table C.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comments
2007419820986293504	22:47:45.6	+58:06:49	IDS	-	4.5	3.58	abs	abs
2206875422143048960	22:50:57.1	+61:23:38	IDS	-	3.23	-4.98	P	-
2013474586693428096	22:54:03.8	+58:54:02	CAFOS	-	-	-	-	-
2014636396825359488	22:54:15.8	+60:47:49	IDS	Yes	-45.78	-46.32	s	-
2013507915639151488	22:58:59.1	+59:18:02	CAFOS	Yes	-	-	-	-
2207681742123365120	23:02:28.5	+63:16:31	CAFOS	Yes	-	-	-	-
2208975695512792064	23:10:11.3	+65:55:59	CAFOS	-	-	-	-	-
2014090042628166656	23:12:26.4	+60:58:13	IDS	-	3.05	2.51	abs	abs
2013834955933319680	23:16:07.6	+60:01:44	IDS	Yes	-76.36	-76.9	s	-
2015682857022899840	23:27:57.8	+61:46:08	IDS	Yes	-49.66	-60.2	d	-
2012261211191786624	23:40:28.0	+61:02:00	IDS	-	3.51	2.59	abs	abs
2016307791941936896	23:42:26.7	+63:37:39	IDS	Yes	-123.0	-	s	emiss, B[e]

Notes: The first three columns were taken from the catalogue of Table 3.3. Hβ comlum is 'Yes' if Hβ emission was observed. Hα EW_{corr} has been corrected from the typical absorption EW values of each spectral sub-type (from Joner and Hintz, 2015). The Hα line profile classification is as follows: single-peaked (s), double-peaked (d) and showing a P-Cygni profile (P), both regular or inverse. 'Comments' column indicates the possible detected contaminants (see Sect. 4.5): 'Emiss' marks sources with a strong emission line spectrum. 'Abs' marks sources with the Hα line in absorption.'B[e]' marks B[e] (FS CMa) stars. 'Evolved' marks evolved stars. 'Cont' indicate other types of contaminants (see main text for further discussion).

Table C.3 Distances, photometry, extinctions, absolute magnitudes and stellar parameters of the 145 observed PMS candidates.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
429587679827887360	0.247 ± 0.031	3560^{+440}_{-360}	$2.639^{+0.077}_{-0.077}$	$-3.07^{+0.31}_{-0.33}$	$9.1^{+1.5}_{-1.2}$	$3.79^{+0.18}_{-0.20}$	$0.115^{+0.049}_{-0.036}$	12.72	12.33	11.77
431934385541454080	0.361 ± 0.038	2560^{+280}_{-230}	-	-	-	-	-	14.17	12.63	11.46
428909457258627200	0.356 ± 0.027	2590^{+200}_{-170}	$1.71^{+0.06}_{-0.10}$	$1.09^{+0.25}_{-0.22}$	$2.40^{+0.25}_{-0.22}$	$1.56^{+0.11}_{-0.14}$	$5.1^{+1.8}_{-1.4}$	15.19	14.87	14.40
430481140413947520	0.211 ± 0.017	4130^{+320}_{-280}	$3.08^{+0.07}_{-0.15}$	$-2.76^{+0.30}_{-0.23}$	$8.8^{+1.2}_{-1.6}$	$3.77^{+0.18}_{-0.30}$	$0.135^{+0.068}_{-0.035}$	13.90	13.40	12.77
430854252809011200	0.804 ± 0.034	1202^{+52}_{-48}	$3.74^{+0.09}_{-0.11}$	$-1.16^{+0.20}_{-0.18}$	$4.47^{+0.36}_{-0.31}$	$2.311^{+0.073}_{-0.079}$	$0.59^{+0.16}_{-0.14}$	14.07	12.98	11.96
430707601141581952	2.729 ± 0.034	$362.6^{+4.6}_{-4.5}$	$0.70^{+0.31}_{-0.52}$	$4.16^{+0.55}_{-0.33}$	$1.42^{+0.17}_{-0.40}$	$0.22^{+0.18}_{-0.24}$	$6.7^{+3.3}_{-3.3}$	13.31	12.65	11.90
425559107889037568	0.740 ± 0.031	1302^{+54}_{-50}	$2.803^{+0.027}_{-0.057}$	$1.47^{+0.14}_{-0.12}$	$2.48^{+0.13}_{-0.15}$	$1.261^{+0.047}_{-0.055}$	$2.52^{+0.54}_{-0.37}$	15.98	14.85	13.86
523968761523650304	0.306 ± 0.025	2970^{+240}_{-210}	$2.26^{+0.14}_{-0.19}$	$-1.95^{+0.35}_{-0.31}$	$5.45^{+0.75}_{-0.65}$	$2.67^{+0.16}_{-0.16}$	$0.35^{+0.17}_{-0.13}$	13.20	12.68	12.02
426213008067891968	0.325 ± 0.021	2820^{+180}_{-160}	$1.69^{+0.06}_{-0.10}$	$-0.37^{+0.22}_{-0.19}$	$3.18^{+0.20}_{-0.22}$	$2.13^{+0.10}_{-0.13}$	$1.88^{+0.39}_{-0.32}$	13.89	13.58	13.11
509013483546651008	0.395 ± 0.023	2350^{+140}_{-130}	$1.55^{+0.20}_{-0.28}$	$0.58^{+0.40}_{-0.32}$	$2.53^{+0.42}_{-0.36}$	$1.62^{+0.13}_{-0.16}$	$3.0^{+1.5}_{-1.1}$	14.48	13.99	13.38
508164248253696384	0.193 ± 0.025	4390^{+540}_{-440}	$2.14^{+0.11}_{-0.14}$	$-0.79^{+0.36}_{-0.36}$	$3.73^{+0.53}_{-0.46}$	$2.20^{+0.15}_{-0.16}$	$1.12^{+0.54}_{-0.39}$	15.05	14.56	13.91
506799479443438080	2.092 ± 0.025	$471.6^{+5.8}_{-5.6}$	-	-	-	-	-	13.55	13.34	12.98
514573748208687616	0.179 ± 0.031	4630^{+770}_{-590}	$2.97^{+0.49}_{-0.34}$	$-3.35^{+0.63}_{-0.83}$	$8.8^{+4.4}_{-2.0}$	$3.63^{+0.18}_{-0.30}$	$0.106^{+0.053}_{-0.053}$	13.46	12.95	12.22
465485914143803136	0.456 ± 0.023	2060^{+100}_{-90}	$2.37^{+0.19}_{-0.17}$	$-0.10^{+0.27}_{-0.29}$	$3.01^{+0.39}_{-0.31}$	$1.91^{+0.13}_{-0.11}$	$2.02^{+0.75}_{-0.63}$	14.43	13.85	13.11
464144544317484544	0.159 ± 0.020	5160^{+580}_{-480}	$2.42^{+0.06}_{-0.10}$	$-2.90^{+0.31}_{-0.29}$	$7.25^{+0.74}_{-0.69}$	$3.13^{+0.14}_{-0.17}$	$0.151^{+0.064}_{-0.046}$	13.57	13.08	12.44
464675226177231744	0.497 ± 0.036	1900^{+140}_{-130}	$2.416^{+0.032}_{-0.046}$	$-1.24^{+0.19}_{-0.19}$	$4.34^{+0.30}_{-0.29}$	$2.66^{+0.11}_{-0.11}$	$0.85^{+0.17}_{-0.15}$	13.00	12.58	11.99
466147613983225472	0.390 ± 0.023	2390^{+140}_{-130}	$2.75^{+0.05}_{-0.13}$	$-1.68^{+0.25}_{-0.17}$	$4.95^{+0.30}_{-0.44}$	$2.79^{+0.09}_{-0.21}$	$0.56^{+0.15}_{-0.10}$	13.47	12.95	12.27
462017332914172800	0.256 ± 0.023	3480^{+300}_{-260}	$2.19^{+0.17}_{-0.20}$	$-1.36^{+0.36}_{-0.35}$	$4.72^{+0.72}_{-0.59}$	$2.40^{+0.15}_{-0.15}$	$0.51^{+0.25}_{-0.20}$	14.13	13.53	12.82
442287245287225856	0.479 ± 0.037	1960^{+160}_{-130}	$2.34^{+0.14}_{-0.19}$	$-1.27^{+0.34}_{-0.31}$	$4.34^{+0.60}_{-0.52}$	$2.40^{+0.16}_{-0.16}$	$0.71^{+0.36}_{-0.26}$	13.08	12.54	11.87
462339210641162496	2.120 ± 0.033	$465.5^{+7.2}_{-7.0}$	$0.72^{+0.28}_{-0.22}$	$2.96^{+0.26}_{-0.32}$	$1.50^{+0.14}_{-0.16}$	$0.67^{+0.13}_{-0.11}$	$17.0^{+8.5}_{-5.7}$	12.37	12.02	11.54
250764453016220800	0.347 ± 0.031	2650^{+240}_{-200}	$3.35^{+0.11}_{-0.09}$	$-2.54^{+0.27}_{-0.30}$	$6.46^{+0.68}_{-0.55}$	$2.96^{+0.17}_{-0.14}$	$0.213^{+0.086}_{-0.065}$	13.69	12.94	12.08

Table C.3 continued.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
272076905372370816	0.176 ± 0.025	4760^{+630}_{-510}	$3.47^{+0.10}_{-0.07}$	$-3.19^{+0.32}_{-0.37}$	$10.7^{+5.4}_{-1.5}$	$4.04^{+0.18}_{-0.21}$	$0.084^{+0.042}_{-0.028}$	14.24	13.67	12.95
258477905041318144	0.415 ± 0.040	2250^{+230}_{-190}	$2.69^{+0.14}_{-0.15}$	$-1.54^{+0.34}_{-0.35}$	$4.71^{+0.66}_{-0.54}$	$2.53^{+0.18}_{-0.16}$	$0.57^{+0.28}_{-0.21}$	13.52	12.91	12.18
205118464010485632	0.335 ± 0.048	2750^{+440}_{-330}	$2.062^{+0.077}_{-0.077}$	$-1.92^{+0.36}_{-0.40}$	$6.6^{+1.1}_{-1.0}$	$3.36^{+0.18}_{-0.23}$	$0.30^{+0.15}_{-0.10}$	12.56	12.34	11.88
201374425057424512	0.883 ± 0.028	1098^{+35}_{-33}	$1.21^{+0.20}_{-0.28}$	$1.91^{+0.35}_{-0.27}$	$1.70^{+0.18}_{-0.12}$	$1.09^{+0.11}_{-0.14}$	$8.9^{+3.1}_{-2.4}$	13.71	13.32	12.78
181458215025292160	0.708 ± 0.042	1360^{+85}_{-76}	$1.84^{+0.13}_{-0.11}$	$-0.42^{+0.23}_{-0.26}$	$3.26^{+0.34}_{-0.25}$	$2.19^{+0.18}_{-0.15}$	$1.81^{+0.38}_{-0.41}$	12.41	12.09	11.58
3349361015210583168	1.480 ± 0.045	663^{+21}_{-20}	$0.93^{+0.30}_{-0.36}$	$2.42^{+0.43}_{-0.37}$	$1.80^{+0.70}_{-0.39}$	$0.88^{+0.16}_{-0.17}$	$5.9^{+3.0}_{-3.0}$	13.00	12.46	11.81
3348898739291597568	1.463 ± 0.055	672^{+26}_{-24}	$0.88^{+0.52}_{-0.52}$	$5.44^{+0.60}_{-0.60}$	$1.05^{+0.17}_{-0.31}$	$-0.24^{+0.18}_{-0.28}$	$10.9^{+5.5}_{-5.5}$	16.31	15.45	14.56
3123854268434443648	0.311 ± 0.040	2930^{+400}_{-310}	$2.52^{+0.30}_{-0.36}$	$-2.21^{+0.61}_{-0.58}$	$6.8^{+1.0}_{-1.1}$	$2.73^{+0.18}_{-0.25}$	$0.100^{+0.050}_{-0.048}$	13.64	12.64	11.67
3131337377071123200	0.541 ± 0.051	1760^{+180}_{-150}	$1.60^{+0.19}_{-0.17}$	$-0.48^{+0.37}_{-0.40}$	$3.42^{+0.61}_{-0.45}$	$2.06^{+0.18}_{-0.15}$	$1.40^{+0.70}_{-0.56}$	12.76	12.35	11.82
3049153761980978560	0.758 ± 0.022	1271^{+37}_{-35}	$1.00^{+0.09}_{-0.10}$	$1.53^{+0.16}_{-0.16}$	$2.06^{+0.19}_{-0.10}$	$1.324^{+0.094}_{-0.082}$	$7.7^{+3.8}_{-2.3}$	13.25	13.05	12.74
3044960396431590784	0.899 ± 0.037	1080^{+45}_{-41}	$0.93^{+0.20}_{-0.28}$	$1.61^{+0.37}_{-0.29}$	$1.82^{+0.25}_{-0.15}$	$1.21^{+0.12}_{-0.15}$	$7.4^{+1.8}_{-2.4}$	13.05	12.71	12.25
3045350894857953664	0.773 ± 0.033	1249^{+54}_{-50}	$1.11^{+0.20}_{-0.25}$	$1.60^{+0.33}_{-0.29}$	$1.82^{+0.20}_{-0.13}$	$1.22^{+0.13}_{-0.14}$	$7.5^{+1.5}_{-2.0}$	13.55	13.19	12.71
3045461567573658368	0.805 ± 0.035	1201^{+54}_{-49}	$1.32^{+0.27}_{-0.25}$	$1.10^{+0.34}_{-0.36}$	$2.31^{+0.49}_{-0.35}$	$1.41^{+0.14}_{-0.13}$	$3.6^{+1.8}_{-1.6}$	13.35	12.82	12.17
5546155247158673792	0.231 ± 0.019	3840^{+300}_{-260}	$2.65^{+0.13}_{-0.11}$	$-2.00^{+0.26}_{-0.30}$	$5.41^{+0.59}_{-0.44}$	$2.81^{+0.18}_{-0.16}$	$0.40^{+0.13}_{-0.12}$	14.07	13.56	12.86
5544618168572613504	4.21 ± 0.22	237^{+14}_{-12}	$1.21^{+0.52}_{-0.52}$	$5.03^{+0.64}_{-0.64}$	$1.15^{+0.22}_{-0.44}$	$-0.07^{+0.18}_{-0.30}$	$6.0^{+3.0}_{-3.0}$	14.07	13.11	12.16
5543776728634411008	2.107 ± 0.028	$468.2^{+6.3}_{-6.2}$	$0.32^{+0.22}_{-0.30}$	$3.78^{+0.33}_{-0.25}$	$1.25^{+0.11}_{-0.12}$	$0.34^{+0.10}_{-0.14}$	42^{+21}_{-21}	12.77	12.45	12.02
5524619735681532160	1.030 ± 0.023	945^{+21}_{-20}	$1.622^{+0.049}_{-0.049}$	$1.18^{+0.10}_{-0.10}$	$2.35^{+0.12}_{-0.12}$	$1.497^{+0.058}_{-0.053}$	$5.2^{+1.1}_{-0.9}$	12.97	12.68	12.20
5328760495308550400	2.974 ± 0.019	$333.0^{+2.2}_{-2.1}$	$0.38^{+0.61}_{-0.61}$	$5.01^{+0.63}_{-0.63}$	$1.15^{+0.22}_{-0.49}$	$-0.06^{+0.18}_{-0.30}$	$5.7^{+2.9}_{-2.9}$	13.72	13.00	12.21
5312087505304128000	0.650 ± 0.017	1475^{+39}_{-37}	$0.64^{+0.31}_{-0.52}$	$2.34^{+0.58}_{-0.36}$	$2.2^{+0.5}_{-1.1}$	$0.95^{+0.18}_{-0.25}$	$0.79^{+0.40}_{-0.40}$	14.48	13.82	13.11
5258744904832986880	0.442 ± 0.018	2124^{+87}_{-80}	$1.202^{+0.049}_{-0.049}$	$0.82^{+0.13}_{-0.14}$	$2.32^{+0.21}_{-0.09}$	$1.636^{+0.063}_{-0.072}$	$4.48^{+0.42}_{-0.59}$	13.88	13.66	13.31
5350683490177538944	0.326 ± 0.018	2810^{+150}_{-140}	$1.55^{+0.09}_{-0.10}$	$-0.47^{+0.21}_{-0.21}$	$3.29^{+0.26}_{-0.21}$	$2.11^{+0.11}_{-0.10}$	$1.65^{+0.38}_{-0.36}$	13.60	13.32	12.82
5254137160888722816	0.193 ± 0.023	4490^{+520}_{-430}	$2.07^{+0.10}_{-0.07}$	$-2.42^{+0.29}_{-0.34}$	$8.8^{+2.6}_{-1.2}$	$3.74^{+0.18}_{-0.18}$	$0.149^{+0.057}_{-0.059}$	13.12	12.91	12.52

Table C.3 continued.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
5338434587073245056	0.374 ± 0.020	2480^{+130}_{-120}	$1.11^{+0.09}_{-0.10}$	$0.95^{+0.21}_{-0.21}$	$2.20^{+0.27}_{-0.12}$	$1.55^{+0.10}_{-0.09}$	$5.05^{+0.64}_{-0.76}$	14.25	14.04	13.69
5337981412138546432	0.491 ± 0.016	1926^{+63}_{-59}	$1.58^{+0.20}_{-0.28}$	$0.22^{+0.35}_{-0.27}$	$2.86^{+0.44}_{-0.39}$	$1.76^{+0.11}_{-0.14}$	$2.2^{+1.1}_{-0.8}$	13.70	13.22	12.58
5338316767522194688	0.216 ± 0.017	4080^{+310}_{-270}	$2.24^{+0.10}_{-0.03}$	$-1.52^{+0.18}_{-0.26}$	$5.11^{+0.77}_{-0.32}$	$2.97^{+0.18}_{-0.09}$	$0.57^{+0.11}_{-0.17}$	14.10	13.77	13.25
5337719389118815488	0.338 ± 0.020	2730^{+160}_{-150}	$1.57^{+0.19}_{-0.36}$	$-0.56^{+0.48}_{-0.32}$	$3.43^{+0.60}_{-0.52}$	$2.12^{+0.18}_{-0.22}$	$1.44^{+0.72}_{-0.60}$	13.55	13.19	12.72
5337668772877422464	0.424 ± 0.024	2210^{+130}_{-120}	$0.95^{+0.20}_{-0.28}$	$0.68^{+0.40}_{-0.32}$	$2.44^{+0.44}_{-0.35}$	$1.58^{+0.13}_{-0.16}$	$3.4^{+1.7}_{-1.3}$	13.69	13.36	12.88
5337250722275063936	0.391 ± 0.019	2380^{+110}_{-100}	$1.14^{+0.13}_{-0.11}$	$0.07^{+0.20}_{-0.23}$	$2.80^{+0.31}_{-0.19}$	$1.98^{+0.18}_{-0.13}$	$2.72^{+0.48}_{-0.56}$	13.27	13.10	12.79
5333514409948908288	0.361 ± 0.023	2570^{+160}_{-150}	$3.33^{+0.15}_{-0.04}$	$-2.52^{+0.17}_{-0.29}$	$7.4^{+1.4}_{-0.6}$	$3.51^{+0.18}_{-0.12}$	$0.201^{+0.046}_{-0.068}$	13.42	12.87	12.09
5333596770242078976	0.408 ± 0.018	2290^{+100}_{-90}	$1.27^{+0.05}_{-0.13}$	$0.13^{+0.22}_{-0.14}$	$3.04^{+0.41}_{-0.41}$	$2.08^{+0.10}_{-0.19}$	$2.39^{+0.87}_{-0.49}$	13.37	13.20	12.89
5333977613578341248	0.421 ± 0.036	2230^{+200}_{-170}	$1.40^{+0.19}_{-0.36}$	$2.22^{+0.53}_{-0.38}$	$1.77^{+0.22}_{-0.12}$	$1.01^{+0.18}_{-0.13}$	$15.0^{+7.5}_{-7.5}$	15.67	15.37	14.92
5333545642950621696	0.406 ± 0.018	2300^{+100}_{-90}	-	-	-	-	-	13.74	13.41	12.68
5333575329766038144	0.397 ± 0.019	2350^{+110}_{-100}	$1.38^{+0.11}_{-0.09}$	$0.64^{+0.19}_{-0.21}$	$2.37^{+0.34}_{-0.13}$	$1.71^{+0.13}_{-0.10}$	$4.16^{+0.60}_{-0.75}$	14.12	13.88	13.47
5334867393402370816	0.232 ± 0.023	3820^{+370}_{-310}	$2.80^{+0.36}_{-0.31}$	$-2.66^{+0.49}_{-0.56}$	$7.6^{+1.2}_{-0.9}$	$2.93^{+0.18}_{-0.21}$	$0.057^{+0.028}_{-0.027}$	14.26	13.05	11.97
6054820460425919616	0.155 ± 0.018	5340^{+580}_{-480}	$3.77^{+0.22}_{-0.34}$	$-4.01^{+0.54}_{-0.44}$	$10.8^{+2.8}_{-2.0}$	$3.84^{+0.18}_{-0.30}$	$0.057^{+0.028}_{-0.025}$	14.12	13.40	12.46
5858981977697486592	0.379 ± 0.022	2450^{+150}_{-130}	$0.21^{+0.20}_{-0.25}$	$1.77^{+0.37}_{-0.33}$	$1.76^{+0.21}_{-0.12}$	$1.16^{+0.14}_{-0.15}$	$8.3^{+3.5}_{-2.1}$	14.09	13.93	13.69
5868283055734381184	0.471 ± 0.022	2002^{+92}_{-84}	$1.45^{+0.11}_{-0.09}$	$0.27^{+0.19}_{-0.21}$	$2.61^{+0.19}_{-0.15}$	$1.86^{+0.13}_{-0.11}$	$3.19^{+0.48}_{-0.51}$	13.48	13.23	12.80
5868479043676144384	0.438 ± 0.022	2140^{+110}_{-100}	$1.69^{+0.13}_{-0.30}$	$0.07^{+0.40}_{-0.24}$	$2.81^{+0.40}_{-0.36}$	$1.99^{+0.18}_{-0.27}$	$2.7^{+1.0}_{-0.6}$	13.70	13.41	12.95
5868455988315314560	0.329 ± 0.021	2790^{+170}_{-160}	$1.84^{+0.36}_{-0.31}$	$-0.42^{+0.43}_{-0.49}$	$4.8^{+0.9}_{-1.0}$	$2.03^{+0.18}_{-0.19}$	$0.26^{+0.13}_{-0.13}$	14.55	13.65	12.74
5866357054949304832	0.303 ± 0.038	3000^{+400}_{-320}	$4.758^{+0.067}_{-0.037}$	$-4.30^{+0.28}_{-0.34}$	$11.9^{+1.6}_{-1.2}$	$3.97^{+0.18}_{-0.15}$	$0.043^{+0.013}_{-0.011}$	13.97	12.85	11.84
5866510505529424896	0.559 ± 0.019	1702^{+56}_{-53}	$1.490^{+0.049}_{-0.049}$	$0.43^{+0.12}_{-0.12}$	$2.491^{+0.092}_{-0.071}$	$1.780^{+0.063}_{-0.041}$	$3.60^{+0.31}_{-0.33}$	13.35	13.07	12.65
5854532425868957184	0.569 ± 0.024	1673^{+69}_{-64}	$2.02^{+0.11}_{-0.09}$	$0.77^{+0.18}_{-0.20}$	$2.36^{+0.35}_{-0.16}$	$1.66^{+0.12}_{-0.11}$	$4.33^{+0.71}_{-0.92}$	14.33	13.92	13.37
5884829984147822976	2.400 ± 0.026	$411.8^{+4.4}_{-4.3}$	$0.55^{+0.36}_{-0.31}$	$4.79^{+0.33}_{-0.39}$	$0.98^{+0.29}_{-0.06}$	$-0.06^{+0.17}_{-0.13}$	28^{+14}_{-14}	13.95	13.42	12.77
5935057668182928256	0.480 ± 0.040	1970^{+170}_{-150}	$2.21^{+0.25}_{-0.10}$	$-1.24^{+0.27}_{-0.44}$	$4.7^{+2.4}_{-0.5}$	$2.86^{+0.18}_{-0.20}$	$0.73^{+0.37}_{-0.37}$	12.80	12.45	11.96

Table C.3 continued.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
5941445658948980864	1.479 ± 0.026	663^{+12}_{-12}	$0.50^{+0.16}_{-0.08}$	$3.75^{+0.12}_{-0.20}$	$1.18^{+0.10}_{-0.03}$	$0.349^{+0.080}_{-0.046}$	$18.9^{+3.0}_{-4.4}$	13.80	13.36	12.82
5940955031917319040	0.874 ± 0.052	1112^{+69}_{-62}	$1.80^{+0.09}_{-0.10}$	$0.70^{+0.22}_{-0.23}$	$2.30^{+0.16}_{-0.14}$	$1.65^{+0.12}_{-0.11}$	$4.40^{+0.70}_{-0.71}$	13.09	12.74	12.18
5969536355649568640	0.635 ± 0.030	1509^{+72}_{-66}	$1.52^{+0.06}_{-0.10}$	$0.89^{+0.20}_{-0.16}$	$2.49^{+0.22}_{-0.26}$	$1.64^{+0.09}_{-0.12}$	$4.0^{+1.4}_{-0.6}$	13.59	13.30	12.90
5966697730168623488	0.595 ± 0.028	1605^{+75}_{-69}	$1.30^{+0.11}_{-0.09}$	$1.06^{+0.19}_{-0.21}$	$2.44^{+0.23}_{-0.27}$	$1.55^{+0.13}_{-0.11}$	$4.3^{+2.0}_{-0.7}$	13.61	13.39	13.00
5972078735771910784	5.358 ± 0.039	$185.7^{+1.4}_{-1.3}$	$1.61^{+0.52}_{-0.52}$	$4.86^{+0.54}_{-0.54}$	$1.16^{+0.23}_{-0.45}$	$-0.00^{+0.18}_{-0.27}$	$4.3^{+2.1}_{-2.1}$	13.85	12.81	11.74
4054609866596686336	0.843 ± 0.037	1149^{+51}_{-47}	$1.50^{+0.17}_{-0.20}$	$1.70^{+0.29}_{-0.27}$	$1.80^{+0.20}_{-0.11}$	$1.19^{+0.11}_{-0.12}$	$8.0^{+2.0}_{-1.4}$	13.91	13.50	12.93
4054618520934162944	0.888 ± 0.058	1097^{+76}_{-67}	$1.28^{+0.20}_{-0.25}$	$0.95^{+0.38}_{-0.35}$	$2.20^{+0.37}_{-0.26}$	$1.48^{+0.15}_{-0.16}$	$4.5^{+2.0}_{-1.6}$	12.83	12.44	11.92
4094703381988286592	0.314 ± 0.048	2920^{+500}_{-380}	$3.21^{+0.07}_{-0.15}$	$-3.09^{+0.45}_{-0.42}$	$9.8^{+1.9}_{-2.2}$	$3.92^{+0.18}_{-0.30}$	$0.100^{+0.050}_{-0.035}$	12.99	12.45	11.79
4097464668003385216	2.125 ± 0.036	$464.4^{+8.1}_{-7.8}$	$1.70^{+0.30}_{-0.34}$	$2.88^{+0.38}_{-0.34}$	$1.69^{+0.56}_{-0.38}$	$0.70^{+0.15}_{-0.15}$	$6.5^{+3.2}_{-3.2}$	13.70	12.91	12.06
4145745975926657664	0.642 ± 0.035	1494^{+84}_{-76}	$1.39^{+0.10}_{-0.04}$	$1.57^{+0.15}_{-0.22}$	$1.87^{+0.13}_{-0.09}$	$1.224^{+0.088}_{-0.057}$	$6.7^{+1.1}_{-1.1}$	14.30	13.84	13.22
4090478474203479296	0.253 ± 0.041	3540^{+640}_{-480}	$3.451^{+0.027}_{-0.057}$	$-1.64^{+0.37}_{-0.39}$	$5.96^{+0.49}_{-0.56}$	$2.50^{+0.14}_{-0.15}$	$0.162^{+0.077}_{-0.037}$	15.82	14.56	13.38
4097686082159430784	2.206 ± 0.064	448^{+13}_{-13}	$1.04^{+0.36}_{-0.31}$	$3.89^{+0.37}_{-0.43}$	$1.35^{+0.40}_{-0.28}$	$0.31^{+0.18}_{-0.16}$	$10.4^{+5.2}_{-5.2}$	13.87	13.19	12.44
4097384579759711744	2.656 ± 0.040	$372.6^{+5.6}_{-5.5}$	$0.69^{+0.30}_{-0.36}$	$4.26^{+0.39}_{-0.34}$	$1.10^{+0.23}_{-0.10}$	$0.15^{+0.14}_{-0.16}$	26^{+13}_{-13}	13.27	12.80	12.20
4153689436387288448	0.441 ± 0.042	2140^{+220}_{-180}	$4.49^{+0.07}_{-0.15}$	$-2.98^{+0.35}_{-0.28}$	$9.4^{+1.4}_{-1.7}$	$3.85^{+0.18}_{-0.30}$	$0.113^{+0.056}_{-0.031}$	14.07	13.16	12.24
4152422554127130240	0.528 ± 0.071	1820^{+280}_{-220}	$3.83^{+0.49}_{-0.34}$	$-3.01^{+0.61}_{-0.81}$	$7.9^{+3.9}_{-1.7}$	$3.49^{+0.18}_{-0.30}$	$0.148^{+0.074}_{-0.074}$	12.93	12.13	11.27
4152405172406857088	0.438 ± 0.067	2170^{+400}_{-290}	$5.78^{+0.49}_{-0.34}$	$-3.77^{+0.65}_{-0.86}$	$10.1^{+5.0}_{-2.3}$	$3.79^{+0.18}_{-0.30}$	$0.071^{+0.036}_{-0.036}$	15.14	13.69	12.53
4098138462472816384	0.444 ± 0.044	2130^{+230}_{-190}	$4.86^{+0.49}_{-0.34}$	$-3.72^{+0.54}_{-0.71}$	$9.9^{+5.0}_{-2.0}$	$3.78^{+0.18}_{-0.30}$	$0.074^{+0.037}_{-0.037}$	13.89	12.77	11.73
4272195138879459200	0.755 ± 0.075	1290^{+140}_{-120}	$5.29^{+0.49}_{-0.34}$	$-2.84^{+0.55}_{-0.72}$	$7.5^{+3.7}_{-1.5}$	$3.42^{+0.18}_{-0.30}$	$0.173^{+0.087}_{-0.087}$	14.23	13.00	11.86
4155634296310906112	0.266 ± 0.042	3380^{+580}_{-440}	$2.90^{+0.06}_{-0.10}$	$-1.69^{+0.40}_{-0.40}$	$4.90^{+0.71}_{-0.60}$	$2.64^{+0.18}_{-0.19}$	$0.53^{+0.26}_{-0.19}$	14.46	13.85	13.09
4256499816550576256	0.284 ± 0.037	3190^{+440}_{-350}	$3.10^{+0.49}_{-0.34}$	$-2.68^{+0.59}_{-0.78}$	$7.1^{+3.5}_{-1.5}$	$3.35^{+0.18}_{-0.30}$	$0.20^{+0.10}_{-0.10}$	13.52	12.94	12.22
4256500851683628544	0.286 ± 0.028	3170^{+320}_{-270}	$3.62^{+0.49}_{-0.34}$	$-2.80^{+0.53}_{-0.71}$	$7.3^{+3.7}_{-1.4}$	$3.39^{+0.18}_{-0.30}$	$0.182^{+0.091}_{-0.091}$	14.04	13.33	12.49
4253225574005033088	4.988 ± 0.038	$199.3^{+1.5}_{-1.5}$	$0.19^{+0.52}_{-0.52}$	$6.07^{+0.54}_{-0.54}$	$0.85^{+0.18}_{-0.14}$	$-0.49^{+0.18}_{-0.22}$	27^{+13}_{-13}	13.44	12.76	12.02

Table C.3 continued.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
4259271891523989376	0.316 ± 0.048	2880^{+480}_{-360}	$3.316^{+0.032}_{-0.046}$	$-3.60^{+0.34}_{-0.36}$	$9.2^{+1.2}_{-1.0}$	$3.61^{+0.17}_{-0.17}$	$0.084^{+0.033}_{-0.026}$	12.68	12.02	11.23
4265423830878562944	0.383 ± 0.042	2430^{+290}_{-230}	$3.10^{+0.09}_{-0.11}$	$-0.23^{+0.33}_{-0.33}$	$3.26^{+0.43}_{-0.38}$	$1.94^{+0.14}_{-0.13}$	$1.54^{+0.65}_{-0.48}$	15.67	14.80	13.86
4267078153499687296	1.723 ± 0.060	572^{+20}_{-19}	$2.235^{+0.027}_{-0.057}$	$2.22^{+0.13}_{-0.10}$	$1.95^{+0.13}_{-0.11}$	$0.965^{+0.041}_{-0.053}$	$4.77^{+0.88}_{-0.84}$	14.19	13.24	12.35
2019217237173395968	1.874 ± 0.019	$525.7^{+5.3}_{-5.2}$	$1.64^{+0.30}_{-0.36}$	$3.04^{+0.38}_{-0.33}$	$1.47^{+0.58}_{-0.22}$	$0.64^{+0.14}_{-0.15}$	$10.2^{+5.1}_{-5.1}$	14.02	13.29	12.48
4315228344710978816	0.467 ± 0.024	2020^{+110}_{-100}	$0.76^{+0.09}_{-0.11}$	$1.58^{+0.21}_{-0.20}$	$1.83^{+0.12}_{-0.08}$	$1.227^{+0.081}_{-0.083}$	$7.4^{+0.9}_{-1.3}$	14.16	13.87	13.50
4318785810234714752	0.226 ± 0.037	3870^{+650}_{-490}	-	-	-	-	-	12.45	12.30	11.25
2020080113265080832	0.426 ± 0.026	2200^{+130}_{-120}	$2.705^{+0.035}_{-0.085}$	$-0.03^{+0.21}_{-0.16}$	$3.49^{+0.20}_{-0.24}$	$1.856^{+0.064}_{-0.081}$	$1.12^{+0.27}_{-0.17}$	15.37	14.39	13.46
2020089630912890368	0.405 ± 0.022	2310^{+130}_{-120}	$2.520^{+0.032}_{-0.046}$	$-0.88^{+0.16}_{-0.15}$	$3.90^{+0.20}_{-0.21}$	$2.519^{+0.081}_{-0.091}$	$1.17^{+0.18}_{-0.15}$	13.91	13.45	12.84
2032083584842678144	0.227 ± 0.021	3910^{+360}_{-300}	$1.149^{+0.027}_{-0.057}$	$-0.10^{+0.23}_{-0.22}$	$3.96^{+0.24}_{-0.24}$	$1.891^{+0.086}_{-0.090}$	$0.66^{+0.14}_{-0.13}$	14.61	14.01	13.30
2027980718477508480	1.181 ± 0.025	827^{+17}_{-17}	$2.12^{+0.11}_{-0.14}$	$1.14^{+0.18}_{-0.15}$	$2.04^{+0.11}_{-0.09}$	$1.437^{+0.071}_{-0.085}$	$5.88^{+0.65}_{-0.63}$	13.37	12.85	12.24
1828276425855506304	0.437 ± 0.031	2150^{+160}_{-140}	$1.81^{+0.31}_{-0.52}$	$-0.94^{+0.66}_{-0.46}$	$5.6^{+0.7}_{-2.8}$	$2.26^{+0.18}_{-0.29}$	$0.077^{+0.039}_{-0.039}$	13.52	12.53	11.57
2031912537697905536	0.385 ± 0.028	2420^{+180}_{-160}	$2.400^{+0.027}_{-0.057}$	$-1.97^{+0.20}_{-0.18}$	$6.43^{+0.29}_{-0.28}$	$2.634^{+0.073}_{-0.079}$	$0.126^{+0.019}_{-0.021}$	13.33	12.35	11.40
2035402872974695936	0.113 ± 0.020	6900^{+1200}_{-900}	$2.197^{+0.069}_{-0.055}$	$-3.43^{+0.35}_{-0.40}$	$8.91^{+0.87}_{-0.70}$	$3.21^{+0.16}_{-0.14}$	$0.046^{+0.017}_{-0.013}$	13.72	12.97	12.14
2071705173505640448	0.483 ± 0.088	2030^{+510}_{-340}	-	-	-	-	-	13.73	11.32	9.96
1823174039068377216	1.114 ± 0.021	875^{+16}_{-16}	$0.45^{+0.28}_{-0.22}$	$2.88^{+0.26}_{-0.32}$	$1.55^{+0.13}_{-0.21}$	$0.70^{+0.13}_{-0.11}$	$12.9^{+6.4}_{-2.1}$	13.32	13.04	12.62
1834095999134055680	0.220 ± 0.027	3970^{+490}_{-400}	$2.591^{+0.067}_{-0.037}$	$-3.42^{+0.27}_{-0.32}$	$8.9^{+1.1}_{-0.8}$	$3.61^{+0.18}_{-0.14}$	$0.100^{+0.030}_{-0.027}$	12.63	12.17	11.56
2059037012843484032	0.551 ± 0.019	1727^{+59}_{-55}	$1.66^{+0.28}_{-0.22}$	$0.64^{+0.30}_{-0.36}$	$2.64^{+0.49}_{-0.37}$	$1.59^{+0.15}_{-0.12}$	$2.6^{+1.3}_{-1.0}$	14.07	13.49	12.78
1836558703328498944	0.581 ± 0.030	1643^{+88}_{-79}	$4.30^{+0.04}_{-0.14}$	$-1.80^{+0.25}_{-0.15}$	$5.71^{+0.39}_{-0.38}$	$2.56^{+0.06}_{-0.10}$	$0.249^{+0.062}_{-0.062}$	14.93	13.59	12.41
2062526622213061760	0.794 ± 0.038	1218^{+59}_{-54}	$1.22^{+0.19}_{-0.17}$	$1.12^{+0.27}_{-0.29}$	$2.05^{+0.20}_{-0.14}$	$1.43^{+0.13}_{-0.12}$	$5.7^{+1.1}_{-1.4}$	13.09	12.77	12.33
2061354783350172672	0.186 ± 0.024	4540^{+560}_{-450}	$3.16^{+0.49}_{-0.34}$	$-4.00^{+0.57}_{-0.75}$	$10.8^{+5.4}_{-2.2}$	$3.88^{+0.18}_{-0.30}$	$0.057^{+0.029}_{-0.029}$	13.06	12.45	11.72
2068318948239115008	0.928 ± 0.029	1046^{+32}_{-31}	$0.84^{+0.20}_{-0.28}$	$1.62^{+0.35}_{-0.26}$	$1.82^{+0.23}_{-0.14}$	$1.21^{+0.11}_{-0.14}$	$7.4^{+1.7}_{-2.3}$	12.88	12.56	12.13
2067485110409323648	1.266 ± 0.016	772^{+10}_{-10}	$0.38^{+0.36}_{-0.31}$	$3.68^{+0.33}_{-0.39}$	$1.46^{+0.38}_{-0.34}$	$0.39^{+0.17}_{-0.14}$	$8.2^{+4.1}_{-4.1}$	13.99	13.50	12.89

Table C.3 continued.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
2066412811690449792	0.662 ± 0.030	1447^{+65}_{-60}	$5.12^{+0.49}_{-0.34}$	$-2.27^{+0.43}_{-0.59}$	$6.2^{+3.1}_{-1.1}$	$3.18^{+0.18}_{-0.30}$	$0.30^{+0.15}_{-0.15}$	14.85	13.65	12.56
2066439165609688192	0.723 ± 0.025	1331^{+46}_{-43}	$4.06^{+0.10}_{-0.09}$	$-0.41^{+0.16}_{-0.17}$	$3.23^{+0.20}_{-0.17}$	$2.10^{+0.11}_{-0.08}$	$1.76^{+0.31}_{-0.30}$	15.26	14.27	13.28
2067030702869079424	1.284 ± 0.038	763^{+23}_{-22}	$2.393^{+0.032}_{-0.046}$	$0.53^{+0.11}_{-0.10}$	$2.91^{+0.13}_{-0.12}$	$1.965^{+0.060}_{-0.082}$	$3.27^{+0.36}_{-0.47}$	12.77	12.34	11.77
2066784304885540992	0.252 ± 0.034	3450^{+460}_{-370}	$5.11^{+0.49}_{-0.34}$	$-5.22^{+0.58}_{-0.76}$	$16.4^{+8.2}_{-3.9}$	$4.37^{+0.18}_{-0.30}$	$0.020^{+0.010}_{-0.010}$	13.77	12.58	11.49
2166981120209304576	1.482 ± 0.021	$662.2^{+9.3}_{-9.1}$	$2.37^{+0.45}_{-0.36}$	$1.47^{+0.39}_{-0.48}$	$1.88^{+0.50}_{-0.15}$	$1.28^{+0.18}_{-0.17}$	$7.1^{+3.6}_{-2.9}$	13.56	12.94	12.19
1969788039090648832	0.337 ± 0.020	2730^{+160}_{-150}	$1.91^{+0.06}_{-0.10}$	$-0.69^{+0.22}_{-0.18}$	$3.53^{+0.22}_{-0.24}$	$2.25^{+0.11}_{-0.12}$	$1.41^{+0.32}_{-0.26}$	13.77	13.40	12.89
2164505844663760768	0.430 ± 0.030	2180^{+150}_{-140}	$2.49^{+0.49}_{-0.34}$	$-1.89^{+0.48}_{-0.64}$	$5.5^{+2.8}_{-1.1}$	$3.04^{+0.18}_{-0.30}$	$0.43^{+0.22}_{-0.22}$	12.71	12.29	11.70
2164554394961279360	1.287 ± 0.021	760^{+12}_{-12}	$1.17^{+0.28}_{-0.22}$	$2.34^{+0.26}_{-0.32}$	$1.55^{+0.20}_{-0.08}$	$0.92^{+0.13}_{-0.10}$	$10.9^{+5.3}_{-3.5}$	13.37	12.91	12.32
1971703422709930496	1.413 ± 0.023	694^{+11}_{-11}	$0.87^{+0.30}_{-0.34}$	$3.54^{+0.37}_{-0.33}$	$1.35^{+0.48}_{-0.20}$	$0.44^{+0.14}_{-0.15}$	$12.0^{+6.0}_{-6.0}$	14.22	13.61	12.99
1972408072220900864	1.382 ± 0.028	709^{+15}_{-14}	$1.67^{+0.30}_{-0.36}$	$1.89^{+0.41}_{-0.35}$	$2.10^{+0.86}_{-0.46}$	$1.10^{+0.15}_{-0.16}$	$4.0^{+2.0}_{-2.0}$	13.56	12.81	12.00
2178435351312223488	1.030 ± 0.015	945^{+14}_{-14}	$1.95^{+0.06}_{-0.10}$	$1.35^{+0.13}_{-0.09}$	$2.17^{+0.12}_{-0.09}$	$1.451^{+0.063}_{-0.089}$	$6.9^{+3.5}_{-1.2}$	13.56	13.18	12.66
1978002902777895552	1.167 ± 0.033	837^{+24}_{-22}	$1.43^{+0.20}_{-0.28}$	$1.19^{+0.34}_{-0.26}$	$2.08^{+0.29}_{-0.23}$	$1.38^{+0.11}_{-0.14}$	$5.1^{+2.1}_{-1.6}$	12.69	12.23	11.65
2173413297587679744	0.636 ± 0.028	1506^{+67}_{-62}	$0.89^{+0.27}_{-0.25}$	$0.64^{+0.34}_{-0.36}$	$2.70^{+0.57}_{-0.40}$	$1.59^{+0.15}_{-0.13}$	$2.4^{+1.2}_{-1.1}$	12.85	12.42	11.87
1981413759654971008	0.190 ± 0.030	4420^{+670}_{-520}	$3.50^{+0.49}_{-0.34}$	$-4.44^{+0.61}_{-0.80}$	$12.6^{+6.3}_{-2.8}$	$4.06^{+0.18}_{-0.30}$	$0.039^{+0.019}_{-0.019}$	12.81	12.28	11.32
1980664098879084288	0.211 ± 0.031	4050^{+570}_{-450}	$2.50^{+0.49}_{-0.34}$	$-2.97^{+0.59}_{-0.78}$	$7.8^{+3.9}_{-1.7}$	$3.47^{+0.18}_{-0.30}$	$0.153^{+0.076}_{-0.076}$	12.96	12.56	11.95
2198056166359436160	0.270 ± 0.034	3330^{+430}_{-340}	$3.91^{+0.05}_{-0.10}$	$-4.07^{+0.34}_{-0.31}$	$14.5^{+7.2}_{-1.6}$	$4.50^{+0.18}_{-0.23}$	$1.31^{+0.66}_{-0.66}$	13.11	12.45	11.66
2203217037728935296	1.151 ± 0.032	849^{+24}_{-23}	$1.22^{+0.11}_{-0.09}$	$1.57^{+0.15}_{-0.17}$	$2.06^{+0.22}_{-0.10}$	$1.34^{+0.13}_{-0.09}$	$8.3^{+4.2}_{-1.7}$	12.67	12.44	12.09
2005453516247868416	0.363 ± 0.031	2550^{+220}_{-190}	$0.78^{+0.11}_{-0.09}$	$-0.11^{+0.26}_{-0.29}$	$2.93^{+0.30}_{-0.23}$	$2.00^{+0.18}_{-0.13}$	$2.33^{+0.58}_{-0.54}$	12.83	12.70	12.48
2199663691020796928	0.381 ± 0.025	2440^{+160}_{-140}	$2.33^{+0.22}_{-0.30}$	$-1.65^{+0.44}_{-0.36}$	$5.80^{+0.69}_{-0.82}$	$2.50^{+0.15}_{-0.17}$	$0.21^{+0.10}_{-0.08}$	13.46	12.61	11.73
2004435162317441280	0.245 ± 0.019	3630^{+270}_{-240}	$0.88^{+0.06}_{-0.10}$	$-0.51^{+0.25}_{-0.22}$	$3.33^{+0.24}_{-0.25}$	$2.19^{+0.10}_{-0.14}$	$1.67^{+0.40}_{-0.33}$	13.31	13.17	12.93
2204517656901678848	1.037 ± 0.021	939^{+19}_{-18}	$1.917^{+0.027}_{-0.057}$	$2.20^{+0.10}_{-0.07}$	$1.95^{+0.10}_{-0.09}$	$0.970^{+0.028}_{-0.036}$	$4.78^{+0.66}_{-0.70}$	14.80	13.98	13.11
2200017424528999936	0.151 ± 0.019	5420^{+620}_{-510}	$3.96^{+0.49}_{-0.34}$	$-3.76^{+0.55}_{-0.73}$	$10.0^{+5.0}_{-2.0}$	$3.78^{+0.18}_{-0.30}$	$0.072^{+0.036}_{-0.036}$	14.69	13.87	12.96

Table C.3 continued.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
2208194286334441344	1.124 ± 0.027	868^{+21}_{-20}	$2.08^{+0.06}_{-0.11}$	$2.00^{+0.16}_{-0.11}$	$1.84^{+0.12}_{-0.10}$	$1.080^{+0.045}_{-0.070}$	$9.9^{+3.2}_{-2.4}$	13.58	13.77	12.38
2007423050801656960	0.207 ± 0.020	4180^{+370}_{-320}	$2.461^{+0.067}_{-0.037}$	$-2.24^{+0.21}_{-0.25}$	$6.10^{+0.61}_{-0.43}$	$3.14^{+0.15}_{-0.11}$	$0.315^{+0.074}_{-0.074}$	13.74	13.33	12.74
2007419820986293504	0.243 ± 0.030	3620^{+440}_{-360}	$2.45^{+0.07}_{-0.15}$	$-3.65^{+0.38}_{-0.32}$	$11.7^{+2.0}_{-2.4}$	$4.14^{+0.18}_{-0.30}$	$0.062^{+0.031}_{-0.016}$	11.92	11.59	11.10
2206875422143048960	1.272 ± 0.022	769^{+14}_{-13}	$2.52^{+0.10}_{-0.04}$	$2.27^{+0.07}_{-0.14}$	$1.62^{+0.10}_{-0.03}$	$0.951^{+0.055}_{-0.030}$	$10.39^{+0.39}_{-0.73}$	15.01	14.22	13.39
2013474586693428096	2.079 ± 0.029	$474.5^{+6.7}_{-6.5}$	$1.57^{+0.22}_{-0.30}$	$2.97^{+0.34}_{-0.26}$	$1.37^{+0.23}_{-0.08}$	$0.66^{+0.10}_{-0.13}$	$14.4^{+7.2}_{-6.2}$	13.29	12.92	11.93
2014636396825359488	0.231 ± 0.033	3770^{+550}_{-430}	$5.42^{+0.49}_{-0.34}$	$-4.02^{+0.60}_{-0.79}$	$10.9^{+5.4}_{-2.3}$	$3.88^{+0.18}_{-0.30}$	$0.056^{+0.028}_{-0.028}$	15.59	14.28	13.15
2013507915639151488	0.291 ± 0.020	3110^{+210}_{-190}	$3.16^{+0.15}_{-0.04}$	$-2.50^{+0.18}_{-0.30}$	$7.3^{+1.5}_{-0.6}$	$3.49^{+0.18}_{-0.13}$	$0.206^{+0.050}_{-0.073}$	13.68	13.13	12.43
2207681742123365120	1.239 ± 0.017	789^{+11}_{-10}	$1.92^{+0.19}_{-0.17}$	$1.67^{+0.20}_{-0.22}$	$1.91^{+0.27}_{-0.12}$	$1.21^{+0.11}_{-0.09}$	$7.3^{+3.1}_{-1.5}$	13.56	13.08	12.45
2208975695512792064	1.376 ± 0.016	$711.6^{+8.1}_{-7.9}$	$2.16^{+0.51}_{-0.67}$	$1.93^{+0.69}_{-0.53}$	$1.81^{+0.91}_{-0.35}$	$1.08^{+0.18}_{-0.28}$	$6.7^{+3.4}_{-3.4}$	14.09	13.35	12.44
2014090042628166656	0.313 ± 0.026	2920^{+240}_{-210}	$1.85^{+0.49}_{-0.34}$	$-2.05^{+0.50}_{-0.67}$	$5.8^{+2.9}_{-1.2}$	$3.10^{+0.18}_{-0.30}$	$0.37^{+0.19}_{-0.19}$	12.40	12.13	11.70
2013834955933319680	0.282 ± 0.016	3200^{+180}_{-160}	$3.70^{+0.49}_{-0.34}$	$-3.02^{+0.45}_{-0.61}$	$7.9^{+4.0}_{-1.4}$	$3.50^{+0.18}_{-0.30}$	$0.146^{+0.073}_{-0.073}$	13.93	13.21	12.33
2015682857022899840	0.324 ± 0.022	2830^{+190}_{-170}	$2.52^{+0.10}_{-0.09}$	$0.02^{+0.22}_{-0.24}$	$2.81^{+0.22}_{-0.19}$	$1.93^{+0.13}_{-0.11}$	$2.57^{+0.54}_{-0.51}$	15.33	14.80	14.10
2012261211191786624	0.223 ± 0.018	3940^{+310}_{-270}	$4.10^{+0.07}_{-0.15}$	$-3.51^{+0.31}_{-0.24}$	$11.1^{+1.6}_{-1.9}$	$4.07^{+0.18}_{-0.28}$	$0.070^{+0.035}_{-0.017}$	14.34	13.56	12.71
2016307791941936896	1.251 ± 0.018	781^{+11}_{-11}	$4.30^{+0.49}_{-0.34}$	$-0.61^{+0.37}_{-0.53}$	$3.9^{+1.9}_{-0.8}$	$2.52^{+0.18}_{-0.30}$	$1.28^{+0.64}_{-0.64}$	14.08	13.15	12.20

Notes: The first three and the last three columns were taken from the catalogue of Table 3.3. Distances from Bailer-Jones et al. (2018). See Chapter 4 for the derivation of the extinction, absolute magnitude and stellar parameters. A maximum of 50% relative error was set for the uncertainties of the mass, luminosity and age.

C.2 Observed classical Be candidates

Table C.4 Derived spectral types and effective temperatures for the 14 observed CBe candidates.

Gaia source id	RA h:m:s	DEC deg:m:s	Probability CBe	Flag	SIMBAD ObType/SpType	SpType	T_{eff} (K)	Comments
431592948519739136	00:01:45.6	+63:27:58	0.89	-	Em*/B	B1.5	24800 ± 3000	-
522471059195179008	01:11:12.2	+61:06:06	0.88	-	Em*/B5	B5	15700 ± 1000	-
506475222292587264	01:44:22.3	+59:05:26	0.88	-	Em*	B5	15700 ± 1000	-
509678889933494912	01:45:06.9	+60:39:30	0.87	-	Em*	B5	15700 ± 1000	-
469722985282680064	04:00:58.1	+57:22:48	0.87	-	-	B2	20600 ± 2500	-
253403143486587136	04:26:22.2	+44:00:31	0.89	-	Em*/OB	-	21000 ± 1000	-
252274593578484992	04:30:00.7	+42:23:00	0.94	-	Em*	-	21000 ± 1000	-
2058442688129432960	20:01:53.7	+33:57:49	0.90	-	Star	B	21000 ± 11000	-
2164054323342523520	21:11:00.9	+45:39:41	0.88	-	Be*/B1-3Ve	B2	20600 ± 2500	-
2165027631639736448	21:21:16.9	+49:01:47	0.89	W	-	B3	17000 ± 1000	-
1981502162946440832	21:59:25.2	+53:37:32	0.86	W	-	B5	15700 ± 1000	-
2006954899366779136	22:46:04.7	+56:27:12	0.90	-	-	B3	17000 ± 1000	-
2016138123559082624	23:52:20.4	+63:17:07	0.93	-	-	B2	20600 ± 2500	-
2013028635958122240	23:57:22.9	+62:48:15	0.91	W	Em*	B5	15700 ± 1000	-

Notes: The first five columns were taken from the catalogue of Table 3.4. The 'HAeBe', 'ID AllW/ID' and 'ID IPH/VPH', 'Var', and 'W3W4' warning flags of Chapter 3 have been abbreviated to H, ID, V, and W respectively. For several objects the temperature was derived directly from model fitting the spectra, and thus no spectral type was derived for those sources. 'SIMBAD' column lists the object types and spectral types tabulated at the SIMBAD database at the time of writing. 'Comments' column indicates the possible detected contaminants (see Sect. 4.5): 'Emiss' marks sources with a strong emission line spectrum. 'Abs' marks sources with the Hα line in absorption.'B[e]' marks B[e] (FS CMa) stars. 'Evolved' marks evolved stars. 'Cont' indicate other types of contaminants (see main text for further discussion).

Table C.5 Hydrogen line properties of the 14 observed CBe candidates.

Gaia source id	RA h:m:s	DEC deg:m:s	Instrument	Hβ	Hα EW_{obs} (Å)	Hα EW_{corr} (Å)	Hα line shape	Comment
431592948519739136	00:01:45.6	+63:27:58	IDS	Yes	-37.68	-38.6	s	-
522471059195179008	01:11:12.2	+61:06:06	IDS	Yes	-39.27	-44.0	s	-
506475222292587264	01:44:22.3	+59:05:26	IDS	Yes	-57.18	-61.91	s	-
509678889933494912	01:45:06.9	+60:39:30	IDS	Yes	-13.2	-17.93	s	-
469722985282680064	04:00:58.1	+57:22:48	CAFOS	Yes	-	-	-	-
253403143486587136	04:26:22.2	+44:00:31	CAFOS	Yes	-	-	-	-
252274593578484992	04:30:00.7	+42:23:00	CAFOS	Yes	-	-	-	-
2058442688129432960	20:01:53.7	+33:57:49	IDS	Yes	-20.68	-21.22	s	-
2164054323342523520	21:11:00.9	+45:39:41	CAFOS	Yes	-	-	-	-
2165027631639736448	21:21:16.9	+49:01:47	IDS	Yes	-11.47	-15.63	s	-
1981502162946440832	21:59:25.2	+53:37:32	IDS	Yes	-24.19	-28.92	d	-
2006954899366779136	22:46:04.7	+56:27:12	CAFOS	-	-	-	-	-
2016138123559082624	23:52:20.4	+63:17:07	IDS	-	-7.65	-8.15	d	-
2013028635958122240	23:57:22.9	+62:48:15	IDS	Yes	-44.63	-49.36	s	-

Notes: The first three columns were taken from the catalogue of Table 3.4. Hβ comlum is 'Yes' if Hβ emission was observed. Hα EW_{corr} has been corrected from the typical absorption EW values of each spectral sub-type (from Joner and Hintz, 2015). The Hα line profile classification is as follows: single-peaked (s), double-peaked (d) and showing a P-Cygni profile (P), both regular or inverse. 'Comments' column indicates the possible detected contaminants (see Sect. 4.5): 'Emiss' marks sources with a strong emission line spectrum. 'Abs' marks sources with the Hα line in absorption.'B[e]' marks B[e] (FS CMa) stars. 'Evolved' marks evolved stars. 'Cont' indicate other types of contaminants (see main text for further discussion).

Table C.6 Distances, photometry, extinctions, absolute magnitudes and stellar parameters of the 14 observed CBe candidates.

Gaia source id	Parallax (mas)	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	Log(L) ($L_\odot$)	Age (Myr)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
431592948519739136	0.257 ± 0.025	3470^{+340}_{-280}	$2.06^{+0.07}_{-0.15}$	$-2.76^{+0.34}_{-0.27}$	$8.3^{+4.2}_{-2.3}$	$3.79^{+0.18}_{-0.30}$	27^{+14}_{-14}	12.24	12.00	11.60
522471059195179008	0.401 ± 0.038	2320^{+230}_{-200}	$1.643^{+0.067}_{-0.037}$	$-1.85^{+0.23}_{-0.28}$	$4.43^{+0.79}_{-0.40}$	$2.98^{+0.18}_{-0.13}$	139^{+37}_{-52}	11.86	11.62	11.26
506475222292587264	0.186 ± 0.029	4450^{+660}_{-520}	$1.853^{+0.067}_{-0.037}$	$-2.97^{+0.30}_{-0.36}$	$5.94^{+0.77}_{-0.59}$	$3.46^{+0.18}_{-0.16}$	69^{+19}_{-16}	12.39	12.12	11.69
509678889933494912	0.327 ± 0.033	2790^{+290}_{-240}	$1.884^{+0.067}_{-0.037}$	$-1.58^{+0.24}_{-0.28}$	$4.54^{+0.46}_{-0.54}$	$2.89^{+0.16}_{-0.13}$	116^{+58}_{-24}	12.82	12.54	12.10
469722985282680064	0.288 ± 0.028	3150^{+310}_{-260}	$1.94^{+0.15}_{-0.04}$	$-1.61^{+0.23}_{-0.36}$	$5.6^{+1.7}_{-0.4}$	$3.13^{+0.18}_{-0.13}$	57^{+11}_{-29}	13.08	12.82	12.42
253403143486587136	0.356 ± 0.045	2580^{+340}_{-270}	$1.98^{+0.15}_{-0.04}$	$-1.84^{+0.29}_{-0.43}$	$5.8^{+1.7}_{-0.5}$	$3.22^{+0.18}_{-0.15}$	57^{+12}_{-28}	12.45	12.19	11.78
252274593578484992	0.399 ± 0.046	2330^{+290}_{-230}	$1.84^{+0.15}_{-0.04}$	$-1.54^{+0.27}_{-0.41}$	$5.5^{+1.7}_{-0.5}$	$3.10^{+0.18}_{-0.15}$	57^{+12}_{-28}	12.36	12.13	11.76
2058442688129432960	0.364 ± 0.029	2560^{+210}_{-180}	$2.22^{+0.49}_{-0.34}$	$-2.30^{+0.50}_{-0.67}$	$5.2^{+2.6}_{-1.5}$	$3.20^{+0.18}_{-0.30}$	94^{+47}_{-47}	12.30	11.96	11.43
2164054323342523520	0.140 ± 0.019	5690^{+670}_{-550}	$2.30^{+0.15}_{-0.04}$	$-2.74^{+0.26}_{-0.40}$	$6.7^{+3.3}_{-0.9}$	$3.57^{+0.18}_{-0.15}$	51^{+21}_{-26}	13.67	13.33	12.84
2165027631639736448	0.320 ± 0.033	2850^{+300}_{-250}	$2.616^{+0.058}_{-0.033}$	$-2.31^{+0.23}_{-0.27}$	$5.3^{+1.0}_{-0.5}$	$3.27^{+0.17}_{-0.12}$	88^{+21}_{-32}	12.98	12.58	11.95
1981502162946440832	0.179 ± 0.032	4610^{+770}_{-590}	$2.508^{+0.067}_{-0.037}$	$-3.77^{+0.33}_{-0.40}$	$7.1^{+1.2}_{-0.7}$	$3.74^{+0.18}_{-0.15}$	47^{+11}_{-13}	12.50	12.05	11.47
2006954899366779136	0.277 ± 0.029	3240^{+340}_{-280}	$1.832^{+0.058}_{-0.033}$	$-1.81^{+0.23}_{-0.27}$	$5.21^{+0.55}_{-0.46}$	$3.09^{+0.17}_{-0.14}$	80^{+37}_{-16}	12.83	12.57	12.18
2016138123559082624	0.252 ± 0.023	3530^{+310}_{-270}	$2.22^{+0.15}_{-0.04}$	$-2.21^{+0.21}_{-0.34}$	$6.2^{+1.6}_{-0.6}$	$3.37^{+0.18}_{-0.12}$	53^{+26}_{-27}	13.07	12.75	12.28
2013028635958122240	0.321 ± 0.031	2840^{+280}_{-230}	$1.770^{+0.067}_{-0.037}$	$-1.73^{+0.22}_{-0.27}$	$4.35^{+0.84}_{-0.35}$	$2.93^{+0.18}_{-0.11}$	142^{+38}_{-55}	12.56	12.31	11.90

Notes: The first three and the last three columns were taken from the catalogue of Table 3.4. Distances from Bailer-Jones et al. (2018). See Chapter 4 for the derivation of the extinction, absolute magnitude and stellar parameters. A maximum of 50% relative error was set for the uncertainties of the mass, luminosity and age.

Appendix D

SIMBAD set

This appendix contains Tables D.1, D.2, and D.3. These tables tabulate the main compiled and derived parameters of the SIMBAD set of sources described in Chapter 4. In particular, they list coordinates, SIMBAD spectral type, astrometric quality, distance, Gaia DR2 G, G_{BP} and G_{RP} photometry and my derivations of A_G, absolute magnitude, and mass:

1. Table D.1 lists the 103 sources in the PMS catalogue.
2. Table D.2 lists the 56 sources in the CBe catalogue.
3. Table D.3 lists the 28 sources in the Either catalogue.

Table D.1 Compiled and derived parameters of the 103 sources of the SIMBAD set in the PMS catalogue.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
427077052412203776	00:39:52.4	+60:17:19	A0-1IVe	Yes	2520^{+230}_{-190}	$2.52^{+0.14}_{-0.15}$	$1.01^{+0.33}_{-0.33}$	$2.20^{+0.37}_{-0.17}$	16.11	15.54	14.85
512405339480772992	01:31:30.5	+63:09:14	B6-7Ve	Yes	2360^{+220}_{-190}	$2.492^{+0.053}_{-0.063}$	$0.75^{+0.24}_{-0.24}$	$2.94^{+0.31}_{-0.30}$	15.56	15.11	14.51
459182413984008448	02:17:57.8	+59:05:52	F7Ie	-	3600^{+2000}_{-1300}	$2.196^{+0.061}_{-0.065}$	$-1.5^{+1.1}_{-1.0}$	$5.6^{+1.7}_{-1.5}$	14.30	13.49	12.57
458936638775961728	02:25:58.8	+58:36:26	B9Ve	Yes	3170^{+530}_{-410}	$2.24^{+0.13}_{-0.06}$	$0.45^{+0.36}_{-0.47}$	$2.55^{+0.66}_{-0.13}$	15.62	15.19	14.61
465534670614128000	02:32:43.7	+61:28:10	A2	Yes	1953^{+84}_{-78}	$2.17^{+0.11}_{-0.14}$	$0.19^{+0.22}_{-0.20}$	$2.68^{+0.21}_{-0.19}$	14.32	13.81	13.16
465482959206323072	02:34:59.4	+61:19:59	A4	Yes	1830^{+110}_{-100}	$2.81^{+0.06}_{-0.11}$	$0.99^{+0.23}_{-0.19}$	$2.12^{+0.13}_{-0.13}$	15.89	15.12	14.33
465846725755990784	02:36:05.6	+61:22:05	A3	Yes	2090^{+110}_{-100}	$2.38^{+0.14}_{-0.06}$	$0.32^{+0.17}_{-0.25}$	$2.61^{+0.25}_{-0.15}$	14.90	14.30	13.58
465658606190143616	02:39:50.9	+61:18:29	A2e	Yes	2230^{+140}_{-120}	$2.16^{+0.11}_{-0.14}$	$0.15^{+0.26}_{-0.24}$	$2.74^{+0.24}_{-0.24}$	14.55	14.05	13.40
464931073088436096	02:47:48.6	+60:57:51	A0-1Ve	Yes	2130^{+110}_{-100}	$2.11^{+0.14}_{-0.15}$	$0.33^{+0.26}_{-0.26}$	$2.55^{+0.24}_{-0.19}$	14.52	14.07	13.46
467257570973658752	02:51:30.4	+62:10:52	B7-8Ve	-	6300^{+1700}_{-1200}	$2.69^{+0.06}_{-0.15}$	$-0.73^{+0.61}_{-0.57}$	$3.69^{+0.79}_{-0.71}$	16.43	15.96	15.27
464664475880853760	02:52:05.1	+60:24:04	G3.0	-	4500^{+2400}_{-1600}	$2.769^{+0.018}_{-0.010}$	$1.1^{+1.0}_{-0.9}$	$2.95^{+0.85}_{-0.75}$	18.19	17.14	16.02
464676192551516800	02:53:01.0	+60:35:44	A6.5	Yes	1970^{+100}_{-90}	$2.075^{+0.045}_{-0.045}$	$0.96^{+0.15}_{-0.16}$	$2.18^{+0.12}_{-0.10}$	15.09	14.51	13.81
464671485267733760	02:53:06.2	+60:27:45	F3.0	-	2100^{+1500}_{-700}	$2.772^{+0.055}_{-0.066}$	$4.4^{+1.0}_{-1.3}$	-	19.45	18.73	17.55
464674543284312832	02:53:36.1	+60:35:41	G4.0	-	2800^{+1000}_{-600}	$2.523^{+0.010}_{-0.025}$	$2.42^{+0.56}_{-0.67}$	$2.00^{+0.50}_{-0.30}$	18.20	17.19	16.15
464671794501976192	02:53:39.3	+60:30:18	G8.0	-	-	$2.268^{+0.024}_{-0.038}$	-	-	20.28	20.54	18.26
464675574076417408	02:53:43.6	+60:42:07	A1.0	-	3680^{+900}_{-630}	$2.37^{+0.09}_{-0.11}$	$0.82^{+0.51}_{-0.57}$	$2.22^{+0.39}_{-0.20}$	16.51	16.02	15.31
464675093040094976	02:53:47.4	+60:38:47	F3.0	Yes	2240^{+190}_{-160}	$2.170^{+0.055}_{-0.066}$	$1.21^{+0.23}_{-0.23}$	$2.23^{+0.19}_{-0.18}$	15.87	15.13	14.27
464685881999247488	02:54:28.9	+60:38:19	K1.5	-	3300^{+1300}_{-800}	$2.232^{+0.049}_{-0.049}$	$2.36^{+0.66}_{-0.80}$	$2.30^{+0.54}_{-0.35}$	18.21	17.16	16.08
464686294316066176	02:55:02.2	+60:40:17	F5.0	Yes	1900^{+380}_{-270}	$2.315^{+0.035}_{-0.085}$	$3.20^{+0.42}_{-0.43}$	$1.33^{+0.11}_{-0.08}$	17.71	16.92	15.99
464693780436748800	02:56:22.8	+60:46:56	K3.0	-	3300^{+2200}_{-1300}	$2.25^{+0.29}_{-0.23}$	$4.5^{+1.3}_{-1.4}$	$1.33^{+0.56}_{-0.51}$	20.30	19.27	17.99
466194098411007616	02:57:21.5	+60:53:09	G6.5	-	1800^{+2000}_{-800}	$4.157^{+0.021}_{-0.021}$	$3.8^{+1.3}_{-1.7}$	$1.38^{+0.69}_{-0.44}$	20.68	19.17	17.75

Table D.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
464641210037937920	02:57:35.9	+60:37:49	K0.5	-	1690^{+860}_{-440}	$2.351^{+0.046}_{-0.046}$	$4.79^{+0.71}_{-0.94}$	$1.05^{+0.46}_{-0.18}$	19.13	18.28	16.99
464624824741993088	02:58:07.8	+60:29:14	F6.0	Yes	1940^{+220}_{-180}	$2.530^{+0.085}_{-0.061}$	$2.35^{+0.28}_{-0.32}$	$1.65^{+0.20}_{-0.15}$	17.22	16.31	15.36
466140604596689280	02:58:56.9	+60:42:13	A2.0	Yes	2410^{+280}_{-230}	$2.22^{+0.11}_{-0.14}$	$1.33^{+0.36}_{-0.35}$	$2.01^{+0.24}_{-0.11}$	15.99	15.46	14.81
466128333870122752	02:59:02.1	+60:36:24	K2.0	-	2190^{+820}_{-490}	$2.13^{+0.10}_{-0.29}$	$3.77^{+0.84}_{-0.79}$	$1.60^{+0.40}_{-0.45}$	18.62	17.60	16.51
466127135579300480	02:59:40.8	+60:34:54	K1.0	-	2500^{+1600}_{-800}	$2.16^{+0.09}_{-0.10}$	$4.4^{+0.9}_{-1.2}$	$1.25^{+0.63}_{-0.37}$	19.36	18.50	17.29
466124077562607488	02:59:44.0	+60:32:29	A2.5	Yes	2140^{+150}_{-130}	$2.178^{+0.068}_{-0.068}$	$1.24^{+0.20}_{-0.21}$	$1.99^{+0.11}_{-0.06}$	15.59	15.06	14.39
466126585823413248	03:00:44.8	+60:40:09	K3.5	Yes	830^{+140}_{-110}	$2.17^{+0.11}_{-0.11}$	$6.51^{+0.41}_{-0.46}$	$0.75^{+0.10}_{-0.03}$	19.32	18.28	16.99
463121070787796864	03:01:21.6	+60:28:57	B9Ve	Yes	2220^{+110}_{-100}	$2.95^{+0.13}_{-0.06}$	$-0.69^{+0.16}_{-0.24}$	$3.54^{+0.34}_{-0.18}$	14.60	13.99	13.24
463129110966547584	03:01:22.8	+60:39:40	G7.0	-	1730^{+510}_{-330}	$2.669^{+0.042}_{-0.024}$	$3.67^{+0.48}_{-0.60}$	$1.45^{+0.35}_{-0.25}$	18.65	17.53	16.44
463121654903339008	03:01:29.3	+60:31:14	K4.0	-	2120^{+820}_{-480}	$2.25^{+0.23}_{-0.28}$	$3.76^{+0.83}_{-0.94}$	$1.05^{+0.35}_{-0.28}$	18.83	17.64	16.41
463112510919178752	03:03:20.8	+60:29:28	B7.0	-	2380^{+640}_{-430}	$4.066^{+0.032}_{-0.046}$	$1.19^{+0.48}_{-0.55}$	-	17.92	17.14	16.09
462785234405806720	03:10:46.3	+59:30:04	B5-6Ve	Yes	2150^{+220}_{-180}	$3.863^{+0.085}_{-0.051}$	$-0.04^{+0.24}_{-0.29}$	$3.70^{+0.47}_{-0.44}$	16.30	15.49	14.60
449791072846755072	03:26:20.5	+58:42:41	G0e	Yes	979^{+75}_{-66}	$4.162^{+0.027}_{-0.057}$	$2.07^{+0.21}_{-0.19}$	$2.03^{+0.19}_{-0.14}$	17.76	16.19	14.97
450175455241251456	03:27:58.5	+58:58:34	A2e	Yes	815^{+52}_{-46}	$4.49^{+0.11}_{-0.14}$	$2.17^{+0.26}_{-0.24}$	$1.79^{+0.12}_{-0.05}$	17.40	16.21	15.11
449511693813662464	03:28:47.9	+57:55:56	A2e	Yes	1126^{+29}_{-28}	$2.67^{+0.11}_{-0.14}$	$1.34^{+0.19}_{-0.16}$	$2.01^{+0.24}_{-0.11}$	14.92	14.27	13.52
449886524200720384	03:30:03.0	+58:05:35	F:e	Yes	1310^{+190}_{-150}	$5.76^{+0.39}_{-0.35}$	$0.77^{+0.61}_{-0.68}$	$2.7^{+1.0}_{-0.7}$	19.15	17.11	15.76
449903356177893248	03:34:00.7	+58:16:38	G5e	Yes	1121^{+62}_{-56}	$3.23^{+0.03}_{-0.10}$	$2.29^{+0.22}_{-0.14}$	$2.10^{+0.27}_{-0.15}$	16.97	15.77	14.55
448592605173283712	03:34:21.2	+56:43:44	A5	Yes	1084^{+28}_{-27}	$2.89^{+0.11}_{-0.00}$	$1.20^{+0.06}_{-0.17}$	$2.00^{+0.11}_{-0.02}$	15.08	14.27	13.42
188262538442656512	05:10:02.2	+38:49:18	B	-	7300^{+2300}_{-1600}	$2.18^{+0.51}_{-0.47}$	$-3.8^{+1.0}_{-1.1}$	$10.1^{+5.1}_{-3.2}$	13.00	12.74	12.16
181174059990205184	05:22:51.0	+33:25:47	F7:Ve	-	3940^{+920}_{-650}	$1.118^{+0.061}_{-0.065}$	$0.85^{+0.45}_{-0.51}$	$2.75^{+0.54}_{-0.41}$	15.50	14.94	14.29
181176533891335936	05:22:52.2	+33:29:58	F6:Ve	-	4000^{+930}_{-660}	$1.811^{+0.085}_{-0.061}$	$-0.01^{+0.45}_{-0.54}$	$3.52^{+0.73}_{-0.52}$	15.52	14.82	14.01
181176636970539648	05:22:58.1	+33:30:41	A3:Ve	Yes	3170^{+600}_{-440}	$1.33^{+0.14}_{-0.06}$	$1.28^{+0.38}_{-0.51}$	$1.97^{+0.29}_{-0.11}$	15.43	15.11	14.61

Table D.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
181175743617347328	05:23:04.3	+33:28:46	B4Ve	Yes	3710^{+650}_{-490}	$2.985^{+0.033}_{-0.067}$	$-1.46^{+0.37}_{-0.38}$	$4.95^{+0.65}_{-0.65}$	14.89	14.38	13.66
182636406157011072	05:28:06.4	+34:25:01	B1V	-	1800^{+1100}_{-500}	$2.10^{+0.10}_{-0.07}$	$-1.1^{+0.8}_{-1.1}$	$6.6^{+3.3}_{-0.3}$	12.47	12.27	11.85
189195959387017600	05:39:52.1	+36:20:13	F0:	-	5000^{+2300}_{-1500}	$4.92^{+0.04}_{-0.14}$	$-2.99^{+0.94}_{-0.86}$	$8.0^{+1.8}_{-1.8}$	16.91	15.42	14.10
3400052933657222272	05:47:03.8	+21:00:35	G5III-IV	Yes	1870^{+230}_{-180}	$3.42^{+0.03}_{-0.10}$	$0.13^{+0.33}_{-0.28}$	$3.91^{+0.36}_{-0.32}$	16.29	14.91	13.79
3437613792753603072	06:07:23.9	+30:11:44	Be	-	10600^{+2600}_{-2000}	$2.97^{+0.51}_{-0.47}$	$-5.8^{+0.9}_{-1.0}$	21^{+10}_{-8}	12.71	12.27	11.49
3436756074901231872	06:10:17.3	+29:25:17	B9	Yes	1053^{+36}_{-34}	$2.75^{+0.13}_{-0.06}$	$0.53^{+0.13}_{-0.20}$	$2.57^{+0.63}_{-0.12}$	13.95	13.39	12.69
3326710044885698432	06:40:42.2	+09:40:11	K4V	Yes	696^{+23}_{-22}	$0.97^{+0.23}_{-0.28}$	$5.24^{+0.34}_{-0.30}$	$1.10^{+0.09}_{-0.20}$	16.27	15.41	14.47
3326684687399179392	06:41:08.8	+09:23:43	K4V	-	666^{+57}_{-49}	$1.07^{+0.23}_{-0.28}$	$3.95^{+0.44}_{-0.41}$	$1.05^{+0.33}_{-0.29}$	15.06	14.14	13.21
3107263943641210496	06:43:35.7	-01:09:30	B2V	-	-	$2.56^{+0.15}_{-0.04}$	-	-	14.33	14.13	13.37
3125925237244613248	06:44:37.7	+01:19:33	B[e]	-	9700^{+2900}_{-2100}	$3.74^{+0.51}_{-0.47}$	$-4.1^{+1.0}_{-1.1}$	$11.3^{+5.6}_{-3.7}$	15.12	14.60	13.52
3127268943529262848	06:44:53.0	+02:58:05	A:e	-	183^{+21}_{-17}	$2.52^{+0.54}_{-0.39}$	$8.95^{+0.61}_{-0.78}$	-	17.97	17.78	16.52
3125506014076025088	06:44:55.8	+00:13:17	F2	-	4300^{+1000}_{-700}	$0.944^{+0.069}_{-0.055}$	$1.15^{+0.44}_{-0.52}$	$2.24^{+0.46}_{-0.31}$	15.69	15.28	14.72
3125726156919234176	06:46:56.4	+01:16:41	B0.5Ve	Yes	2220^{+380}_{-290}	$4.936^{+0.050}_{-0.050}$	$-1.69^{+0.35}_{-0.39}$	$8.3^{+1.7}_{-0.8}$	15.91	14.98	13.92
3129183674313574784	06:47:04.2	+04:37:18	B?	-	6100^{+1800}_{-1200}	$3.39^{+0.51}_{-0.47}$	$-2.1^{+1.0}_{-1.1}$	$6.0^{+3.0}_{-1.9}$	15.73	15.23	14.30
5597659331053579008	08:06:34.9	-28:32:01	G5III	-	8000^{+1900}_{-1400}	$1.84^{+0.03}_{-0.10}$	$-3.66^{+0.51}_{-0.47}$	$9.5^{+1.1}_{-1.1}$	13.56	12.71	11.82
5329885506170824832	08:48:45.5	-46:05:09	Beqp	-	7100^{+1300}_{-1000}	$4.38^{+0.51}_{-0.47}$	$-5.79^{+0.80}_{-0.88}$	21^{+10}_{-7}	13.77	12.83	11.86
6054145325942642048	12:15:20.5	-63:39:13	B[e]	-	2900^{+340}_{-270}	$2.89^{+0.51}_{-0.47}$	$-1.74^{+0.69}_{-0.75}$	$5.4^{+2.7}_{-1.4}$	13.84	13.46	12.66
5853602028546852352	14:11:46.3	-64:16:24	Be	-	5500^{+2900}_{-1800}	$3.57^{+0.51}_{-0.47}$	$-2.4^{+1.3}_{-1.4}$	$6.6^{+3.3}_{-2.6}$	15.47	14.90	13.96
5934909852598399232	16:14:15.4	-51:46:04	B[e]	-	2300^{+2000}_{-900}	$6.40^{+0.51}_{-0.47}$	$-1.1^{+1.5}_{-1.9}$	$4.4^{+2.2}_{-1.9}$	18.54	17.13	15.64
5964315500792772352	16:53:24.7	-44:52:53	Bepec	-	1690^{+220}_{-180}	$3.57^{+0.51}_{-0.47}$	$-0.23^{+0.71}_{-0.77}$	$4.0^{+2.0}_{-1.5}$	15.08	14.48	13.57
5973991306187697152	17:14:45.0	-36:18:38	B1-B5	-	-	$6.55^{+0.37}_{-0.19}$	-	-	15.92	14.27	13.00
4146594871922064000	18:19:17.1	-13:54:51	A1Ve	Yes	1690^{+100}_{-90}	$1.82^{+0.09}_{-0.11}$	$1.12^{+0.23}_{-0.22}$	$2.10^{+0.27}_{-0.09}$	14.47	14.08	13.54

Table D.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
4273117977040507648	18:27:53.8	-00:02:34	K2:	Yes	423^{+12}_{-11}	$2.81^{+0.10}_{-0.29}$	$4.07^{+0.35}_{-0.16}$	$1.46^{+0.09}_{-0.22}$	16.27	15.01	13.83
4273036059117893632	18:29:40.2	+00:15:13	K2:	Yes	532^{+89}_{-67}	$1.62^{+0.10}_{-0.29}$	$8.13^{+0.59}_{-0.44}$	-	19.22	18.38	17.36
4154600965877954688	18:34:03.0	-11:32:32	B[e]	-	3180^{+880}_{-580}	$3.91^{+0.51}_{-0.47}$	$-0.9^{+0.9}_{-1.0}$	$4.3^{+2.1}_{-1.4}$	16.24	15.49	14.56
4268419106711771776	19:02:00.3	+02:09:11	A7V-F0V	Yes	1620^{+100}_{-90}	$1.73^{+0.25}_{-0.23}$	$0.67^{+0.35}_{-0.38}$	$2.48^{+0.46}_{-0.33}$	13.96	13.45	12.75
4261761087116725248	19:02:30.0	-02:27:57	B[e]	-	7600^{+3200}_{-2200}	$3.92^{+0.51}_{-0.47}$	$-1.8^{+1.2}_{-1.3}$	$5.5^{+2.7}_{-2.0}$	17.00	16.53	15.32
4512787627579740032	19:11:18.8	+15:32:33	Be	-	10400^{+3400}_{-2500}	$5.80^{+0.51}_{-0.47}$	$-7.4^{+1.1}_{-1.1}$	40^{+20}_{-20}	14.93	13.49	12.32
4293406612283985024	19:11:49.6	+04:58:58	A7Ib:	-	3800^{+1100}_{-800}	$5.02^{+0.09}_{-0.11}$	$-5.33^{+0.58}_{-0.66}$	$14.5^{+3.1}_{-2.2}$	14.14	12.61	11.39
4268281088010981248	19:13:22.6	+03:25:00	B1I	-	5700^{+1200}_{-800}	$3.33^{+0.10}_{-0.07}$	$-2.64^{+0.42}_{-0.50}$	$9.0^{+2.7}_{-1.4}$	14.95	14.48	13.73
4513629514212361984	19:14:59.7	+17:22:46	B9V	-	1750^{+160}_{-130}	$3.34^{+0.13}_{-0.06}$	$-1.30^{+0.23}_{-0.32}$	$4.31^{+0.52}_{-0.31}$	13.95	13.25	12.40
2018024881202947200	19:26:37.7	+21:09:27	O9-B0I	-	7600^{+1900}_{-1400}	$3.90^{+0.22}_{-0.12}$	$-3.15^{+0.55}_{-0.71}$	$14.0^{+7.0}_{-4.8}$	15.49	15.16	14.05
2020388040934202880	19:39:35.5	+24:06:27	B3-6I	-	-	$7.54^{+0.12}_{-0.10}$	-	-	18.47	17.32	14.99
2020112789377651712	19:44:05.3	+23:26:48	B:e:	Yes	1940^{+130}_{-110}	$4.06^{+0.51}_{-0.47}$	$-1.00^{+0.60}_{-0.64}$	$4.4^{+2.2}_{-1.1}$	15.26	14.50	13.50
2033745393571783808	19:47:51.9	+31:18:18	B0I	-	8300^{+3600}_{-2300}	$3.24^{+0.05}_{-0.10}$	$-1.42^{+0.79}_{-0.82}$	-	16.72	16.42	15.59
2059843023592531712	20:04:18.6	+35:54:26	A0	Yes	1617^{+80}_{-73}	$2.02^{+0.10}_{-0.09}$	$-0.54^{+0.19}_{-0.20}$	$3.37^{+0.24}_{-0.20}$	12.96	12.53	11.97
2055181605358559872	20:06:39.9	+33:14:28	B1[e]+K?	-	4350^{+930}_{-670}	$6.64^{+0.10}_{-0.07}$	$-7.19^{+0.43}_{-0.52}$	66^{+33}_{-33}	14.23	12.63	11.39
2059181529904135424	20:11:43.4	+36:31:08	B5ev	-	2900^{+1300}_{-700}	$2.087^{+0.067}_{-0.037}$	$-1.76^{+0.65}_{-0.84}$	$5.3^{+1.7}_{-1.0}$	13.01	12.68	12.19
2080664612696881152	20:14:38.6	+41:56:14	G7+K4	Yes	995^{+19}_{-18}	$0.278^{+0.042}_{-0.024}$	$3.589^{+0.065}_{-0.083}$	$1.489^{+0.061}_{-0.039}$	14.33	13.86	13.29
2068285722372121344	20:20:49.3	+40:33:03	B1.5V	Yes	1761^{+93}_{-84}	$5.35^{+0.07}_{-0.15}$	$-4.25^{+0.26}_{-0.18}$	$14.0^{+1.6}_{-2.2}$	13.51	12.33	11.27
2060969232708934784	20:21:28.4	+37:26:20	B2e	Yes	1564^{+54}_{-50}	$5.06^{+0.15}_{-0.04}$	$-2.64^{+0.11}_{-0.23}$	$7.7^{+1.3}_{-0.5}$	14.52	13.39	12.35
2056853691965982720	20:29:47.9	+35:59:26	B2	-	2330^{+330}_{-260}	$5.47^{+0.15}_{-0.04}$	$-2.76^{+0.30}_{-0.44}$	$7.9^{+2.1}_{-0.8}$	15.73	14.55	13.35
2066405218188247936	20:39:00.9	+41:53:59	K1.5	-	$306.0^{+2.4}_{-2.4}$	$0.332^{+0.049}_{-0.049}$	$5.593^{+0.066}_{-0.066}$	$0.850^{+0.040}_{-0.011}$	13.91	13.35	12.70
2066407107973861888	20:39:04.3	+41:57:03	F6.5	Yes	$557.7^{+8.6}_{-8.4}$	$0.164^{+0.031}_{-0.031}$	$3.670^{+0.063}_{-0.064}$	$1.219^{+0.023}_{-0.014}$	12.88	12.57	12.15

Table D.1 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
2167366464672809216	20:41:53.7	+46:49:20	B3:	-	6800^{+1900}_{-1300}	$5.909^{+0.058}_{-0.033}$	$-6.80^{+0.50}_{-0.58}$	32^{+10}_{-6}	14.73	13.29	12.09
2066987443960692864	20:46:36.8	+43:44:35	Be	-	950^{+200}_{-140}	$3.43^{+0.51}_{-0.47}$	$0.05^{+0.82}_{-0.92}$	$3.6^{+1.8}_{-1.3}$	14.02	13.38	12.57
2066913742320709760	20:46:46.9	+43:17:35	dK0	-	5700^{+1700}_{-1200}	$5.828^{+0.079}_{-0.092}$	$-4.03^{+0.63}_{-0.65}$	$10.6^{+1.8}_{-1.5}$	17.95	15.57	14.13
2067074305377614720	20:49:32.2	+44:17:03	F0e	Yes	891^{+54}_{-49}	$4.24^{+0.04}_{-0.14}$	$2.96^{+0.27}_{-0.17}$	$1.48^{+0.12}_{-0.06}$	18.19	16.95	15.71
2163782228572361856	20:55:44.3	+46:33:13	B[e]	-	6100^{+1900}_{-1400}	$4.79^{+0.51}_{-0.47}$	$-2.9^{+1.0}_{-1.1}$	$7.7^{+3.8}_{-2.5}$	16.72	15.85	14.60
2162236693175188992	20:59:17.2	+44:17:46	Be	Yes	2290^{+130}_{-120}	$4.74^{+0.51}_{-0.47}$	$-2.94^{+0.58}_{-0.63}$	$7.8^{+3.9}_{-1.7}$	14.66	13.60	12.58
2179194048688570240	21:33:17.8	+57:48:13	F3	-	1880^{+670}_{-400}	$1.714^{+0.055}_{-0.066}$	$1.85^{+0.58}_{-0.72}$	$1.82^{+0.51}_{-0.29}$	15.48	14.93	14.11
2178395631452868224	21:37:38.8	+57:29:37	K3.5e	Yes	875^{+35}_{-32}	$2.04^{+0.11}_{-0.11}$	$4.45^{+0.20}_{-0.20}$	$1.25^{+0.10}_{-0.16}$	17.37	16.20	15.10
2178562551060090624	21:40:36.9	+58:14:38	K2e	Yes	979^{+27}_{-25}	$2.41^{+0.10}_{-0.29}$	$2.38^{+0.35}_{-0.16}$	$2.20^{+0.21}_{-0.77}$	15.92	14.75	13.68
2178562654139305856	21:40:41.2	+58:15:11	K4e	Yes	1540^{+430}_{-280}	$3.04^{+0.23}_{-0.28}$	$3.61^{+0.71}_{-0.77}$	$1.04^{+0.36}_{-0.27}$	18.99	17.58	16.18
2001249121215647488	22:15:40.4	+52:15:56	A2e	Yes	3680^{+620}_{-470}	$1.11^{+0.11}_{-0.14}$	$1.65^{+0.43}_{-0.45}$	$2.04^{+0.20}_{-0.27}$	15.85	15.58	15.21
2006046771487587328	22:24:13.8	+56:11:33	G8	-	2810^{+230}_{-200}	$1.397^{+0.024}_{-0.038}$	$-0.35^{+0.20}_{-0.20}$	$4.69^{+0.20}_{-0.09}$	14.00	13.29	12.40
2007418171718891904	22:47:22.4	+58:01:21	Ae	-	4700^{+2200}_{-1500}	$4.06^{+0.54}_{-0.39}$	$-3.5^{+1.2}_{-1.4}$	$8.9^{+4.3}_{-3.1}$	15.07	13.95	12.87
2207277877757956352	22:53:55.1	+62:43:37	BIII	-	8100^{+2300}_{-1600}	$5.25^{+0.51}_{-0.47}$	$-5.2^{+0.9}_{-1.0}$	$16.6^{+8.3}_{-5.5}$	15.82	14.57	13.48
2014648388374019968	22:54:07.6	+60:50:23	B0Ve	Yes	3420^{+450}_{-360}	$6.19^{+0.05}_{-0.10}$	$-4.62^{+0.34}_{-0.31}$	23^{+12}_{-10}	15.66	14.24	13.08
2014636053227982976	22:54:36.6	+60:48:36	B0Ve	Yes	3090^{+270}_{-230}	$5.74^{+0.05}_{-0.10}$	$-4.80^{+0.27}_{-0.23}$	20^{+10}_{-6}	14.63	13.38	12.28

Notes: 'SIMBAD' column lists the spectral types tabulated at the SIMBAD database at the time of writing. 'Astrom' column value is 'Yes' if the source has $\varpi/\sigma(\varpi) \geq 5$ and RUWE < 1.4. Distances from Bailer-Jones et al. (2018). See Chapter 4 for the derivation of the extinction, absolute magnitude and mass. A maximum of 50% relative error was set for the uncertainty of the mass.

Table D.2 Compiled and derived parameters of the 56 sources of the SIMBAD set in the CBe catalogue.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
432172258012607488	00:04:52.7	+65:05:50	B:e	Yes	3160^{+220}_{-190}	$3.03^{+0.51}_{-0.47}$	$-2.33^{+0.61}_{-0.65}$	$5.3^{+2.7}_{-1.8}$	13.75	13.20	12.50
430675371706383872	00:35:26.2	+62:15:16	B	Yes	3600^{+380}_{-320}	$2.46^{+0.51}_{-0.47}$	$-2.73^{+0.67}_{-0.73}$	$5.7^{+2.9}_{-1.9}$	12.92	12.52	11.95
427784931739195136	00:42:22.6	+62:38:21	B	Yes	2630^{+240}_{-200}	$3.13^{+0.51}_{-0.47}$	$-2.93^{+0.64}_{-0.70}$	$6.0^{+3.0}_{-1.9}$	12.88	12.30	11.58
427006786745454720	00:43:53.2	+59:56:42	B1Ve	-	10700^{+2300}_{-1800}	$2.69^{+0.10}_{-0.07}$	$-2.55^{+0.46}_{-0.52}$	$8.9^{+4.4}_{-1.5}$	15.67	15.29	14.76
523655164499151744	00:51:55.2	+62:55:07	B	Yes	3870^{+480}_{-390}	$2.24^{+0.51}_{-0.47}$	$-2.48^{+0.70}_{-0.76}$	$5.7^{+2.8}_{-2.1}$	13.06	12.70	12.19
426510803924862592	00:55:56.4	+60:15:06	A8	Yes	2440^{+190}_{-170}	$0.10^{+0.11}_{-0.10}$	$0.36^{+0.26}_{-0.27}$	$2.28^{+0.12}_{-0.30}$	12.54	12.39	12.15
426276264348003456	01:03:58.1	+59:53:11	B3-4Ve	Yes	5780^{+880}_{-690}	$2.095^{+0.075}_{-0.084}$	$-1.99^{+0.36}_{-0.38}$	$5.18^{+0.86}_{-0.94}$	14.24	13.92	13.45
414092369841623296	01:15:20.3	+58:50:03	B5-7Ve	Yes	3280^{+200}_{-180}	$1.77^{+0.10}_{-0.08}$	$-1.26^{+0.20}_{-0.23}$	$4.11^{+0.55}_{-0.58}$	13.36	13.09	12.67
512579539048321408	01:27:02.2	+63:02:23	B	Yes	2790^{+230}_{-200}	$2.65^{+0.51}_{-0.47}$	$-2.56^{+0.63}_{-0.68}$	$5.7^{+2.9}_{-2.0}$	12.78	12.32	11.71
509981809696980096	01:33:28.7	+61:07:59	B4Ve	Yes	3500^{+430}_{-350}	$2.066^{+0.033}_{-0.067}$	$-2.19^{+0.30}_{-0.29}$	$5.30^{+0.59}_{-0.79}$	12.90	12.59	12.12
509938001029116928	01:40:34.0	+61:09:43	B	Yes	2840^{+280}_{-230}	$1.93^{+0.51}_{-0.47}$	$-1.77^{+0.65}_{-0.71}$	$4.9^{+2.5}_{-1.8}$	12.71	12.43	11.99
509534510315192704	01:43:23.3	+59:53:07	B5-6Ve	Yes	3990^{+330}_{-290}	$2.947^{+0.085}_{-0.051}$	$-2.20^{+0.21}_{-0.26}$	$4.95^{+0.71}_{-0.44}$	14.31	13.75	13.06
515067630783462016	02:08:26.9	+64:22:42	A1V	Yes	4510^{+680}_{-530}	$1.50^{+0.09}_{-0.11}$	$-2.20^{+0.38}_{-0.40}$	$3.91^{+0.42}_{-0.37}$	12.89	12.58	12.12
515107080065239040	02:19:32.3	+63:43:47	B2-3Ve	Yes	4690^{+440}_{-370}	$2.70^{+0.20}_{-0.09}$	$-3.04^{+0.27}_{-0.39}$	$7.0^{+3.5}_{-1.1}$	13.46	13.02	12.42
459097064392907264	02:23:25.5	+58:30:03	B	Yes	3390^{+540}_{-410}	$2.41^{+0.51}_{-0.47}$	$-2.66^{+0.75}_{-0.83}$	$5.9^{+2.9}_{-2.1}$	12.78	12.39	11.83
458889222336467584	02:24:31.8	+58:09:32	B	Yes	2280^{+210}_{-180}	$2.41^{+0.51}_{-0.47}$	$-1.75^{+0.65}_{-0.70}$	$4.9^{+2.5}_{-1.8}$	12.83	12.45	11.88
467441666153712128	02:58:56.7	+63:35:58	B5-6Ve	Yes	5710^{+890}_{-690}	$2.420^{+0.085}_{-0.051}$	$-2.61^{+0.33}_{-0.40}$	$5.27^{+0.90}_{-0.51}$	14.01	13.59	13.02
447329197595237248	03:06:12.1	+54:32:53	Bp	Yes	2570^{+240}_{-200}	$3.17^{+0.51}_{-0.47}$	$-2.37^{+0.65}_{-0.70}$	$5.5^{+2.7}_{-1.9}$	13.45	12.85	12.13
463295854477414912	03:07:26.8	+61:33:37	Bp	-	5100^{+1100}_{-800}	$2.98^{+0.51}_{-0.47}$	$-3.96^{+0.83}_{-0.92}$	$8.0^{+4.0}_{-3.0}$	13.10	12.55	11.87
462035023880343936	03:13:44.8	+59:27:58	Bp	Yes	3230^{+340}_{-280}	$2.95^{+0.51}_{-0.47}$	$-3.14^{+0.67}_{-0.72}$	$6.4^{+3.2}_{-2.1}$	12.89	12.36	11.67
251943984175934720	03:52:21.2	+52:28:36	Bp	Yes	2140^{+180}_{-150}	$2.52^{+0.51}_{-0.47}$	$-2.05^{+0.63}_{-0.68}$	$4.8^{+2.4}_{-1.5}$	12.55	12.12	11.54

Table D.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
251768410211245696	03:56:14.1	+52:26:03	F0e	Yes	2600^{+300}_{-250}	$1.46^{+0.04}_{-0.14}$	$-0.86^{+0.36}_{-0.28}$	$2.82^{+0.21}_{-0.22}$	13.16	12.67	12.03
468432742744697856	04:01:28.5	+55:05:24	Bp	Yes	3700^{+540}_{-420}	$3.07^{+0.51}_{-0.47}$	$-3.13^{+0.73}_{-0.80}$	$6.3^{+3.2}_{-2.2}$	13.36	12.78	12.08
274939204718646656	04:13:36.9	+52:44:35	Bp	-	4670^{+910}_{-670}	$2.58^{+0.51}_{-0.47}$	$-3.56^{+0.81}_{-0.89}$	$7.1^{+3.5}_{-2.5}$	12.79	12.36	11.76
187875544704682496	05:16:58.6	+39:12:51	B	Yes	2480^{+360}_{-280}	$2.47^{+0.51}_{-0.47}$	$-2.17^{+0.73}_{-0.80}$	$5.2^{+2.6}_{-1.8}$	12.69	12.27	11.71
181176155934193280	05:23:09.3	+33:30:02	B1.5Ve	Yes	3630^{+740}_{-540}	$2.54^{+0.07}_{-0.15}$	$-2.87^{+0.50}_{-0.47}$	$8.4^{+4.2}_{-2.3}$	12.81	12.46	11.94
3445423280247717504	05:29:29.0	+28:25:36	B	-	4600^{+1100}_{-800}	$2.20^{+0.51}_{-0.47}$	$-2.7^{+0.9}_{-1.0}$	$5.7^{+2.8}_{-2.0}$	13.13	12.79	12.29
3448514011733266816	05:41:58.6	+33:09:49	A1/5e	Yes	1980^{+170}_{-140}	$0.70^{+0.33}_{-0.17}$	$-0.12^{+0.33}_{-0.51}$	$2.53^{+0.44}_{-0.34}$	12.26	12.07	11.75
3443617710356727296	05:55:03.2	+29:10:01	A1/3	Yes	4250^{+680}_{-520}	$0.76^{+0.20}_{-0.19}$	$-0.39^{+0.48}_{-0.52}$	$2.63^{+0.48}_{-0.37}$	13.67	13.51	13.20
3449916988569002880	06:02:00.0	+30:57:29	B	-	5700^{+1900}_{-1300}	$1.89^{+0.51}_{-0.47}$	$-3.6^{+1.0}_{-1.1}$	$7.2^{+3.6}_{-2.9}$	12.29	12.03	11.60
3426505632935655680	06:02:44.4	+24:47:44	F8:	-	4000^{+1000}_{-700}	$0.097^{+0.065}_{-0.094}$	$-1.01^{+0.52}_{-0.56}$	$3.01^{+0.43}_{-0.36}$	12.40	12.12	11.66
3331579090395206272	06:17:59.2	+11:16:46	B	-	3700^{+2600}_{-1600}	$1.96^{+0.51}_{-0.47}$	$-2.7^{+1.7}_{-1.7}$	$6.1^{+3.0}_{-3.0}$	12.42	12.14	11.69
3330427386323978368	06:23:07.9	+09:55:03	B	-	6600^{+1800}_{-1300}	$1.32^{+0.51}_{-0.47}$	$-3.1^{+0.9}_{-1.0}$	$6.3^{+3.2}_{-2.3}$	12.47	12.30	12.06
3132155684301328512	06:35:00.0	+06:28:51	Bp	-	4500^{+1000}_{-700}	$2.85^{+0.51}_{-0.47}$	$-4.0^{+0.9}_{-1.0}$	$8.1^{+4.0}_{-3.1}$	12.65	12.14	11.49
3119355242950146816	06:40:54.1	-00:37:07	B0IV	-	6100^{+1500}_{-1100}	$2.45^{+0.05}_{-0.10}$	$-3.57^{+0.52}_{-0.53}$	$12.3^{+6.2}_{-2.7}$	13.08	12.79	12.34
3107289232414212736	06:41:16.5	-01:07:40	B0IV	-	6800^{+1300}_{-1000}	$3.43^{+0.05}_{-0.10}$	$-4.40^{+0.43}_{-0.42}$	34^{+2}_{-17}	13.71	13.17	12.49
3107448700253675392	06:45:03.4	-00:34:14	K3V	-	4100^{+1000}_{-700}	$-0.05^{+0.29}_{-0.23}$	$0.84^{+0.63}_{-0.77}$	$1.70^{+0.85}_{-0.76}$	14.38	13.85	13.19
3107060297771500032	06:45:09.3	-01:15:21	O9IV	-	5400^{+1300}_{-900}	$3.86^{+0.20}_{-0.05}$	$-4.70^{+0.46}_{-0.68}$	39^{+3}_{-19}	13.45	12.82	12.04
3129237374282787584	06:47:07.7	+04:42:32	B	-	7100^{+2400}_{-1800}	$2.17^{+0.51}_{-0.47}$	$-3.8^{+1.1}_{-1.1}$	$7.7^{+3.8}_{-3.1}$	12.91	12.58	12.08
3106136634984970112	06:48:34.4	-02:18:31	F4III	Yes	1730^{+140}_{-120}	$0.312^{+0.066}_{-0.035}$	$0.71^{+0.19}_{-0.23}$	$1.99^{+0.10}_{-0.07}$	12.52	12.22	11.80
3106114885277650944	06:49:27.0	-02:33:30	B6III	Yes	2600^{+270}_{-230}	$1.899^{+0.037}_{-0.032}$	$-1.33^{+0.23}_{-0.25}$	$4.20^{+0.26}_{-0.40}$	12.93	12.65	12.18
3106222710430617728	06:50:04.8	-02:15:10	B1V	-	4400^{+1100}_{-800}	$2.99^{+0.10}_{-0.07}$	$-3.56^{+0.49}_{-0.59}$	$10.7^{+5.4}_{-2.3}$	13.12	12.66	12.06
2028814423187785856	19:52:09.5	+29:23:34	B1Ve	Yes	3470^{+420}_{-340}	$2.87^{+0.10}_{-0.07}$	$-3.24^{+0.29}_{-0.35}$	$10.1^{+5.0}_{-1.6}$	12.76	12.33	11.76

Table D.2 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
2030977991509511936	20:01:23.7	+31:58:00	B	Yes	2590^{+190}_{-160}	$2.22^{+0.51}_{-0.47}$	$-1.36^{+0.61}_{-0.66}$	$4.6^{+2.3}_{-1.7}$	13.27	12.92	12.41
2166419853889615360	20:53:14.3	+46:34:59	B:	Yes	4670^{+530}_{-440}	$3.29^{+0.51}_{-0.47}$	$-3.46^{+0.68}_{-0.74}$	$6.9^{+3.5}_{-2.4}$	13.80	13.17	12.42
2163411482678514688	20:53:57.9	+46:23:09	B3V	Yes	2350^{+160}_{-140}	$1.759^{+0.058}_{-0.033}$	$-1.76^{+0.17}_{-0.20}$	$5.11^{+0.49}_{-0.31}$	12.09	11.86	11.48
2065653598916388352	20:58:47.5	+41:46:37	O9.5-B0IV-Ve	-	8000^{+1200}_{-900}	$4.41^{+0.20}_{-0.11}$	$-4.77^{+0.38}_{-0.50}$	24^{+12}_{-12}	14.99	14.16	13.29
2162212675717093120	21:01:40.6	+44:19:44	Be	Yes	3620^{+240}_{-210}	$2.94^{+0.51}_{-0.47}$	$-2.23^{+0.60}_{-0.64}$	$5.2^{+2.6}_{-1.7}$	14.03	13.50	12.82
2164338267900758400	21:07:49.1	+46:25:23	B:	-	910^{+430}_{-220}	$1.75^{+0.51}_{-0.47}$	$1.4^{+1.1}_{-1.3}$	$2.3^{+1.2}_{-0.3}$	13.09	12.91	12.46
2162489920134840960	21:09:18.1	+44:58:44	G:	Yes	2010^{+110}_{-100}	$0.98^{+0.18}_{-0.30}$	$0.04^{+0.41}_{-0.30}$	$2.42^{+0.35}_{-0.26}$	13.13	12.53	11.82
2200102018206234752	22:19:51.4	+58:08:53	B0Ve	Yes	3640^{+620}_{-470}	$4.17^{+0.05}_{-0.10}$	$-4.70^{+0.40}_{-0.39}$	33^{+4}_{-16}	13.02	12.28	11.44
2200100948745913600	22:20:10.1	+58:06:34	B1Ve	-	1900^{+1100}_{-500}	$3.51^{+0.10}_{-0.07}$	$-2.0^{+0.8}_{-1.1}$	$8.3^{+4.1}_{-1.6}$	13.40	12.82	12.10
2007840865221064192	22:24:57.3	+57:17:33	B	Yes	3040^{+340}_{-280}	$2.05^{+0.51}_{-0.47}$	$-1.98^{+0.68}_{-0.74}$	$5.0^{+2.5}_{-1.7}$	12.79	12.48	12.02
2013491427269041536	22:54:53.2	+59:09:33	B1Ve	Yes	2930^{+260}_{-220}	$4.09^{+0.10}_{-0.07}$	$-3.86^{+0.24}_{-0.28}$	$10.4^{+5.2}_{-1.1}$	13.32	12.56	11.72
2013491599067728896	22:54:56.7	+59:10:22	B3Ve	-	4130^{+370}_{-310}	$3.440^{+0.058}_{-0.033}$	$-3.28^{+0.20}_{-0.24}$	$6.81^{+0.70}_{-0.38}$	13.89	13.24	12.46
2015971616264073600	23:49:46.8	+62:42:37	B5Ve	Yes	3140^{+170}_{-150}	$2.864^{+0.067}_{-0.037}$	$-1.95^{+0.14}_{-0.18}$	$4.58^{+0.66}_{-0.26}$	13.93	13.40	12.73

Notes: 'SIMBAD' column lists the spectral types tabulated at the SIMBAD database at the time of writing. 'Astrom' column value is 'Yes' if the source has $\varpi/\sigma(\varpi) \geq 5$ and RUWE < 1.4. Distances from Bailer-Jones et al. (2018). See Chapter 4 for the derivation of the extinction, absolute magnitude and mass. A maximum of 50% relative error was set for the uncertainty of the mass.

Table D.3 Compiled and derived parameters of the 28 sources of the SIMBAD set in the Either catalogue.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
524688980302817408	01:17:20.9	+63:47:28	A0-1IIe	-	5800^{+1100}_{-800}	$3.79^{+0.14}_{-0.15}$	$-2.55^{+0.49}_{-0.53}$	$6.5^{+1.3}_{-1.0}$	15.98	15.08	14.11
519352324516039680	01:35:49.9	+66:12:43	B1Ve	-	6900^{+1600}_{-1100}	$4.95^{+0.10}_{-0.07}$	$-6.67^{+0.46}_{-0.54}$	34^{+17}_{-11}	13.51	12.47	11.50
457992330084893568	02:30:58.7	+57:15:04	B	Yes	2130^{+110}_{-100}	$1.91^{+0.51}_{-0.47}$	$-0.15^{+0.57}_{-0.62}$	$3.8^{+1.8}_{-1.3}$	13.68	13.40	12.97
465526596075681536	02:32:08.0	+61:22:04	k4V	Yes	$348.2^{+3.9}_{-3.8}$	$0.79^{+0.23}_{-0.28}$	$6.70^{+0.30}_{-0.25}$	$0.728^{+0.076}_{-0.052}$	16.03	15.19	14.32
465533571102525312	02:32:59.0	+61:25:45	A2	Yes	2200^{+100}_{-90}	$2.33^{+0.11}_{-0.14}$	$-0.67^{+0.23}_{-0.20}$	$3.57^{+0.33}_{-0.29}$	13.92	13.37	12.68
461168441217250048	02:59:50.8	+58:23:33	B5Ve	Yes	3000^{+230}_{-200}	$3.228^{+0.067}_{-0.037}$	$-1.50^{+0.19}_{-0.23}$	$4.82^{+0.50}_{-0.29}$	14.73	14.11	13.34
486469646880886656	03:35:38.0	+61:07:13	B	Yes	2210^{+100}_{-100}	$3.54^{+0.51}_{-0.47}$	$-1.56^{+0.57}_{-0.61}$	$5.1^{+2.6}_{-1.2}$	14.40	13.69	12.90
251868671929099776	03:58:13.8	+52:43:11	F0e	Yes	3970^{+430}_{-350}	$2.12^{+0.04}_{-0.14}$	$-1.05^{+0.35}_{-0.26}$	$4.54^{+0.51}_{-0.50}$	14.72	14.06	13.28
3425450307939737344	06:12:33.2	+23:49:42	B	Yes	1990^{+190}_{-160}	$2.03^{+0.51}_{-0.47}$	$-1.40^{+0.65}_{-0.70}$	$4.8^{+2.4}_{-1.2}$	12.37	12.12	11.61
3331395304451482752	06:25:18.1	+12:32:11	Be	-	4900^{+1000}_{-700}	$3.86^{+0.51}_{-0.47}$	$-3.38^{+0.81}_{-0.90}$	$9.0^{+4.5}_{-2.5}$	14.76	13.96	13.11
3326706020501406976	06:39:43.5	+09:36:04	K0	Yes	601^{+90}_{-70}	$4.168^{+0.079}_{-0.092}$	$4.31^{+0.36}_{-0.38}$	$1.24^{+0.21}_{-0.21}$	19.10	17.37	16.09
3119548688277389568	06:40:13.8	+00:07:13	B1V	-	6300^{+1100}_{-800}	$2.77^{+0.10}_{-0.07}$	$-3.08^{+0.38}_{-0.45}$	$10.5^{+5.2}_{-1.6}$	14.08	13.68	13.13
3125618714017476096	06:45:29.2	+00:55:42	G8III	-	4500^{+1500}_{-1000}	$0.325^{+0.024}_{-0.038}$	$-1.23^{+0.56}_{-0.65}$	$5.60^{+0.85}_{-0.60}$	12.83	12.37	11.75
3113995433003934080	06:55:05.1	+01:00:57	B	Yes	3020^{+540}_{-400}	$1.20^{+0.51}_{-0.47}$	$-1.70^{+0.78}_{-0.86}$	$5.3^{+2.7}_{-1.5}$	12.02	11.90	11.66
5543176193429801472	08:15:31.2	-35:46:08	B1e	Yes	3850^{+310}_{-270}	$2.81^{+0.10}_{-0.07}$	$-2.18^{+0.23}_{-0.27}$	$9.0^{+1.7}_{-1.6}$	13.96	13.56	12.99
4155000844481174656	18:33:27.8	-10:35:24	B0.5Ve	-	3900^{+2700}_{-1500}	$5.021^{+0.050}_{-0.050}$	$-6.3^{+1.1}_{-1.2}$	31^{+15}_{-14}	12.66	11.63	10.63
4318134628803970816	19:32:55.1	+14:13:37	B0-1I	-	5600^{+2600}_{-1700}	$4.55^{+0.10}_{-0.12}$	$-4.56^{+0.91}_{-0.92}$	$17.2^{+8.6}_{-5.5}$	14.46	13.73	12.67
2020091069715624960	19:43:28.1	+23:21:12	G0	Yes	583^{+12}_{-11}	$0.026^{+0.027}_{-0.057}$	$3.37^{+0.10}_{-0.07}$	$1.342^{+0.058}_{-0.042}$	12.54	12.22	11.77
1820874341793491840	19:45:05.5	+16:40:04	B7	-	11200^{+3000}_{-2200}	$0.973^{+0.032}_{-0.046}$	$-1.44^{+0.53}_{-0.54}$	$4.62^{+0.92}_{-0.72}$	14.88	14.79	14.57
2031789976512554496	19:45:33.7	+29:32:04	A0	Yes	2030^{+130}_{-110}	$0.41^{+0.10}_{-0.09}$	$0.33^{+0.21}_{-0.23}$	$2.56^{+0.19}_{-0.16}$	12.34	12.28	12.15
2067929686763411584	20:33:27.1	+41:35:58	Be	Yes	1630^{+130}_{-110}	$6.22^{+0.51}_{-0.47}$	$-4.01^{+0.62}_{-0.67}$	$11.0^{+5.5}_{-2.5}$	14.86	13.27	12.05

Table D.3 continued.

Gaia source id	RA h:m:s	DEC deg:m:s	SIMBAD SpType	Astrom	Distance (pc)	A_G (mag)	M_G (mag)	Mass ($M_\odot$)	G_{BP} (mag)	G (mag)	G_{RP} (mag)
2066322960974338816	20:40:36.9	+41:27:06	F7.5	Yes	$569.5^{+9.2}_{-8.9}$	$0.109^{+0.033}_{-0.033}$	$3.894^{+0.067}_{-0.067}$	$1.170^{+0.036}_{-0.014}$	13.09	12.78	12.36
2163392417338150912	20:54:34.4	+46:08:55	B	Yes	3740^{+350}_{-300}	$3.68^{+0.51}_{-0.47}$	$-3.68^{+0.65}_{-0.70}$	$9.9^{+5.0}_{-2.3}$	13.61	12.87	12.04
2163485875830774400	21:01:40.4	+46:06:52	B:	-	4530^{+860}_{-640}	$2.60^{+0.51}_{-0.47}$	$-3.10^{+0.80}_{-0.89}$	$8.2^{+4.1}_{-2.3}$	13.22	12.78	12.18
2161955630510182528	21:02:02.5	+43:30:46	B9.5Ve	-	4480^{+810}_{-610}	$4.44^{+0.06}_{-0.10}$	$-3.27^{+0.42}_{-0.42}$	$8.2^{+1.2}_{-1.0}$	15.53	14.43	13.41
2005907644593787264	22:18:45.8	+55:27:25	A3	-	-	$0.67^{+0.14}_{-0.06}$	-	-	12.55	12.56	12.06
2200101331011488640	22:20:13.1	+58:07:43	B0-1Ve	Yes	3170^{+440}_{-350}	$3.25^{+0.10}_{-0.12}$	$-2.97^{+0.37}_{-0.38}$	$10.7^{+5.3}_{-2.2}$	13.28	12.78	12.13
2014636431185097600	22:54:27.1	+60:48:52	B0Ve	Yes	3390^{+230}_{-210}	$5.41^{+0.05}_{-0.10}$	$-4.14^{+0.24}_{-0.19}$	$13.9^{+7.0}_{-0.3}$	15.06	13.92	12.87

Notes: 'SIMBAD' column lists the spectral types tabulated at the SIMBAD database at the time of writing. 'Astrom' column value is 'Yes' if the source has $\varpi/\sigma(\varpi) \geq 5$ and RUWE < 1.4. Distances from Bailer-Jones et al. (2018). See Chapter 4 for the derivation of the extinction, absolute magnitude and mass. A maximum of 50% relative error was set for the uncertainty of the mass.

It is reasonable to hope that in a not too distant future we shall be competent to understand so simple a thing as a star.

The Internal Constitution of the Stars, 1926

Sir Arthur Stanley Eddington