Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

Johanna Carolina Andrea Coronado Martínez
Born in Iquique, Chile

Oral examination: 29 July 2020

Small-Scale Structure of the Milky Way's Stellar Orbit Distribution
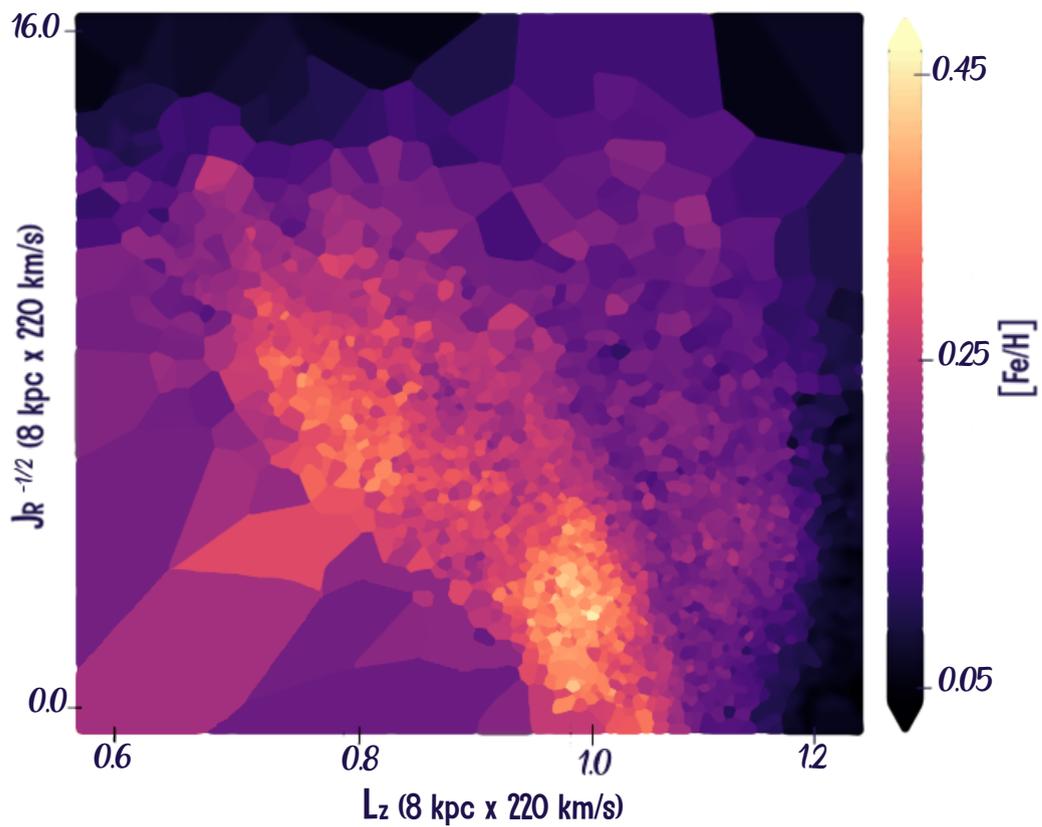
Referees: Prof. Dr. Hans-Walter Rix

Prof. Dr. Andreas Just

Abstract

The exact processes behind the formation and evolution of galaxies are interesting puzzles in modern astrophysics. Our Galaxy offers us the unique opportunity to be studied in detail, as we can obtain the 3D positions, 3D velocities and also the chemical information on a star-by-star basis. Different Galactic surveys have advanced in the effort of studying the Milky Way. The Gaia mission in particular provides the full 6D stellar position-velocity phase-space measurements for millions of its stars. By combining Gaia with chemical information from spectroscopic surveys, we can obtain a detailed physical picture of our Galaxy. In this thesis, we set out to investigate the stellar orbit distribution of the Milky Way, while also adding their chemical information ([Fe/H]) in a chemical tagging generalization approach. We first make use of the spectroscopic information from LAMOST, in combination with parallaxes and proper motions from Gaia. We develop a method to obtain improved spectrophotometric distances (with errors less than 6%) for 150 000 main sequence stars. With more precise distances at hand, we investigate the small-scale structure in the orbit distribution of the Galactic disc for $\sim$ 600 000 main sequence stars in LAMOST $\times$ Gaia. Most stars disperse from their birth sites and siblings, in orbit and orbital phase, becoming 'field stars'. We explore and provide direct observational evidence for this process in the Milky Way disc, by quantifying the probability that orbit similarity among stars implies indistinguishable metallicity. We define the orbit similarity among pairs of stars through their distance in action-angle space $\Delta(J, \theta)$ and their abundance similarity by Δ[Fe/H]. By grouping such star pairs into associations with a friend-of-friends algorithm linked by $\Delta(J, \theta)$, we find that hundreds of mono-abundance groups –some clusters, some spread across the sky– are over an order-of-magnitude more abundant than expected for a smooth phase-space distribution, suggesting that we are witnessing the 'dissolution' of stellar birth associations into the field. We finally explore a significantly larger sample of 6.2 million stars with radial velocities in Gaia, which is not limited to main sequence stars. Although this sample does not have [Fe/H] information, we are able to recover the same major groups found in the previous sample in both action and angle space. Moreover, we are able to identify other known associations by simple inspection, opening up the possibility for this method to be applied to further characterize dissolving associations across the Galaxy.

Zusammenfassung

Die genauen Prozesse hinter der Entstehung und Entwicklung von Galaxien sind noch offene Rätsel in der modernen Astrophysik. Unsere Galaxie bietet die einzigartige Möglichkeit, sie im Detail zu untersuchen, da wir die 3D Positionen, die 3D Geschwindigkeiten und auch die chemischen Informationen auf einzelner Sternenbasis erhalten können. Verschiedene galaktische Durchmusterungen sind bei den Bemühungen zur Untersuchung der Milchstraße vorangekommen. Insbesondere die Gaia-Mission liefert die vollständigen 6D Sternpositions-Geschwindigkeits-Phasenraummessungen für Millionen seiner Sterne. Durch die Kombination von Gaia mit chemischen Informationen aus spektroskopischen Untersuchungen können wir ein detailliertes physikalisches Bild unserer Galaxie erhalten. In dieser Arbeit wollen wir die Verteilung der Sternenbahnen in der Milchstraße untersuchen und gleichzeitig ihre chemischen Informationen ([Fe/H]) in einem Verallgemeinerungsansatz mit chemischen Markierungen hinzufügen. Zunächst nutzen wir die spektroskopische Information von LAMOST in Kombination mit Parallaxen und Eigenbewegungen von Gaia. Wir entwickeln eine Methode, um verbesserte spektrophotometrische Abstände (mit Fehlern von weniger als 6%) für 150 000 Hauptreihensterne zu erhalten. Mit den vorliegenden, präziseren Entfernungen untersuchen wir die kleinskalige Struktur in der Bahnverteilung der galaktischen Scheibe mittels $\sim$ 600 000 Hauptreihensterne in LAMOST $\times$ Gaia. Die meisten Sterne lösen sich von ihren Geburtsorten und Geschwistern in der Umlaufbahn und in der Orbitalphase auf und werden zu 'Feldsternen'. Wir erforschen und liefern direkte Beobachtungen für diesen Prozess in der Milchstraßenscheibe, indem wir die Wahrscheinlichkeit quantifizieren, dass Bahnähnlichkeit zwischen Sternen eine nicht unterscheidbare Metallisierung impliziert. Wir definieren die Bahnähnlichkeit zwischen Sternenpaaren durch ihren Abstand im Aktionswinkelraum $\Delta(J, \theta)$ und ihre Häufigkeitsähnlichkeit durch Δ[Fe/H]. Durch Gruppierung solcher Sternenpaare in Assoziationen mit einem Freund-von-Freunden Algorithmus, der durch $\Delta(J, \theta)$ verknüpft ist, finden wir Hunderte von Mono-Häufigkeitsgruppen –einige Haufen, einige über den Himmel verteilt– in einer Größenordnung, die für eine glatte Phasen-Raum-Verteilung reichlicher sind als erwartet, was darauf hindeutet, dass wir die 'Auflösung' von stellaren Geburtsassoziationen in das Feld beobachten. Schließlich untersuchen wir eine deutlich größere Probe von 6.2 Millionen Sternen mit Radialgeschwindigkeiten in Gaia, diesmal nicht auf Hauptreihensterne beschränkt. Obwohl diese Probe keine [Fe/H]–Information enthält, sind wir in der Lage, die gleichen Hauptgruppen, die in der vorherigen Probe gefunden wurden, sowohl im Aktions– als auch im Winkelraum wiederzufinden. Darüber hinaus sind wir in der Lage, andere bekannte Assoziationen durch einfache Inspektion zu identifizieren, was die Möglichkeit eröffnet, diese Methode zur weiteren Charakterisierung der sich auflösenden Assoziationen in der gesamten Galaxie anzuwenden.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

Marie Curie.

Acknowledgements

I have been struggling to decide where to put the acknowledgements of this thesis: at the beginning or at the end? But, I think it is better to keep them at the beginning, otherwise people won't read them, and I have reached this point thanks to the help of many people that I want to thank.

First to my supervisor, Hans-Walter for your infinite patience, for understanding when I was struggling and immediately going to the blackboard in your "full-professor-mode", explaining things step by step. I don't have enough words to express how grateful I am for your generosity. Thanks to Wilma Trick, my "PhD sister" who has been like a second supervisor to me, with your guidance and advice constantly helping me to improve the quality of my research. I am also lucky to call you my friend. To Andy Gould: thanks for our conversations (about science and life) and for showing me how to be a better scientist, giving me your very valuable input. Thanks also to everyone at the MPIA Milky Way group for our friendly group meetings, special thanks to Sven Buder, Jan Rybizki, Morgan Fouesneau and Coryn Bailer-Jones. To IMPRS and to Christian Fendt, for allowing me to come to Heidelberg, where I have encountered great scientists and wonderful people. I will be forever grateful to the IMPRS program that offers the opportunity to meet people from all over the world.

To all the friends that I have made in these almost four years. It is hard to be far away from your family (12 000 kms!) but it becomes a bit easier when you have amazing people around you: Josha my first nemesis and office mate, thanks for making my days happier with your random laughs in the office. Arianna, the most positive person I have met in my life, it's impossible to be sad around you. Asmita thanks for your kindness and for being the "organized one" so we can always follow your lead. Also, thanks for preparing amazing indian food for us. Ivana, we had such good times. I will always remember Hallhuber, when we dressed like twins. Christos/Carlos, you have the best sense of humor, thanks for all the knowledge you share with us as "the ancient wise one". Sepideh, thanks for being such a sweet person and for your friendship all of these years. Melanie: thanks for making the best cakes and to share them with me, and also for translating the abstract. To "the best office at MPIA": Christian, Christoph and Manuel, it was a short time, but I will always remember our afternoon tea/coffee/gossip time. To all the members (past and present) regularly attending student coffee: Alex, Sara, Hector, Christina, Neven, Yulong, Johannes, Paula, Giancarlo, Camille, Jacob, Irina, Oliver, Vince (and I hope I am not forgetting anyone). Thanks for the conversations, laughs and all of the good times we had.

To my chilean friends here in Heidelberg: Jorge, María José, Mauricio and Natalia. Thanks to you guys Heidelberg now feels like home. To my chilean friends all over the world: Cynthia, Camila, Diego, Francisco, Carolina, Claudia, Scarlet, Rodrigo L., Rodrigo C., thanks for your friendship throughout all of these years and constant support. To the HITS gang, Freeke and Rüdiger thanks for the amazing times we spent playing Pandemic Legacy. Also thanks to Sebastián for cooking amazing Colombian food and for sharing your love of board games with me and Felipe.

To Leonardo Vanzi and Krzysztof Helminiak who helped and guided me during the early stages of my scientific career. Also Luca Sbordone, Sonia Duffau and José Miguel Fernández for your advice and friendship. Thanks also to Julio Chanamé for your guidance during my master studies.

During my PhD I had the opportunity to visit La Silla many times. I want to thank everyone there, for always being super nice and welcoming. I had the best times observing there. Special thanks to Javier Alarcón and Mauricio Martínez for all their help and conversations we had during the many long observing nights.

To my parents and brother in Chile, for their constant love and encouragement and for cheering me up in every single step. Giselle, Ángela and Adrián, thanks for your love and welcoming me into your wonderful family.

To my husband Felipe. Thanks for coming with me to Heidelberg so I could pursue my dream. For your love, support and for believing in me, even though I was constantly undermining myself, especially during the last months of my PhD. Thanks for proof reading this thesis, I know it was a difficult task!. Eres el mejor.

To everyone at MPIA. This is the best place to pursue a PhD, and it is not that I am biased (well, maybe I am). I had the best time during these ~ 4 years and I will cherish these memories forever.

Additional thanks to Andreas Just, Eva Grebel and Kurt Roth for agreeing to be part of my committee.

Contents

List of Figures

List of Tables

Abbreviations

MW: Milky Way

DM: Dark Matter

MS: Main Sequence

DR: Data Release

WB: Wide Binaries

RV: Radial Velocity

CMD: Color Magnitude Diagram

TGAS: *Tycho – Gaia* Astrometric Solution

GDR1   Gaia Data Release 1

GDR2   Gaia Data Release 2

LAMOST: Large Sky Area Multi-Object Fibre Spectroscopic Telescope

GALAH: GALactic Archaeology with HERMES

GES: Gaia-ESO survey

RAVE: RAdial Velocity Experiment

APOGEE: Apache Point Observatory Galactic Evolution Experiment

Para mi mamá, María Carolina Martínez.

1 | Introduction

Most likely all of us have gazed into the night sky to look at the stars. Some of us have done it in the backyard with a small telescope, and seen beyond what is visible to the naked eye. Growing up I had the privilege of observing the night sky in the Atacama desert, the perfect place to look at the sky without pollution from city lights. Although I did not know it at the time, many of those luminous points that looked like stars were actually deceiving me. I was also looking at galaxies: collections of millions or billions of stars. Galaxies are extended systems composed of dust, stars and dark matter. All of which is being held together by gravity.

The deeper we look into the night sky, the more galaxies we see. The observable universe contains $\sim 10^{12}$ galaxies (Conselice et al., 2016). We now know that some of them are very similar to our own Galaxy, the Milky Way (MW), while others are quite different. But, before the 20th century, we did not know the existence of other galaxies besides our own. In the past, astronomers had classified other galaxies as *nebulae* based on their fuzzy, diffuse structures. For example, Andromeda was classified as the *Andromeda Nebula* and spiral galaxies as *spiral nebulae*. The study of galaxies remained more or less descriptive until the late 20th century with the development of technology and the advent of photography. These allowed to study the morphologies and structures of external galaxies, and the development of classification schemes. The most notable one done was by Hubble (1926), and later extended by Holmberg (1958), and van den Bergh (1960), amongst others. Additionally, de Vaucouleurs (1959) included bars, rings and other features to the classification.

Ultimately, we need to identify if these structural features are useful to understand galaxies and their formation history. It was Holmberg who established that the physical properties of nearby galaxies correlate with morphology, elliptical galaxies are typically massive, red and with minimal star formation, while spirals tend to be less massive, bluer and showing evidence of ongoing star formation (Holmberg, 1958). Later on, quantitative measurements of the light distribution in galaxies were possible thanks to charged coupled devices (CCDs). This resulted in what today we know as the de

Vaucouleurs profile, where it was identified that massive ellipticals all roughly follow the same light distribution (de Vaucouleurs, 1948). A generalization of this profile was later done by Sérsic (1963), with discs following an exponential light profile. Galaxies were then decomposed into bulge and disc components, with additional features such as bars and rings. Investigation of the three dimensional structure of disc galaxies followed, as well as detailed studies of bulges and discs in spiral galaxies.

The morphology and structure of galaxies in the nearby Universe have been extensively investigated (e.g., Kormendy et al., 2009), essentially because this allows us to obtain clues on their formation and subsequent evolution. For instance, the Hubble sequence was one of the first steps towards having insight on galaxy evolution. At the time, this scheme was considered an evolutionary track, in which spiral galaxies are young, whereas elliptical galaxies are old. Although today we know this is an over simplification, the basic ideas still hold.

Within the current cosmological model, structures form hierarchically; small overdensities collapse and the resulting dark matter (DM) haloes merge to form larger ones which serve as sites of galaxy formation. Thus, undoubtedly, galaxies are the fundamental building blocks of the Universe. Consequently, understanding the formation and evolution of galaxies is of paramount importance, and it is a topic under active research. These are also complex and continuous processes: the structure and composition of galaxies are shaped over billions of years by interactions, collisions or mergers with other galaxies and groups of stars.

Galaxies are the fundamental systems in which stars organise. They have characteristic sizes ($R_{\mathrm{gal}} \sim$ kiloparsec) and masses ($M_{\mathrm{gal}} \sim 10^{10}$ M$_\odot$) (Naab & Ostriker, 2017). In this context, we start by focusing on a specific type of galaxy.

1.1 Disc galaxies

The general picture of galaxy formation we have today is from 40 years ago (White & Rees, 1978; Fall & Efstathiou, 1980). Within the context of the cold DM paradigm, gravity assembles structures in a bottom up fashion; small structures form first, then grow and merge into larger ones. In this scheme, galaxies are formed through the cooling of gas at the centre of DM haloes, which then condense to form stars. In addition to gravity, cooling or dissipation processes are invoked in order for this framework to reproduce the features we observe in galaxies and galaxy clusters. The stars and gas, which account for the luminous (baryonic) matter in galaxies, are mixed with and embedded in non-baryonic and non-relativistic DM, which dominates the total mass of the galaxy and its halo.

In the early universe DM and gas acquire angular momentum through tidal torques and mergers (Peebles, 1969). At this point is when the conservation of angular momentum of the cooling gas within DM haloes could lead to the formation of a rotationally supported galactic disc (Fall & Efstathiou, 1980). In this disc, star formation will start to take place at a rate determined by the local surface density and the dynamical timescale of the disc (Schmidt, 1959; Kennicutt, 1989). Different mechanisms will begin to transform cold into hot gas: stars exploding into supernovas, or winds from massive stars. Feedback is a critical process affecting galaxy evolution, but currently it is a poorly known process.

Overall, disc galaxies show a collection of different morphologies. Although certain properties seem to be common to most disc galaxies (e.g., flat rotation curves), there are some other key properties that differ, such as the surface brightness and scale length (Kautsch et al., 2006).

The different components of disc galaxies will retain different kind of signatures of their formation (Freeman & Bland-Hawthorn, 2002). Generally speaking, the main components of these galaxies are stars, dust and cold gas (atomic and molecular), most of them forming the galactic disc. These stars and gas have a mean metallicity that depend on the luminosity of the galaxy and often show a radial gradient (Zaritsky, Kennicutt, & Huchra, 1994). Additionally these galaxies show the presence of spiral arms (Kennicutt, 1981), commonly a central bar (a fraction of 30% in optical, 70% in infrared, Kruk et al., 2018), and also a bulge. The latter component is present in most of the more luminous disc galaxies, whereas the fainter ones do not show one (Freeman & Bland-Hawthorn, 2002). When observing a disc galaxy face-on we can see that the spiral arms consist mainly of young stars, HII regions, molecular gas and dust (Young & Scoville, 1991). Edge-on galaxy observations have provided a unique opportunity to study the vertical structure of galaxies (Bizyaev et al., 2014), revealing two components: the thin and thick disc (Kautsch et al., 2006). There is also evidence that disc galaxies have a spheroidal halo extending out to large radii. Figure 1.1 shows different examples of face-on spiral galaxies and their shapes.

1.1.1   Properties of disc galaxies

In this section we will mention some of the most relevant properties of disc galaxies.

Surface brightness profiles: An outstanding property of disc galaxies is that the brightness profiles of their stellar discs are very close to exponential distributions over a large range in radii (Freeman, 1970; de Jong, 1996; Ferguson & Clarke, 2001). Thus, an exponential luminosity profile is usually used to model their light distributions.

FIGURE 1.1: Different spiral galaxies observed with the Hubble Space Telescope. On the top left is M101 that has almost twice the diameter of the MW. On the top right is M74 or NGC 628, and it is an example of a 'Grand Design' spiral galaxy. Both galaxies are observed nearly face-on. On the lower left is M83 that shows different star clusters and also supernova remnants. The pink color is due to the absorption of the light from young stars, by diffuse gas clouds. And on the lower right is M100, also a Grand Design-spiral galaxy, showing two prominent spiral arms with young blue stars.

Image credits: Hubble Image: NASA, ESA, K. Kuntz (JHU), F. Bresolin (University of Hawaii), J. Trauger (Jet Propulsion Lab), J. Mould (NOAO), Y.-H. Chu (University of Illinois, Urbana), and STScI; CFHT Image: Canada-France-Hawaii Telescope/ J.-C. Cuillandre/Coelum; NOAO Image: G. Jacoby, B. Bohannan, M. Hanna/ NOAO/AURA/NSF, Judy Schmidt, the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration, AURA,R. Chandar (University of Toledo) and J. Miller (University of Michigan), W. Blair (STScI/Johns Hopkins University) and R. O'Connell (University of Virginia)

In order to take into account the light contribution from both disc and bulge, bulge-disc decompositions of the surface brightness are usually applied (e.g., de Jong, 1996; MacArthur, Courteau, & Holtzman, 2003). In general terms, the final surface brightness distribution of disc galaxies is a superposition of an exponential profile for the disc and a Sérsic profile for the bulge.

Disc vertical structure: Galaxy discs are not infinitesimally thin. The vertical distribution of stars in edge-on disc galaxies can be described by an isothermal sheet, following a $sech^2$ law, with a vertical scale-height that is, to an approximation, independent of galactocentric radius (van der Kruit & Searle 1981; van der Kruit & Freeman 2011; however, recent observations of our own Milky Way show it might not be independent, Bovy et al. 2016a). This component is known as the *thin disc*.

Perpendicular to the disc plane, the stellar density (or luminosity profile) reveals an excess at distances $z \gtrsim 1$ kpc (Burstein, 1979; Tsikoudi, 1980). This exponential excess

of light at large distances above the disc plane is observed in many edge-on disc galaxies (Dalcanton & Bernstein, 2002). This second component is known as the *thick disc*. Almost all galaxies have thick discs, and their scale heights are ~ 2 times larger than the thin discs and thick discs have systematically larger scale lengths than thin discs (Yoachim & Dalcanton, 2006).

Spiral arms: The majority of disc galaxies exhibit some form of spiral arms (Sellwood, 2011), showing a wide variety of spiral structure as mentioned in de Vaucouleurs' classification system (de Vaucouleurs, 1948). These structures occasionally have regular and symmetric patterns, commonly described as 'Grand Design' spirals. Others have no clear symmetry or pattern, or with short spiral arms that are called 'flocculent' spirals, as is the case of the nearby galaxy M33 (Humphreys & Sandage, 1980; Dobbs et al., 2018). The more coherent patterns are normally seen in galaxies with a bar or that have recently suffered a tidal interaction with a passing companion galaxy (e.g., M51 and its neighbour NGC 5195, Kormendy & Norman, 1979; Salo & Laurikainen, 2000; Kendall, Kennicutt, & Clarke, 2011).

Additionally, spiral arms are the site of young stars, molecular clouds and HII regions, thus they are often environments of active star formation (Elmegreen & Elmegreen, 1983; Ferguson et al., 1998; Beuther et al., 2017). What it is unclear is whether spiral arms trigger star formation, or if they simply 'rearrange' young stars or molecular clouds in the galaxy (Bonnell & Dobbs, 2007).

In the past, the main theory for the origin and recurrence of spiral arm features was focused on the presence of density waves in the discs (Lindblad, 1963; Lin & Shu, 1964), where spiral arms are the product of quasi-steady global modes. However, these long lived spiral modes could not be sustained naturally within a disc (Toomre, 1969). Since then, the effects of gravity and disc dynamics have been considered, with spiral arms being the result of gravitationally driven density waves (Toomre, 1981; Sellwood & Carlberg, 1984). In general, numerical simulations do not manage to reproduce the long lived density wave structure, where the spiral arm pattern does not survive more than a few disc rotations (Sellwood, 2011). Thus, arms could be short lived, transient structures resulting from recurring gravitational instabilities (Ragan et al., 2018). However, we do not have direct evidence for this, as it is not possible to observe the time evolution of real galaxies. Consequently, we currently rely on the results of numerical simulations that of course have their limitations.

Bars: A large fraction of disc galaxies have stellar bars. These are elongated structures crossing the face of a galaxy. In the local universe (roughly speaking, closer than ~ 50 Mpc) the fraction of disc galaxies with the presence of bars is $\sim 2/3$ (or $\sim 1/3$ when only considering strongly barred systems, Sheth et al., 2008). For a while now, edge-on disc

galaxies seem to be associated with boxy or peanut shaped bulges (Combes et al., 1990). Once a bar forms, a boxy-peanut bulge will likely form soon after (Martinez-Valpuesta, Shlosman, & Heller, 2006; Fragkoudi et al., 2017). This connection could suggest that bulges are produced by the evolution of galactic bars. Numerical simulations show that bulges form following the dynamical instability of bars (Portail, Wegg, & Gerhard, 2015).

Stellar Haloes: In the cosmological framework of galaxy formation, stellar haloes surrounding the galaxies are the natural outcome of the galaxy evolution process (Eggen, Lynden-Bell, & Sandage, 1962; Steinmetz & Muller, 1995; Bekki & Chiba, 2001). These extended and diffuse stellar structures were assembled by accreting disrupted satellites along the cosmic time (Bullock & Johnston, 2005). The detection of haloes in other galaxies is extremely challenging as these are very faint, with low surface brightness, typically $\gtrsim 7$ mag fainter than the sky (Zibetti & Ferguson, 2004). Stellar haloes in galaxies other than the Milky Way have been detected (e.g, in M33, McConnachie et al. 2006; however Ferguson et al. 2007 excluded the presence of a significant stellar halo in this galaxy). Currently the fraction of disc galaxies that contain a stellar halo it is not clearly known.

Kinematics: Stars and cold gas move in the disc plane on nearly circular orbits. The kinematics of disc galaxies is commonly characterized by the rotation curve $V_{\rm rot}(R)$, which expresses the rotation velocity as a function of galactocentric distance. The shape of the rotation curve is an important component to model the disc's formation and evolution (Bovy et al., 2012). These rotation curves rise steeply in the inner regions and remain roughly flat in the outer parts (Rubin, 1983), exposing the presence of DM in disc galaxies. Disc rotation curves can be measured using a variety of techniques, such as spectroscopy of HII region emission lines (Brand & Blitz, 1993), RR Lyrae (Wegg, Gerhard, & Bieth, 2019), blue horizontal branch stars in the halo (Xue, Rix, & Zhao, 2009), red giant branch and red clump stars in the Galactic disc (Bovy et al., 2012), or masers in high-mass star-forming regions (Reid et al., 2014).

Thus far, we have given a general description and overview of galaxies, but there has been a significant shift in our overall picture for galaxy evolution over the past few years. A combination of large imaging surveys, detailed kinematic studies, and theoretical studies have shown that galaxies live in a state of equilibrium where their ability to form stars is regulated by how much gas is available, and the predominance of outflows rather than by successions of minor and major mergers (Tacconi et al., 2013; Saintonge et al., 2013; Lilly et al., 2013). The resolved structures of galaxies also allow to measure their internal features and how they are assembling. Bulges, discs and bars most likely formed by secular processes produced internally by disc dynamical evolution (Conselice, 2014). Internal and external mechanisms are involved in the disc and bulge evolution (Sachdeva

et al., 2015). State of the art cosmological simulations of Milky Way sized galaxies have shown that the disc forms 'inside out' in a radial sense, and 'upside down' in a vertical sense (Bird et al., 2013; Grand et al., 2016). In these simulations, star particles are born on orbits that become kinematically cooler, leading to a disc that cools and becomes thinner with time. Additionally, galaxy observations are consistent with inside-out growth of discs, where disc galaxies of a given stellar mass appear smaller at higher redshift (van Dokkum et al., 2013; van der Wel et al., 2014; Rodríguez-Puebla et al., 2017), and resolved stellar observations of Local Group galaxies using the Hubble Space Telescope have shown that old populations are usually more centrally concentrated than young populations (Sacchi et al., 2019). Over the next years, as more telescopes start to operate (such as the James Web Space Telescope, Euclid and the Large Synoptic Survey Telescope, among others), many more galaxies will be able to be spatially resolved. These galaxies will also be at higher redshifts, opening up new possibilities for their study.

1.2   The Milky Way as a *model organism*

Thanks to our privileged position within our Galaxy, we can obtain kinematics, 3D positions and 3D velocities (v_{los}, μ_{ra}, μ_{dec}), and also chemical information on a star by star basis. This means that we can study the MW in detail, which should lead us to a better understanding of the mechanisms involved in the formation and evolution of disc galaxies.

The MW seems to be a relatively typical spiral galaxy (Mo, van den Bosch, & White, 2010; Bland-Hawthorn & Gerhard, 2016). Its main stellar component is the disc, which is a flattened structure with a mass of $M_{\mathrm{MW}} = 2.1 \times 10^{11} M_\odot$, estimated from the kinematics of halo globular clusters within 21.1 kpc (Watkins et al., 2019). It also has an exponential radial scale length of ~ 3 kpc (McMillan, 2011a), and a scale height of ~ 0.3 kpc (López-Corredoira et al., 2002). The Sun sits close to the mid-plane of the disc, at ~ 8 kpc from the Galactic center rotating around the center of the Galaxy with $v_{\mathrm{circ}}(R_\odot) \sim 229$ km/s (Eilers et al., 2019). The disc is said to be kinematically cold, because the velocity dispersion of the stars near the Sun ($\sigma_z \sim 25$ km/s) is smaller than the circular velocity (Rix & Bovy, 2013). In the MW, the typical disc metallicity [Fe/H] is near that of the Sun (van der Kruit & Freeman, 2011).

As already discussed, it has been reported that most spirals, and that includes our Galaxy, have a second component which is thicker and surrounds the thin disc (van der Kruit & Freeman, 2011). In the MW, this thick disc was reported as an overdensity of stars at large distances from the Galactic plane (Yoshii, 1982; Gilmore & Reid, 1983), distinct from both halo and thin disc components (Jurić et al., 2008). The thin and thick disc
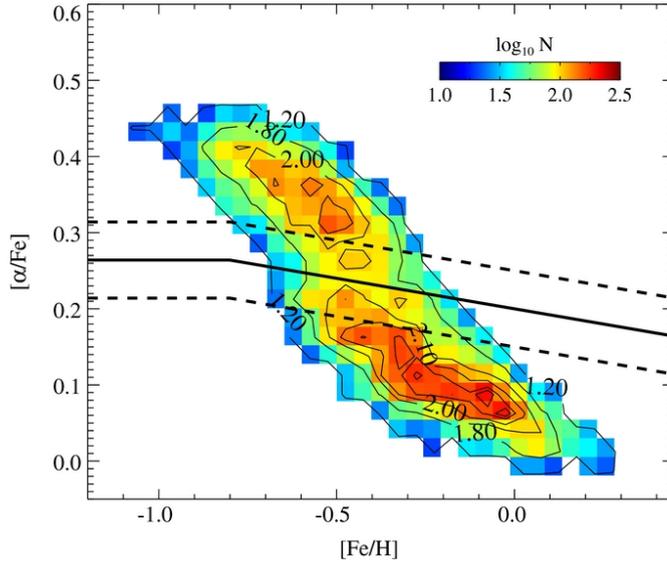
FIGURE 1.2: Example of the distribution of stars in the [α/Fe] vs. [Fe/H] plane for dwarfs stars in SEGUE, from Lee et al. (2011). The solid line indicates the division into likely thin-and thick-disc populations; the dashed lines located $\pm$ 0.05 dex in [α/Fe] on either side of the solid line indicate the adopted dividing points for the high-[α/Fe] (upper-dashed) and low-[α/Fe] (lower-dashed) stars in their sample.

have shown metallicity distributions that seem to overlap but that differ, at given [Fe/H], in their kinematics, age and α content (Navarro et al., 2011). It has been found that the thick disc is composed of relatively old stars (e.g., Bensby, Feltzing, & Lundström, 2003; Fuhrmann, 2008; Adibekyan et al., 2011), metal-poor and α-enhanced (e.g., Fuhrmann, 1998; Fuhrmann, 2008; Feltzing, Bensby, & Lundström, 2003; Haywood, 2008), moving in Galactic orbits with a large scale height and long scale length (e.g., Jurić et al., 2008; Kordopatis et al., 2011).

The separation in these two components, the traditional thin and thick disc, has arisen as an attempt to understand the observed spatial distribution, kinematics and chemical properties of the stars in the disc. The thick and thin discs show up as a bimodal distribution in the [α/Fe]–[Fe/H] plane, with stars divided into 'high-α' and 'low-α' sequences (e.g., Lee et al., 2011; Navarro et al., 2011), as can be seen in Fig. 1.2. However, this separation, or other possible observed trends, will depend on the criteria applied to divide stars into two disc populations (Fuhrmann, 2008; Lee et al., 2011; Bovy, Rix, & Hogg, 2012). For example, Bovy, Rix, & Hogg (2012) found that the mass weighted scale-height distribution varies smoothly when investigating individual elemental-abundance (e.g., [α/Fe] and [Fe/H]), finding that the MW has no distinct thick disc. This points to an early study from Norris (1987), where the thick disc component in the MW was thought to be the tail of a continuous and monotonic scale-height distribution.

Ultimately, finding answers to the thin-disc division, and the observed bimodality in the [α/Fe]–[Fe/H] plane, plays an important role in our understanding of how the

Galaxy formed and evolved. In external galaxies to separate the disc in sub-components may seem adequate, as we can only characterize their general properties. However, in the MW we have the opportunity to observe in great detail the rich information that is present in the disc. For instance, Hayden et al. (2017) have found that the two sequences observed in the [α/Fe] plane for stars in the solar neighborhood is presumably a reflection of their birth radius, rather than a thin or thick disc division. This bimodality is then in the abundances, but no the structure of the disc, which may seem discordant with observations of external edge-on galaxies, where thick disc components are found to be universal (Yoachim & Dalcanton, 2006). However, this is only showing that decomposition in external galaxies has its limitations.

In addition to the disc component, the inner region of the MW contains a central bulge/bar component, with a boxy-peanut shape revealed from photometry and spectroscopic studies (Dwek et al., 1995; Howard et al., 2009; McWilliam & Zoccali, 2010; Wegg & Gerhard, 2013; Ness & Lang, 2016). The estimated mass of this structure is $\sim 10^{10}$ M$_\odot$ (Valenti et al., 2016). The bar extends to a galactocentric radius of ~ 3 kpc, with its longest axis inclined by about 20 degrees with respect to the line from the Sun to the Galactic center (Bissantz & Gerhard, 2002).

Another component in our Galaxy is the Galactic halo. Its underlying and dominant DM has not been directly observed, but its presence can be inferred from different dynamical tracers experiencing its gravitational effect. Some of them are: the kinematics of stellar streams, particularly the Sagittarius stream (Law, Johnston, & Majewski, 2005; Gibbons, Belokurov, & Evans, 2014), the Magellanic Clouds (Busha et al., 2011), high-velocity stars (Piffl et al., 2014; Contigiani, Rossi, & Marchetti, 2019), and the kinematics of globular star clusters and dwarf galaxies, amongst others. The halo also contains the most metal-poor stars in the Galaxy and probably some of the oldest ones. Therefore, it provides us with a picture of the Galaxy in its very early stages of evolution. It extends to over 200 kpc (Zaritsky, 1999), and despite these multiple tracers, there is no consensus in the literature on its total mass (Wang et al., 2015). One estimation is $\sim 10^{12}$ M$_\odot$ (Wilkinson & Evans, 1999). Reports on the halo shape range from spherical (Bovy et al., 2016b) to oblate (Loebman et al., 2014), or prolate (Posti & Helmi, 2019).

Finally, we know the Galaxy has spiral arms based on different tracers (e.g. CO, HI, hot dust; Vallée, 2014). However, its detailed spiral structure has not been well revealed, due to difficulties in obtaining accurate distances of these tracers (Vallée, 2017; Xu, Hou, & Wu, 2018). The MW could have two or three arms, or even more complicated structures (Vallée, 2008).

1.2.1   What questions are to be answered?

Up to this point we have described the structure of spiral galaxies, with an emphasis in our own Galaxy, the MW. Understanding the formation of our Galaxy is still, however, a difficult problem to tackle. For example, we still don't know what are the processes that determine Galactic disc structure, in particular the radial and vertical profiles observed in their stellar distributions. How were the vertical motions of stars acquired? A recent study by Ting & Rix (2019) points towards gradual orbit scattering (e.g., from giant molecular clouds) as the dominant source of disc vertical heating, at least for $R \lesssim 14$ kpc and over the last 8 Gyrs.

Were stars mostly formed in a well-settled disc, or some of them formed in turbulent perturbed discs? (Ceverino et al., 2012). What is the role of feedback from star formation? The latter is thought to be a crucial ingredient for models of the formation and evolution of galaxies. Star formation becomes much more efficient than observed with no feedback, especially in low-mass galaxies (White & Frenk, 1991). However, feedback it is still a poorly understood process, as there is no general consensus on how it must be implemented in numerical simulations (Marasco et al., 2015).

Has the infall of satellites played a role in the formation of a thick disc, and did they deposit debris in the process? (Villalobos & Helmi, 2008). What is the role of radial migration (Sellwood & Binney, 2002; Roškar et al., 2008; Schönrich & Binney, 2009; Minchev et al., 2011). Is radial migration induced by satellite infall (Bird, Kazantzidis, & Weinberg, 2012) or produced by internal processes? These questions are not only relevant for the MW, but also to explain discs in external galaxies.

1.3   Secular orbit evolution

The story of Galaxy formation is incomplete not only for the MW, but for external galaxies as well. To begin with, galaxies do not have two temporally distinct phases of formation and posterior evolution. The processes involved during the early times of galaxy formation were rapid and violent, driven by a combination of dissipative collapse (Eggen, Lynden-Bell, & Sandage, 1962) and mergers (Toomre, 1977), which occur roughly during the first third of the life of the Universe. During this period, the evolution timescale was short, given by the dynamical (or free-fall) timescale t_{dyn}, with $t_{\mathrm{dyn}} \sim (1/G\rho)^{1/2}$, where ρ is the mean density and G is the gravitational constant. The evolution during the remaining two thirds is mostly quiescent and secular (Kormendy & Kennicutt, 2004; Sellwood, 2014). This is the result of the Universe expanding, where as

galaxy clusters virialize and start gaining large internal velocities, then mergers get less common (Toomre, 1977; Conselice et al., 2003).

For the majority a galaxy's life, internal secular processes are dominant. These slow processes produce a gradual restructuring of a galaxy with timescales much longer than t_{dyn}. Additionally, these are important for galaxy evolution, as internally-driven processes must have contributed significantly to the galactic properties we observe in present time. For example, the fact that galaxy discs are thin suggests a continuous period of quiescent evolution, during which a number of processes, such as slow accretion of gas, the presence of the bar, or the spiral arms, could have a large effect on changing and rearranging its structure (Kormendy & Kennicutt, 2004; Sellwood, 2014). However, for secular evolution to have a significant effect, a galaxy must have not experienced major mergers for a long time, as these will erase the signature of secular processes.

One of the most important agents of secular evolution are the spiral arms (Sellwood, 2014), as they redistribute angular momentum and subsequently cause stars to increase their random motions over time. They also cause extensive radial mixing of the gas and stars as they smooth small-scale irregularities in the mass distribution. Bars could also cause similar secular changes. Even though bars show no tendency to evolve in isolated gas-free discs (Miller & Smith, 1979), interaction with gas and other mass components of the galaxy can gradually alter its properties, with evolutionary consequences for the galaxy (Conselice et al., 2003).

This radial mixing of stars and dust show that even in a quiescent regime, a star's present-day orbit may not necessarily reflect its birth orbit, as shown by Sellwood & Binney (2002). In addition, stellar feedback in a galaxy is expected to cause the metallicity of the interstellar medium (ISM) to increase, consequently stars formed at the same radius would have higher metallicities. Moreover, it has been established that the ISM metallicity decreases with increasing radius (Daflon & Cunha, 2004), which in turn would result in coeval stars being progressively more metal-poor as the radius increases. Thus, without stellar migration, we would see a perfect correlation between the ages and metallicities of stars.

This is not the case, however, as it has been observed that stars of a given age show a broad spread in metallicity in the solar neighborhood (Edvardsson et al., 1993; Haywood, 2008; Casagrande et al., 2011). We will discuss this process of radial migration in more detail in the next section.

1.3.1 Radial migration

The discrepancies in the solar neighborhood age-metallicity relation requires either that the metallicity of the ISM was much less homogeneous in the past than it is today, which is probably unlikely, or that stars have have moved away or migrated from their birth radii (Sellwood & Binney, 2002; Haywood, 2008; Schönrich & Binney, 2009; Minchev & Famaey, 2010). Then, effective radial migration or redistribution of angular momentum must be taking place in the MW disc.

Changes in angular momentum can arise only from non-axisymmetries, and Sellwood & Binney (2002) established that efficient radial mixing of stars in galactic discs was caused by resonant interactions with transient spiral waves. Specifically, this implies that even for a star born on a circular orbit, its present-day radius could differ from its birth radius because of orbital heating –changes in the vertical direction increasing the epicycle– or *blurring*. Additionally, there is another process, dubbed *churning*, when the guiding-centre of the orbit changes without changing the angular momentum by interactions with the spiral arms and without causing much *blurring*. Minchev et al. (2011) have also shown that a strong exchange of angular momentum occurs when a stellar disc is perturbed by a central bar and a spiral structure simultaneously.

There is observational evidence that non-axisymmetries can cause perturbations in the motion of stars and gas from the non-circular motions of gas flows in the inner MW (Bissantz, Englmaier, & Gerhard, 2003), and the moving groups in the solar neighborhood containing stars of very different ages (Dehnen, 1998; Famaey et al., 2005; Antoja et al., 2008), suggesting that their clumping in velocity space is most likely due to dynamical perturbations from the bar (e.g., Dehnen, 2000) and/or the spirals (e.g, Quillen & Minchev, 2005; Antoja et al., 2009). All of these effects can be explained by resonances associated with a central bar or spiral arms (Quillen & Minchev, 2005; Minchev & Famaey, 2010).

Finally, we now have quantitative evidence that this migration is overall strong in the Galactic disc, as shown by Frankel et al. (2018). They find that stars migrate by about a half-mass radius over the age of the disc. With their model, they find the Sun's birth radius at ~ 5.2 kpc.

1.3.2 Association – field transition

Most likely, all stars formed in groups, clusters, or hierarchies. If this is true, then most clusters must have dissolved into the Galactic background soon after their formation (Krumholz, McKee, & Bland-Hawthorn, 2019). However, our understanding of the

processes behind this remains underdeveloped (Krumholz, 2014; Renaud, 2018; Adamo & Bastian, 2018).

The Galactic disc has plenty of information encoded about its formation and evolution, although unraveling it can become challenging. There are secular processes associated with this evolution, with radial migration playing an important role (Sellwood & Binney, 2002; Roškar et al., 2008; Minchev & Famaey, 2010). These effects wash out the signatures on the birth site of stars, where stars born together will subsequently disperse from their siblings, in orbit and orbital phase, becoming 'field stars'.

In support of radial migration, the Sun has a metallicity larger by $+0.17$ dex than the average metallicity of stars with solar age in its vicinity (Edvardsson et al., 1993). Based on these findings, Wielen, Fuchs, & Dettbarn (1996) suggested that the Sun has migrated from its birthplace by ~ 2 kpc over the course of its lifetime. More recently, the model of Frankel et al. (2018) finds that the Sun has migrated ~ 3 kpc from its birth radius.

Most stars seem to be born in associations (Carpenter, Heyer, & Snell, 2000; Lada & Lada, 2003; Bressert et al., 2010; Kruijssen, 2012), and the Sun is no exception (Pichardo et al., 2012; Pfalzner, 2013). Hence, if the Sun has in fact migrated, then finding its potential parent cluster and locating the Solar family, i.e. a chemically homogenous group of stars that were born with the Sun, can place constraints on the dynamical and chemical evolution of the Galactic disc in the last ~ 4.5 Gyr.

Furthermore, Bland-Hawthorn, Krumholz, & Freeman (2010) show that clusters up to $\sim 10^4$–$10^5 M_\odot$ in mass are expected to be chemically homogeneous. They discuss that this would allow to *tag* stars belonging to these clusters by measuring their chemical abundances (around 10 different elements). Less abundances are needed if we add age or orbital information. Then, it would be possible to trace groups or clusters that were born together where present-day distribution of its stars could provide strong constraints on the rate of radial diffusion or migration in the Galactic disc.

1.3.3   Abundances as birth tags: Chemical tagging

The process of identifying stars of common birth sites only by their abundance signatures is called *chemical tagging* and was proposed by Freeman & Bland-Hawthorn (2002). A necessary condition for chemical tagging is that star clusters be chemically homogeneous and their abundance distributions are sufficiently distinct and essentially uncorrelated from cluster to cluster (De Silva et al., 2007a; Bland-Hawthorn, Krumholz, & Freeman, 2010; Ting, Conroy, & Goodman, 2015).

For instance, open clusters are good laboratories to test for chemical tagging, as it has been shown that they are chemically homogeneous (De Silva et al., 2007a; De Silva et al., 2007b; Liu et al., 2016a; Bovy, 2016), with moving groups also showing such signature (e.g., Chou et al., 2010). The chemical homogeneity in open clusters and for some moving groups is at the level of the measurement uncertainties (~ 0.03 dex; Bovy, 2016; Ness et al., 2018). Some moving groups may be the debris of in-falling dwarf galaxies that were tidally disrupted in the process of being accreted by the MW (Sagittarius stream; de Boer, Belokurov, & Koposov, 2015). Additionally, there are dispersed clusters that cannot be identified spatially but are still identifiable both chemically and kinematically. Most older dispersed aggregates would now not be recognisable dynamically, and chemical techniques provide the only way to identify their debris.

The hierarchical assembly and secular evolution processes diffuse dynamical memory with time (Sellwood & Binney, 2002; Kormendy & Kennicutt, 2004), so we can expect that these will eventually be effective in erasing the kinematic initial conditions of stars in the Galactic disc. Thus, the dynamical information of dissolved clusters will not be sufficient for a complete reconstruction of our Galaxy's history.

This leaves us with chemical tagging, which aims to reconstruct ancient star groups allowing to find dispersed stellar aggregates in the Galactic disc. However, strict or pure chemical tagging has shown to be a challenging technique and it may not be possible in the MW's main stellar disc component, as shown by the significant incidence of *doppelgangers* stars in the field (Ness et al., 2018). For clusters that are completely phase-mixed chemical tagging is most likely the only prospect to trace their origin. However, for dynamically young groups that still remember their birthplaces in star clusters and associations, a *generalised chemical tagging* approach, combining chemical information with orbits, is needed. This could allow us to identify the dissolution of stellar birth associations into the field.

Recently, Krumholz & Ting (2018) found correlations with distance in the metallicity distribution of the Galaxy (at the 20–30% level out to distances distances of ~ 1–2 kpc), and also with time. This would pose a challenge for chemical tagging studies. If abundances are correlated on scales of kpc and times of hundreds of Myr, then the number of unique chemical signatures may be much smaller than had previously been assumed. These findings may seem to work against the case of strong chemical tagging, however it opens up the possibility to find complex structures in chemical space that can be mapped on to systems covering a very wide range of physical space and time scales.

Finally, Ness et al. (2019) have shown that, for a good fraction of the low–α disc, [Fe/H] and age alone can predict the other abundances measured (in APOGEE), therefore revealing that the abundance-space in the disc is low dimensional.

1.4   Datasets for the Milky Way

Observations are a crucial part in learning how galaxies and stars were formed and evolved to their present structure, specially since theoretical models require calibration with well-studied test cases. In particular, our Galaxy is the perfect laboratory to examine individual stars. Stars record the past in their ages, compositions and kinematics. For example, by detecting stellar streams from phase-space positions, accretion and dissolution events can be inferred. Correlations between the chemical compositions and kinematics of field stars could allow us to deduce the history of star formation and even the past dynamics of the Galactic disc.

In general, surveys mapping the Galaxy are divided in two categories: imaging and spectroscopic. Imaging surveys mainly provide photometry and astrometric solutions, with multi-epoch surveys obtaining proper motions and useful parallaxes (if they have sufficient precision; e.g., Perryman et al., 1997; Munn et al., 2004). On the other hand, spectroscopic surveys collect spectra, usually for considerable fewer objects, providing v_{los}, and allowing estimates of stellar parameters such as T_{eff}, $\log g$ or [Fe/H] (e.g., Nordström et al., 2004; Yanny et al., 2009). With the advent of Gaia, we now have at hand a survey that is not only obtaining astrometric and photometric information on over one billion stars, but also spectra collected by a radial-velocity spectrometer (for the brightest objects, $G < 17$ mag), which will revolutionise our understanding of the Galaxy.

In this section we will describe the datasets that we have used in this thesis.

1.4.1   The Gaia mission

Gaia is a satellite performing a census of more than 1 billion stars, by charting a three-dimensional map of our Galaxy. It was launched in December 2013 and placed close to the L2 lagrangian point. This mission is providing accurate positions, parallaxes and proper motions of these sources, and the main goal is to measure the three-dimensional spatial and the three-dimensional velocity distribution of stars in the MW, and determine their astrophysical properties.

The astrometry is complemented by multicolor photometry in the Gaia magnitude $G-$band, measured for all sources observed by Gaia ($G \approx 20$), and radial velocities which are collected for stars brighter than $G = 17$ (~ 150 million stars, Gaia Collaboration et al., 2016a), revealing in the process the composition, formation and evolution of the MW. This amounts to about 1% of the stellar population in the Galaxy.

As of this date, Gaia has produced two data releases, Gaia Data Release (DR) 1 and 2 (GDR1, GDR2). We will provide an overview of both releases, as this thesis has used data from GDR1 and GDR2.

1.4.1.1 Gaia DR1

Gaia's first data release was published on 14 September 2016, after ~ 1000 days of its launch (14 days after I started this PhD), and it contained the astrometry, and $G-$band photometry for > 1 billion stars. The components of this data released contained:

1. A *primary astrometric* dataset with the positions, proper motions and parallaxes for ~ 2.5 million stars in common between GDR1, the Hipparcos (Perryman et al., 1997) and *Tycho-2* (Hoeg et al., 1997) catalogues. This dataset corresponds to the $Tycho - Gaia$ astrometric solution, or TGAS (Michalik, Lindegren, & Hobbs, 2015). The typical uncertainty is ~ 0.3 mas for the positions and about 1 mas/yr for the proper motions (Gaia Collaboration et al., 2016a). A *secondary astrometric* dataset contained the positions for an additional ~ 1.1 billion sources. For this secondary set the typical uncertainty in the positions in much larger ~ 10 mas. More detailed information and a statistical summary of the astrometry in Gaia DR1 is described in Lindegren et al. (2016).

2. A *photometric* dataset with the mean Gaia G-band magnitudes for all of the sources in GDR1. The brightest source in GDR1 has a magnitude $G = 3.2$, with the majority of the sources being in the range $11.2 \leq G \leq 21$. More detailed information on this dataset can be found in van Leeuwen et al. (2017).

3. The *Cepheids and RR Lyrae* dataset with the G-band light curves for ~ 600 Cepheids and ~ 2500 RR Lyrae.

Besides the astrometry and photometry, Gaia also has a spectroscopic instrument, called the radial-velocity spectrometer (RVS) that collects medium resolution spectra over the wavelength range 8450 – 8720 Å, encompasing the Calcium triplet region (Cropper et al., 2018; Katz et al., 2019). The spectra is being collected for all sources with $G_{RVS} = 16.2$ mag and the primary objective is to determine the radial velocity of the sources, although at the bright end ($G < 12.5$; Recio-Blanco et al., 2016) astrophysical information can be derived directly from the spectra. Results from this instrument are **not** contained in Gaia DR1.

For the purposes of this thesis, data from the TGAS dataset was used, as the information in the full-five parameter space (i.e., including parallax and proper motion) was available

with less than 1 year data thanks to the joint solution between the Hipparcos and *Tycho* catalogues.

1.4.1.2   Gaia DR2

Gaia Data Release 2 was available on 25 April 2018. It contained the five-parameter astrometric solution: positions in the sky (α, δ), parallaxes and proper motions for more than 1.3 billion (10^9) sources with a limiting magnitude of $G = 21$ and a bright limit of $G \approx 3$. Here we will provide a general overview of its contents, but detailed information can be found in Lindegren et al. (2018) and Gaia Collaboration et al. (2018b).

1. Parallax uncertainties are ~ 0.04 mas for sources at $G < 15$, ~ 0.1 mas for sources with $G = 17$, and at the faint end (at $G = 20$) the uncertainty is ~ 0.7 mas. The corresponding uncertainties in the respective proper motion components are up to 0.06 mas/yr (for $G < 15$ mag), 0.2 mas/yr (for $G = 17$ mag) and 1.2 mas/yr (for $G = 20$ mag). In this release, GDR2 parallaxes and proper motions are based only on Gaia data; they do no longer depend on the *Tycho-2* Catalogue.

2. A full six-parameter solution: positions and motions on the sky with parallaxes and radial velocities, all combined with mean G magnitudes for more than 7.2 million stars (the RVS sample). This is possible because median radial velocities (i.e. the median value over the epochs) are available for these stars with a mean G magnitude $4 < G < 13$ and an effective temperature in the range 3550 < T_{eff}< 6900 K. The overall precision of the radial velocities at the bright end is $\sim$ 200-300 m/s while at the faint end the overall precision is approximately 1.2 km/s for a T_{eff} of 4750 K and about 2.5 km/s for a T_{eff} of 6500 K. Thus, there are no radial velocities for 'cool' and 'hot' stars. Additionally, no radial velocities have been determined for detected double-lined spectroscopic binaries; such objects are missing from GDR2.

3. G magnitudes for more than 1.69 billion sources, where GDR2 introduced a new photometric reduction. This results in GDR2 having a different photometric system than GDR1. The broad G passband covers the range 3300–10 500 Å (Evans et al., 2018).

4. G_{BP} and G_{RP} magnitudes for more than 1.38 billion sources (80% of sources in Gaia DR2). The G_{BP} and G_{RP} photometry are derived from the integration of the blue and red photometer (BP and RP) low-resolution spectra covering the ranges 3300–6800 Å and 6300–10500 Å.
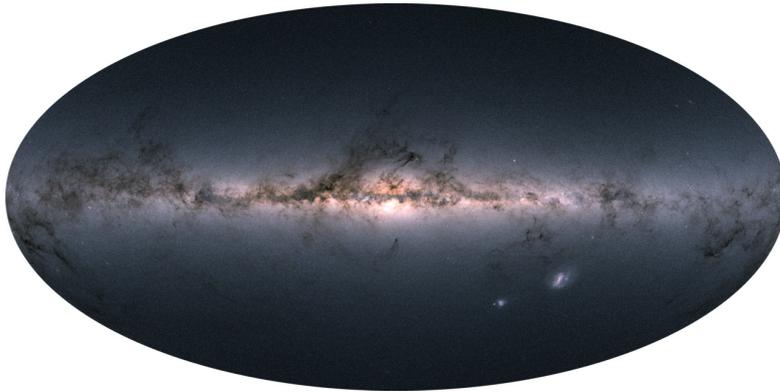
FIGURE 1.3: Gaia's view of the MW. This image shows Gaia's all-sky view of the MW based on measurements of ~ 1.7 billion stars. In the right corner we can also see the Large and Small Magellanic Clouds.

Image Credit:ESA/Gaia/DPAC, CC BY-SA 3.0 IGO

5. *Survey completeness:* the completeness of this release has much improved from the first release, meaning that it is essentially complete between $G = 12$ and $G = 17$. Although a fraction of the bright stars at $G < 7$ is still missing, with no stars brighter than $G = 1.7$ mag appearing in GDR2. The completeness for high proper motion stars has significantly improved with respect to GDR1, but about 20% of stars with proper motion >0.6 arcsec/yr are still missing (Gaia Collaboration et al., 2016a; Evans et al., 2018).

Considerations when using the Gaia DR2 astrometry: the Renormalized Unit Weight Error (RUWE) is a more reliable and informative goodness-of-fit statistic than, for example, the astrometric excess noise. It is expected to be around 1.0 for sources where the single-star model provides a good fit to the astrometric observations. A value significantly greater than 1.0 (e.g., >1.6) could indicate that the source is non-single or otherwise problematic for the astrometric solution.

This thesis has made use of both the RVS sample and the complete sample of GDR2 cross-matched with a spectroscopic survey. In Fig. 1.3 we show Gaia's coverage of the MW based on 1.7 billion stars.

1.4.2 Spectroscopic surveys: an overview

While the Gaia mission will revolutionise our understanding of the structure of the MW and of its formation by providing an unprecedented large volume of high quality positions, parallaxes and proper motions, its spectroscopic measurements are limited to much brighter stars ($G \lesssim 16.2$ mag), and only in the future it will provide some information about their chemical composition (Bailer-Jones et al., 2013). Gaia's spectroscopic limits

mean that precise radial velocities can only be obtained within a distance of $\sim$10 kpc ,even for the brightest stars (de Bruijne, 2012).

Complementing the limited spectroscopic capabilities of Gaia is the motivation of the several ongoing and forthcoming ground-based spectroscopic surveys (high and low resolution) providing radial velocities and some of them chemical abundances for more than 20 chemical species. Some of these surveys are:

1. **RAVE:** The RAdial Velocity Experiment (RAVE, Steinmetz et al., 2006) is a magnitude-limited ($9 < I < 12$) multi-fiber spectroscopic survey of Galactic stars randomly selected in the southern hemisphere. It is a medium resolution spectrograph ($R \sim 7500$) covering the Ca-triplet region (8410–8795 Å), with a typical signal-to-noise-ratio (S/N) and uncertainty for a star in radial velocity of 40 and < 2 km/s, respectively. This survey is currently on its sixth (and last) data release (Steinmetz et al., 2020a), where it has provided wavelength-calibrated and flux-normalized spectra for $\sim$ 500 000 stars. Additionally, they provide spectroscopically derived stellar atmospheric parameters ($T_{\rm eff}$, $\log g$ and the overall metallicity), with abundances of the elements Fe, Al and Ni as well as an overall [α/Fe] ratio (Steinmetz et al., 2020b).

2. **APOGEE:** The Apache Point Observatory Galactic Evolution Experiment (APOGEE, Majewski et al., 2017a) is part of SDSS (Sloan Digital Sky Survey), and is a large-scale infrared (1.51–1.70 μm), high resolution ($R \sim 22\,500$) spectroscopic survey of Galactic stars. It is observing in the H-band, where the extinction is six times smaller than in the V-band, thus making this survey well suited to detect light from stars lying in dusty regions of the MW. APOGEE-1 surveyed $\sim$ 150 000 stars in the Galactic bulge, disc and halo with a typical S/N > 100 delivering stellar parameters including $T_{\rm eff}$, $\log g$, [Fe/H] and [α/Fe]. Additionally, it provides the abundance of 15 chemical species to 0.1 dex precision. It also provides radial velocity measurements with velocity uncertainties of < 100 m/s. This survey is now completed (Majewski et al., 2017b), however it continues through APOGEE-2 collecting data from the duPont telescope at the Las Campanas Observatory. APOGEE-1 predominantly observed red giant stars distributed across several kiloparsecs of the MW disc. APOGEE-2 continues to observe these evolved stars adding the southern hemisphere component.

3. **LAMOST:** The Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) Galactic survey (Deng et al., 2012; Zhao et al., 2012) is an extensive survey, obtaining optical spectra (3700–9000 Å) of a large amount of stars ($\sim 10^7$) with a low resolution spectrograph $R \sim 1800$. The latest data release, LAMOST DR5, has

delivered 9 027 634 optical spectra, of which more than 90% are stellar spectra, with radial velocity measurements. For about 5 million of them, it also provides the basic stellar parameters T_{eff}, $\log g$ and [Fe/H] derived with the LAMOST stellar parameter pipeline (LASP; Wu et al., 2011). Despite the low resolution, it is possible to obtain sensible abundances for $\gtrsim 10$ individual elements from its spectra applying a data-driven model (Xiang et al., 2019). This yields internal abundance precisions of 0.03–0.1 dex for the majority of elements, for stars with S/N ≥ 50 and abundance systematics at the ~ 0.1 dex level.

4. **GALAH:** The GALactic Archaeology with HERMES (GALAH; Zucker et al., 2012; Buder et al., 2018) is a high-resolution ($R \sim 28\,000$) stellar spectroscopic survey. The survey's primary goal is chemical tagging of stars as proposed by Freeman & Bland-Hawthorn (2002). Consequently, the spectrograph has been optimised to measure up to 30 different elements covering a multitude of different nucleosynthesis channels. The second data release of this survey (GALAH DR2; Buder et al., 2018) contains 342 682 stars with stellar parameters (T_{eff}, $\log g$, [Fe/H], [X/Fe]) and abundances for 23 elements. The selection criteria of this survey includes a magnitude cut of $12 < V < 14$ and Galactic latitudes $|b| > 10$ deg, probing mainly FGK stars of the thin and thick disc of the Galaxy.

There are many more spectroscopic surveys, such as the Geneva-Copenhagen Survey (CGS; Nordström et al., 2004) which was one the first homogeneous spectroscopic surveys of the disc encompassing more than 1000 stars. The Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al., 2009), providing spectra for disc stars beyond the solar neighborhood at $R \sim 2000$ with abundances ([Fe/H] and [α/Fe]). The Gaia-ESO (GES, Gilmore et al., 2012) obtaining high-resolution spectra ($\sim 20\,000$) for up to 100 000 stars. Finally, 4MOST (4-metre Multi-Object Spectrograph Telescope) with the ability to perform large spectroscopic surveys of the southern sky, capturing the spectra of 2400 objects simultaneously. Several surveys from 4MOST will complement Gaia, contributing to studies of the stellar halo and bulge of the Galaxy. The spectral resolution ranges from $4000 < R < 21\,000$, with a wavelength coverage of 3700–9500 Å, and it is expected to start its science operations in 2022.

For this thesis we combined the precise astrometry of Gaia with the spectroscopic survey LAMOST, which is the one that provides the largest amount of data and sky coverage, as can be seen from Fig. 1.4.
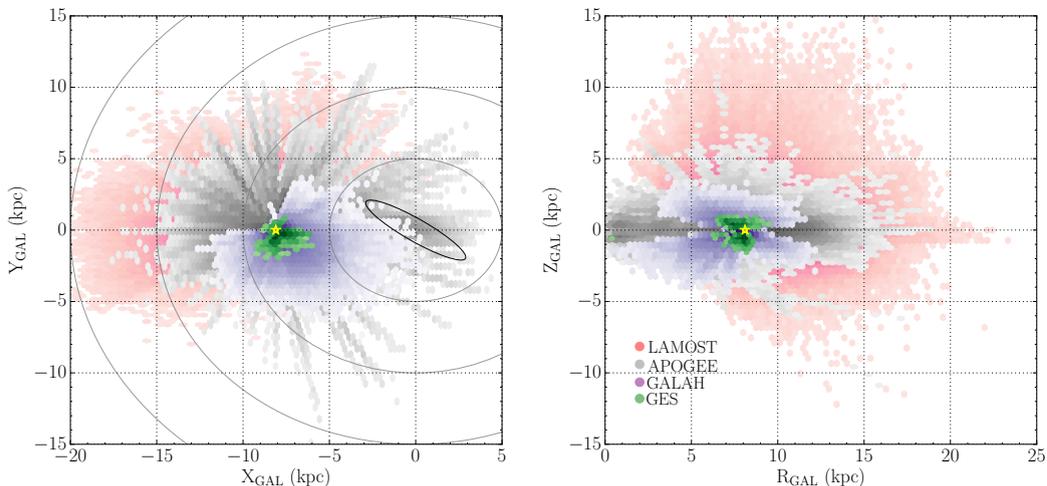
FIGURE 1.4: Survey coverage of the spectroscopic catalogues: LAMOST, GALAH, APOGEE and GES in Galactocentric coordinates. In both panels, the different colors illustrate the different surveys (red: LAMOST DR5, grey: APOGEE DR16, purple: GALAH DR2, and green: Gaia-ESO (GES) DR3 as well as the density of observed stars. To guide the eye, grey circles are placed in multiples of 5 kpc around the Galactic Centre, the location of the Sun is indicated with a yellow star, and the expected location of the Galactic bar (Bland-Hawthorn & Gerhard, 2016) is plotted with a black ellipse. Figure adapted from Queiroz et al. (2019)

1.4.3   How to get element abundances

We have discussed different spectroscopic surveys, where some of them have enough S/N that allow the derivation of element abundances. The latter are very important astrophysical parameters, as a vast amount of fossil information is enclosed in the stellar distribution of chemical elements in the Galaxy (Freeman & Bland-Hawthorn, 2002).

In general terms, metals (elements heavier than hydrogen and helium) are produced in the interior of stars as the product of nuclear fusion reactions and dispersed into the interstellar medium through supernova explosions and winds. This would eventually lead to a trend towards higher metallicities as time passes, with inside-out formation leading to faster chemical evolution or metal enrichment in the inner parts of the Galactic disc (Rix & Bovy, 2013). All supernovae produce iron, however α–elements are produced primarily through supernovae type II. A smooth star formation history will show [α/Fe] decreasing monotonically in time, however mergers could bring gas and produce a star formation burst reinstating a higher [α/Fe], with a dependent relation between [α/Fe] and the star formation history (Gilmore, Wyse, & Kuijken, 1989).

From all of the possible abundances (or [X/H]), the metallicity [Fe/H] is important also for distance estimates. [X/H] determination (normally just [Fe/H], or [M/H] to represent the overall metal content) can be done from photometry or spectroscopy (medium or

high resolution). However, high resolution spectra ($R > 10\,000$) usually enable individual element abundance determinations.

A very nice and detailed review of the accuracy and precision in deriving stellar abundances can be found in Jofré, Heiter, & Soubiran (2019). Here we will provide a brief summary of the methods to derive abundances from spectra.

The most common methods to analyse and determine [X/H] are based on the measurement of equivalent widths (EWs) or the computation of synthetic spectra of absorption lines. EWs are obtained from fitting a gaussian or voigt profile for weak lines and stronger lines, respectively –or by integrating over the line profile–. Synthesis methods consist on varying the abundance of the chemical element in question, until the best fit of a synthetic line profile is found. This latter method might be better suited for crowded spectral regions, or in stars with broad lines.

Recently, machine-learning approaches for measuring abundances have been introduced and applied to stellar surveys (e.g., Ness et al., 2015; Ting et al., 2017; Leung & Bovy, 2019; Xiang et al., 2019). They allow the analysis of large datasets of spectra, where empirical models or neural networks are built, establishing a link between the spectrum and certain labels (abundances) trained on a previously analysed subset of spectra. This is then applied to an entire sample of stars, resulting in precise abundances even for low resolution spectra. Machine-learning methods are very efficient in transferring the known information from the *training sets* to entire datasets. Although, it must be noted that the accuracy of the labels obtained with data-driven methods fully relies on the reference sample.

Finally, Freeman & Bland-Hawthorn (2002) argue that at least ten abundance ratios reflecting different nucleosynthesis channels, at a precision better than 0.05 dex, would be needed to detect structure in chemical space. The most challenging elements are those for which the lines are scarce, too weak, and blended. Some of these belong to heavy neutron-capture elements, which are part of families of other nucleosynthesis channels (different environments and timescales) than the ones typically measured from survey spectra. The most widely measured elements in spectroscopic surveys with high resolution and large wavelength coverage (e.g., GALAH, APOGEE) are C,O, Na, Mg, Al, Si, Ca, Ti, V, Cr, Mn, Fe, Co, and Ni (Jofré, Heiter, & Soubiran, 2019).

All of the efforts from different dedicated spectroscopic surveys show that detailed chemical tagging is therefore possible. However, future methods to trace disrupted clusters will most likely benefit from a combination between elemental abundance information and other methods (e.g., kinematics or dynamical invariants, such as actions).

1.5   How to describe orbits: action-angles

We have presented different surveys in the MW that are currently observing a large amount of stars in the full six dimensional phase-space by combining astrometry and radial velocity measurements. With these datasets at hand, we can start investigating the dynamics of stars.

Disc galaxies, and consequently the MW, are not in a steady state. The Galaxy has strong non-axisymmetric features like a rotating bar or spiral structure in the disc, with infalling objects causing the distribution of matter to evolve, making the gravitational field to be time dependent. Nevertheless, we can make some assumptions and approximations. For instance, we can consider the mass distribution of the disc to be symmetric with respect to the rotation around the axis perpendicular to the disc. Additionally, the DM halo can be approximated as being axisymmetric, at least in the disc plane, and the bar can be considered as an axisymmetric distribution for most orbits outside of the bar region. Finally, since the timescales for secular evolution are quite long, we can work under a general assumption that the Galaxy is in a steady state, and we can investigate the orbits of stars in the Galactic disc by considering an axisymmetric potential $\Phi(R, z)$.

Such framework is governed by the Hamiltonian formalism, which states that an n dimensional dynamical system can be described by the Hamiltonian H, that is expressed in terms of the canonical coordinates $(\boldsymbol{q},\, \boldsymbol{p})$. The solution to the Hamilton equations describe the time evolution of a system. In an axisymmetric potential, H can be expressed in cylindrical coordinates as

$$H = \frac{1}{2}(p_R^2 + p_z^2 + \frac{L_z}{2}) + \Phi(R, z). \tag{1.1}$$

The equations of motion then become

$$
\begin{aligned}
\dot{p}_R &= \ddot{R} = \frac{p_\phi^2}{R^3} - \frac{\partial \Phi}{\partial R}, \\
\dot{p}_\phi &= \frac{\mathrm{d}}{\mathrm{d}t}\left(R^2 \dot{\Phi}\right) = 0, \\
\dot{p}_z &= \ddot{z} = -\frac{\partial \Phi}{\partial z}.
\end{aligned}
\tag{1.2}
$$

Eq 1.2 shows the conservation of the angular momentum component around the potential's symmetry axis z, i.e., $p_\phi = L_z = $ constant. The other two equations describe the coupled oscillations of the star in the R, z directions.

For quasiperiodic orbits there are a set of integrals of motion constant along the orbit (Arnold, 1978). In this context we can consider the action-angle variables (J,θ), which allow us to denote the equations of motion in a simpler form, for which the Hamilton equations then become

$$
\begin{aligned}
\dot{J} &= \frac{\partial H}{\partial \theta} \equiv 0, \\
\dot{\theta} &= \frac{\partial H}{\partial J} \equiv \Omega(J) = \text{const.}
\end{aligned}
\tag{1.3}
$$

The momenta J are constant integrals of motion, and the angles increase linearly with time, $\theta_t = \theta_0 + \Omega \cdot t$ and are 2π periodic. In classical mechanics they are defined as *adiabatic invariants*. This means that the actions remain invariant when a system changes slowly with time. Under this formulation, the description of orbits becomes very easy. Each orbit will be fully determined by a constant J and all of the points in phase-space belonging to the orbit can be mapped on the surface of a torus with coordinates θ, on which an object moves with constant velocity. An additional advantage of these canonical coordinates is that they instantly reduce the complexity of any dynamical dataset by reducing the six phase-space dimensions to three action-angle coordinates. In fact, with this approach, we are now confined to three action integrals J_R, J_z and J_ϕ that quantify the extent of the star's radial and vertical oscillations and the angular momentum, respectively. They are complemented by their respective angles θ_R, θ_z and θ_ϕ. A stellar orbit is then described by the three actions J. Orbits with $J_z = 0$ lie in the Galactic plane, orbits with $J_R = 0$ are circular, and with the appropriate units (dividing J_ϕ by 8 kpc $\times$220 km/s) a circular orbit at the solar circle has $J_\phi = 1$. The orbits of all stars in the Galaxy are described by the distribution function $f(J)$.

Despite the clear advantages of the action-angle coordinate system, in a general potential the calculations may become very difficult. However, they can be easily calculated when the potential is spherical or when using the Stäckel form. In a confocal ellipsoidal coordinate system, this latter potential produces separable Hamilton-Jacobi equations. For the purposes of this thesis, the MW gravitational potential is approximated by an axisymmetric Stäckel potential. A comprehensive and detailed description of action-angles can be found in Binney & Tremaine (2008).

An alternative to action-angle coordinates is a cylindrical coordinate system (R, ϕ, z) that is also useful to cover large regions of the Galaxy.

1.6   Thesis Outline

In this thesis we set out to investigate the stellar orbit distribution of our Galaxy, where we also add the chemical information of stars ([Fe/H]), in a *chemical tagging generalization* approach. Because we need precise distances to study the orbital sub-structure in the Galaxy, in Chapter 2 we develop a probabilistic approach to calculate spectrophotometric distances to $\sim 150\,000$ main sequence stars. We combine the spectroscopic information from LAMOST DR5 and the parallax information from GDR1 (TGAS) where we also take into account the fraction of near-equal binaries. Our method estimates spectroscopic distances with uncertainties of $\sim 6\%$ for single stars. We also present an analysis on how distance uncertainties affect the orbital action calculation, which will become important when investigating possible features in action-angle space.

In Chapter 3 we present an exploration of the orbit sub-structure in GDR2, where we find more signatures in action space than compared velocity space (U, V). Locally, $(1/\varpi < 200$ pc$)$ the J_ϕ (or angular momentum L_z) and J_R actions show prominent features (extended overdensities) in action space, where we also identify well known moving groups. Action space continues to show signatures when we explore larger distances, $1/\varpi > 600$, whereas velocity space at that same distance shows a smooth distribution with no sub-structure present.

In Chapter 4 we apply our model to calculate spectrophotometric distances to $\sim 600\,000$ main sequence stars from GDR2 and LAMOST DR5. We then investigate if the orbit similarity among stars in action-angle space implies indistinguishable metallicities, [Fe/H]. We define the orbit similarity among stars as a distance in action-angle space $\Delta(J, \theta)$ and their abundance similarity by Δ[Fe/H]. First we find an excess of pairs with the same metallicities (Δ[Fe/H] < 0.1) extending to very large separations in $\Delta(J, \theta)$, to nearly 1 kpc distances. By grouping these pairs we find associations with a friends-of-friends algorithm linked by the distance in action–angle space, where we recover known associations (Praesepe, the Pleiades, M67), but we also find groups very extended across the sky, where we identify the recently found Pisces Eridanus stream extending ~ 120 degrees in the sky. This suggests that we are seeing the 'dissolution' of stellar birth associations into the field.

In Chapter 5 we extend the method from Chapter 4 to a larger sample of ~ 6.2 million stars from the GDR2–RVS sample with radial velocities and no cross-match to any spectroscopic survey. Here we make use of the bayesian distances that Schönrich, McMillan, & Eyer (2019) provide for the GDR2–RVS sample. The advantage is that we have a considerable larger sample ($\times 10$ stars), but the drawback is that we have no metallicity information. We recover the largest associations we found in Chapter 4 and we

also recover two more (the Hyades and Coma Berenices) by applying a selection criteria in action space and then investigating their respective angles. We study in more detail some of these groups: The Hyades, the Pisces Eridanus stream and also *unknown* associations that are very dispersed in position and velocity space, but confined in action–angle space.

Finally, in Chapter 6 we present a summary and an outlook with future prospects of the work presented in this thesis.

Astrophysical questions this thesis is posing: As we previously discussed, stars are born in clusters. In the era of Gaia and massive spectroscopic surveys, kinematic and element abundance data are becoming accurate enough that it should be possible to trace stars that are now part of the field back to their birth places, or reconstruct some of the now-dissolved (or in the process of dissolving) structures in which they were born. Several mechanisms (dynamical friction, radial migration or accretion from in falling material) alter the orbits of clusters as they evolve, potentially shaping their demographics. Even after clusters come apart in physical space, their stars remain coherent in action-angle and chemical space for hundreds and thousands of Myr, respectively. Then, it should be possible to reconstruct some of these Milky Way clusters in these spaces.

In this thesis we take the first step towards this goal, and we explore how the orbits of stars are related to their chemical information ([Fe/H]). We find an excess of pairs in action-angle space with indistinguishable metallicities, even to large physical separations. We additionally find hundreds of mono-abundance associations, some clusters and others spread across the sky. These results could potentially constrain how much orbit migration may have have happened, and help us try to answer the fundamental question of how much *dynamical/orbit memory* the Galaxy retains. Additionally, it could provide important clues about cluster formation and subsequent dispersal processes, and the Galactic gravitational potential.

2 | TGAS×LAMOST distances and the role of binarity

Spectrophotometric distances to stars observed by large spectroscopic surveys offer a crucial complement to parallax distances that remain very important also after the future Gaia data releases. Here we present a probabilistic approach to modeling spectroscopic information for a subset of 4000 main sequence stars with good parallaxes ($\sigma_\varpi/\varpi < 0.1$) from the LAMOST × TGAS × 2MASS cross-match, yielding a precise spectroscopic distance estimator with uncertainties of ∼6% for single stars. Unlike previous approaches to this problem, we explicitly account for the individual parallax uncertainties in the model building and fully incorporate the fraction of near-equal binaries of main sequence stars, which would lead to biased distance estimates if neglected. Using this model, we estimate the distance for all (150 000) main sequence stars from LAMOST Data Release 5, without parallax information. As an application, we compute their orbital actions, where our more precise distances result in 5 times smaller action uncertainties. This illustrates how future studies of the Milky Way's orbital structure can benefit from using our model. For the fainter and more distant stars of most current spectroscopic surveys, an approach such as the one presented in this work will deliver better distances than Gaia Data Release 2.

2.1 Introduction

In the age of Gaia we will have access to $\sim 10^9$ stars with some form of parallax and proper motions estimates (Gaia Collaboration et al., 2016b; Gaia Collaboration et al., 2016a). Combined with information from spectroscopic surveys, the full 6D stellar position-velocity phase-space measurements will allow us to study the dynamics and evolution of our Galaxy. For stars with spectra, the distance estimates will often be the dominant source of uncertainty. This is especially true if the distance information comes solely from parallaxes, as those provide only poor distance estimates for the majority of sources in the Gaia catalog (Bailer-Jones, 2015). Fundamentally, stellar distances can be derived either from direct parallax estimates, or from the extinction-corrected flux if the intrinsic luminosities can be inferred from independent astrophysical information. Even after the advent of Gaia Data Release (DR) 2, distance estimates beyond parallaxes will be crucial. Luminosities can be inferred from spectral parameters ($\log g$, T_{eff}, etc.; Queiroz et al., 2018), from objects classification (e.g. RR Lyrae; Sesar et al., 2017), or from asteroseismology (Rodrigues et al., 2014).

In general, useful distance constraints for stars come from both parallaxes and spectra. For example, the Tycho Gaia Astrometric Solution (TGAS) in Gaia DR1 has already provided parallaxes for 2.5 million stars in the solar vicinity ($d \lesssim 200$ pc; Michalik, Lindegren, & Hobbs, 2015; Michalik et al., 2014), but they are not precise enough to be used individually and to probe much larger distances (Queiroz et al., 2018).

Several ground-based dedicated spectroscopic surveys targeting individual stars have become available in the last few years: the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al., 2009), the Apache Point Observatory Galactic Evolution Experiment (APOGEE, Majewski et al. (2017a)), the RAdial Velocity Experiment (RAVE; Steinmetz et al., 2006), the Gaia-ESO (GES; Gilmore et al., 2012), Galactic Archaeology with HERMES (GALAH; Zucker et al., 2012; Buder et al., 2018), the Large sky Area Multi-Object fiber Spectroscopic Telescope (LAMOST; Cui et al., 2012; Zhao et al., 2012), the Experiment for Galactic Undertanding and Exploration (LEGUE; Deng et al., 2012), among others. These have provided valuable data that will allow us to comprehensively study the chemical composition and structure of our Galaxy.

This calls for methods to determine optimal distance estimates that incorporate both parallaxes and spectral information. So far, several works have focused on this goal by making use of the Red Clump as a standard candle (Hawkins et al., 2017; Ruiz-Dern et al., 2018), by making use of all the stars common in TGAS and RAVE (McMillan et al., 2018), or by using several spectroscopic surveys, such as APOGEE, RAVE, GES and GALAH (Queiroz et al., 2018). However, none of these works considered unresolved

binary stars in their modeling for estimating the distances. In this work we follow a similar theoretical approach as the aforementioned authors, but incorporate binarity explicitly, which becomes important for main sequence stars. We illustrate our approach in a study that combines LAMOST data with TGAS.

This is not the first effort to combine these two surveys. Schönrich & Aumer (2017) already worked towards assessing distances in the TGAS × LAMOST cross-match, but only using parallax information. In a different study, Xiang et al. (2017) estimated the absolute magnitude directly from LAMOST spectra, obtaining the distance moduli for 50,000 stars with a TGAS-based magnitude error smaller than 0.2 mag, with a 12 percent error in distance. However, as the authors discuss, these results are obtained with very high signal-to-noise ratio spectra with a median value of 150, which will not be available for the majority of stars in large surveys such as Gaia.

Here, we build a probabilistic model that combines parallax and spectroscopic information, using a subset of stars with precise parallaxes ($\sigma_\varpi/\varpi <0.1$) to build a model for their mean absolute magnitude. This model is then applied to the entire sample of main sequence stars from LAMOST DR5.

One of the direct applications of precise distances is to improve the determination of stellar orbits in the Milky Way, e.g. characterized by their action distribution (J_ϕ, J_R, J_z), see Sec. 2.5.2 for an introduction to orbital actions. In many circumstances it will remain the case that for stars with spectra, the distance uncertainties – based on parallaxes alone – will dominate the uncertainties in calculating the orbits or actions. In Fig. 2.1 we show the distribution of $\sim 150\,000$ main sequence stars from the LAMOST×SDSS/GPS1 cross-match in action space, color-coded by the average metallicity per Voronoi bin. This illustrates the richness of structure in the Galactic disc in terms of stellar orbits and chemical abundances. To characterize the complexity of the stellar disc (e.g. Bovy et al. 2016a; Sanders & Binney 2015) and to explain it in the context of Galaxy formation and evolution (e.g. Minchev et al. 2017; Grand et al. 2018) has been and will be the objective of many studies. Precise estimation of orbits and actions are particularly crucial in action-based dynamical modeling approaches of the Milky Way (e.g. Bovy & Rix 2013; Piffl et al. 2014; Trick, Bovy, & Rix 2016), and for studies investigating orbital properties (e.g. Wojno et al. 2018) or integrating stellar orbits (e.g. Simpson et al. 2019), to just name some very recent efforts. Fig. 2.1 also shows how measurement uncertainties in parallaxes from TGAS translate into widespread uncertainties in action space (see Sec. 2.5.2). In this work we will, drawing on our model for main sequence absolute magnitudes, illustrate the improvement of orbits with better distances.

The structure of the this Chapter is as follows: In Sec. 2.2 we describe the data used, in Sec. 2.3 we present our probabilistic model, while in Sec. 2.4 we show the results of
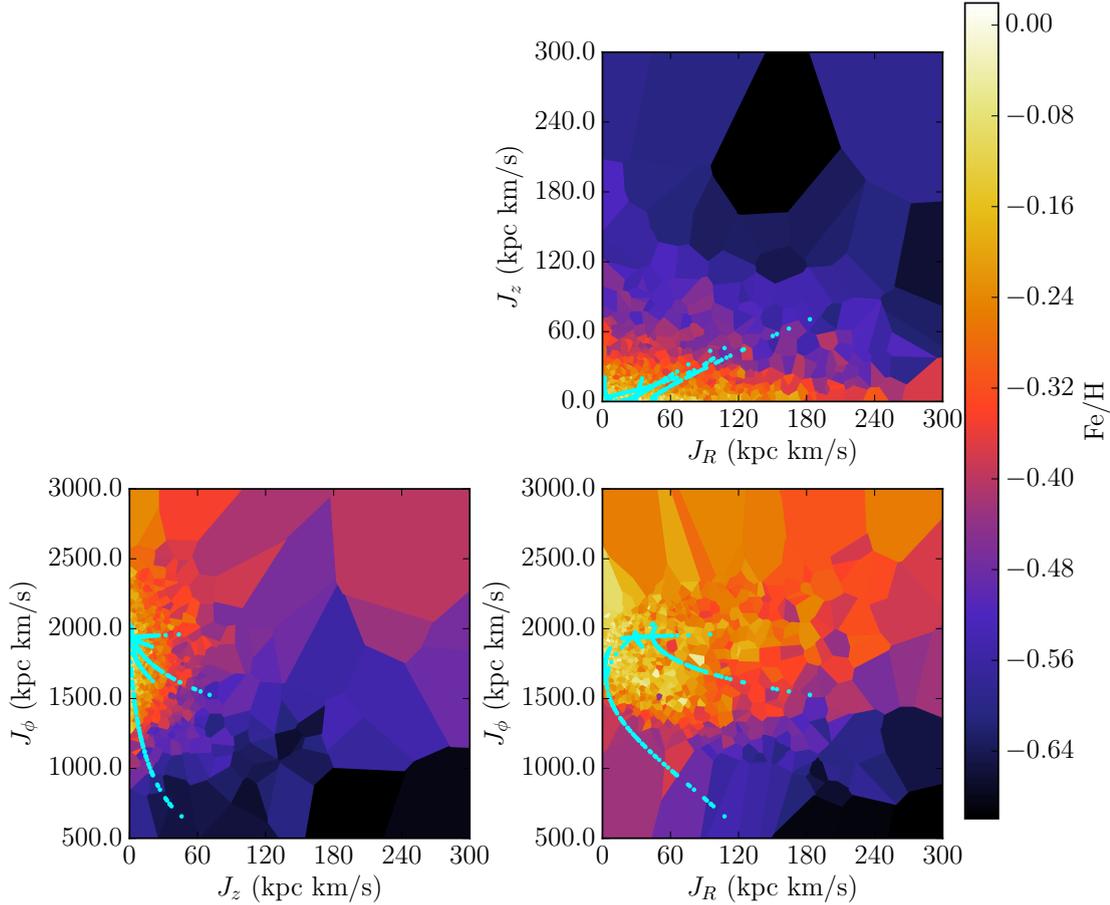
FIGURE 2.1: Action distribution (J_ϕ, J_R, J_z) for $\sim 150\,000$ main sequence stars of the LAMOST sample with proper motions from the GPS1 catalog (Tian et al., 2017) color coded by metallicity. Each cell in this voronoi plot contains 100 stars. Overplotted are the Monte Carlo sampled error ellipses of 5 example stars (in cyan) that result from transforming the measurement uncertainties on the TGAS parallax into action space. This illustrates both the complexity of the stellar disc in action-abundance space, as well as the need for more precise distance estimations. Here, we give just a very short overview of the rich structure in actions and metallicity, of which a detailed description is beyond the scope of this work: Most stars in the Galactic disc are on near circular orbits ($J_R \sim 0, J_z \sim 0$). The overdensity of stars at $J_\phi = R \times v_T \sim 8\,\text{kpc} \times 220\,\text{km/s} = 1760\,\text{kpc km/s}$ is due to the LAMOST survey volume being confined to the solar neighbourhood around $R_\odot \sim 8$ kpc (see also Fig. 2.7). In the vertical action J_z we see the well-known vertical metallicity gradient in the disc (e.g. Ivezić et al. 2008). The low metallicities at $J_\phi < 1300$ kpc km/s are a selection effect of the LAMOST survey, which preferentially selects high-z, low-[Fe/H] stars at smaller radii (see Fig. 2.7). At large J_z the apparent metallicity gradient, rising with increasing J_ϕ, can be traced back to the v_T-vs.-[Fe/H] relation of the thick disc (see e.g. Haywood et al. 2013).
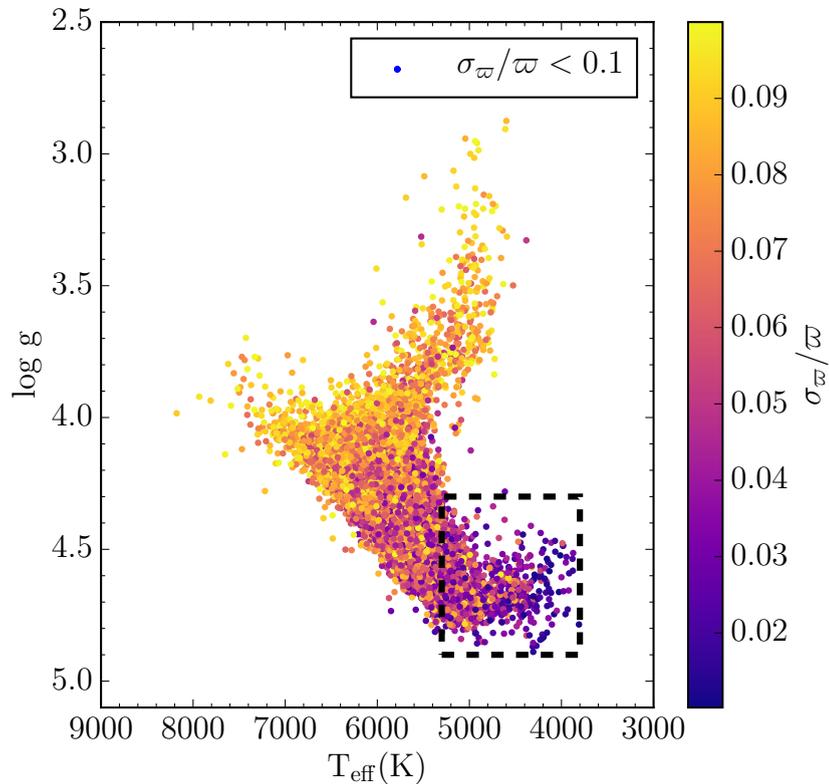
FIGURE 2.2: Distribution in the Kiel diagram of the L⊗T 10% sample. We observe that the coolest stars are most nearby and have the best (fractional) parallaxes. The black dashed rectangle highlights main sequence stars with 3800 K $<T_{\rm eff}$ <5300 K and $\log$ g > 4.2.

the best fit parameters of this model, and show the effects on the computation of orbital actions using the new estimated distances. We finally summarize and discuss the possible implications of our results in Sections 2.5 and 2.6.

2.2   Data description

In this section we describe how we construct the sample that we use to build our model for the mean absolute magnitude, that we then adopt to estimate the distances. First, we obtain the spectroscopic parameters ($T_{\rm eff}, \log g$, [Fe/H]) from LAMOST DR5 (Wu et al., 2011; Wu et al., 2014), and obtain their K–band magnitude from a cross-match with *2MASS*. Then we cross-match this sample with TGAS, which results in $\sim$ 150 000 stars; hereafter, we will refer to this sample as L⊗T. From L⊗T 40 000 stars have "good" parallaxes with $\sigma_{\varpi}/\varpi < 0.1$. We will refer to this subsample as L⊗T 10%, and we show its distribution in the Kiel diagram in Fig. 2.2. The remainder of L⊗T contains stars with
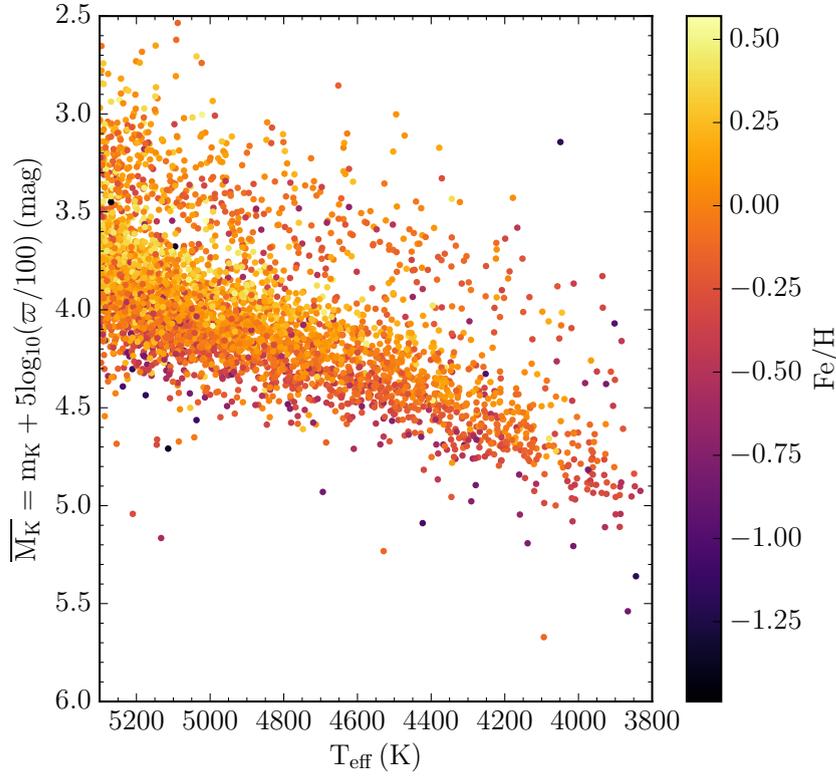
FIGURE 2.3: The mean absolute magnitude $\overline{M_K}$ of MS stars in the L⊗T 10% sample plotted against T_{eff} and color coded by metallicity. The sequence of presumed binary stars appears shifted roughly 0.6 mag towards brighter magnitudes from the primary sequence located at $\overline{M_K} \sim 4$ for $T_{eff} = 5000$ K. The units of ϖ are in milliarcseconds in this figure.

mostly poor parallaxes, hence it is of paramount importance to develop a method to obtain distances for those stars beyond parallax information alone.

From the L⊗T 10% sample, we consider main sequence (MS) stars with $3800\,\text{K} < T_{eff} < 5300\,\text{K}$ and $\log g > 4.2$. This sample is highlighted in Fig. 2.2 with a dashed black rectangle. In Fig. 2.3 we show this subset of ~ 4000 stars where we observe two sequences in the $\overline{M_K}$ vs. T_{eff} plane. The primary sequence of presumed single stars is located at $\overline{M_K} \sim 4$ for $T_{eff} = 5000$ K; while the secondary and less prominent sequence is shifted by about ~ 0.6 mag towards brighter magnitudes, presumably reflecting unresolved binaries of comparable brightness. Based on this subset of MS stars in the L⊗T 10% sample, we proceed to build our model.

2.3   A model for main sequence and binary stars

In this section we will describe how we obtain stellar distances by modeling the mean absolute magnitude of main sequence stars. We start by modeling the K–band absolute magnitude of main sequence stars (of given $\log g$, T_{eff} and [Fe/H]) as a sum of two Gaussians, the first one centered at a mean absolute magnitude $\overline{M_K}$ with dispersion σ_1 and the second one representing the binary sequence and shifted towards brighter magnitudes $\overline{M_K} - 0.6$ with dispersion σ_2.

We follow a Bayesian approach to calculate the probability that our proposed model is associated with the observed data. Then, we can write the probability of the model given the data as:

$$\text{p(model}|\text{data)} = \frac{\text{p(data}|\text{model)p(model)}}{\text{p(data)}}. \tag{2.1}$$

Here, we will explicitly include the individual true distances to the stars as additional free model parameters to be optimized.

2.3.1   Building a model for $\overline{M_K}$

We start by considering that the mean absolute magnitude $\overline{M_K}$ of MS stars does not only depend on the effective temperature, T_{eff}, but also on metallicity, [Fe/H], as illustrated in Fig. 2.3. Hence, we want to determine the mean absolute magnitude using this spectral information. We propose to directly incorporate the spectroscopic parameters (T_{eff}, [Fe/H], $\log g$) and not rely on colors. This will allow us to determine the scatter in the mean absolute magnitude σ directly from the data, as opposed to other works that rely on isochrones, a technique first introduced by Burnett & Binney (2010) and used subsequently by several other works, including Carlin et al. (2015) who implemented it with LAMOST data, but did not provide the obtained distances.

We define the mean absolute magnitude of main sequence stars to be a function of the spectroscopic parameters and expand it up to first order in $\log g$, [Fe/H] and second order in T_{eff} normalizing each parameter by the mean value of the sample: $\overline{T_{\text{eff}}}$ = 4900 K, $\overline{\text{[Fe/H]}}$ = -0.019 and $\overline{\log g}$ = 4.64. This can be expressed as

$$\overline{M_K}(T_{\text{eff}}, \log g, \text{[Fe/H]} \mid \theta_K) = M_0 + \text{a}_{\text{T}} \frac{T_{\text{eff}} - \overline{T_{\text{eff}}}}{\overline{T_{\text{eff}}}}$$

$$+ \text{a}_{\text{T}_2} \left(\frac{T_{\text{eff}} - \overline{T_{\text{eff}}}}{\overline{T_{\text{eff}}}} \right)^2 + \text{a}_{\text{logg}}(\log g - \overline{\log g}) \tag{2.2}$$

$$+ \text{a}_{\text{FeH}} \left(\text{[Fe/H]} - \overline{\text{[Fe/H]}} \right).$$

2.3.2 The probability function including the binary sequence

An unresolved binary system of two identical stars has the same color but twice the luminosity of an equivalent single star. Such systems would form a second sequence or "ridge" in the Color Magnitude Diagram (CMD) (Hurley & Tout, 1998), running almost parallel to the main sequence ~ 0.7 mag brighter. El-Badry et al. (2018) illustrated that the T_{eff}–luminosity tracks of unresolved binary stars with different mass rations q run nearly parallel to the single-star main sequence for $0.8 < q < 1$. As a good fraction of binaries have mass ratios in this range, such a binary "sequence"[1] should be a generic feature. Indeed, Fig. 2.3 shows such a binary sequence among field stars (systems with $0.8 < q < 1$), this is what we observe and what we will model in this work.

We model the distribution of stars in the predicted mean absolute magnitude $\overline{M_K}$ mentioned in Eq. (2.2) as the sum of two Gaussians:

$$p(\overline{M_K}|\theta_M) = (1 - \text{f}_{\text{eqb}}) \cdot \mathcal{N}(\overline{M_K}, \sigma_1) \ + \ \text{f}_{\text{eqb}} \cdot \mathcal{N}(\overline{M_K} - 0.6, \sigma_2). \tag{2.3}$$

We have added the second Gaussian term that accounts for the fraction of near-equal binaries ($0.8 < q < 1$), defined as f_{eqb}. Hereafter, we will use the term binaries to refer to systems in this q-range. We denote the joint posterior probability distribution of the proposed model given the observations as $\text{p}(\theta_M, \{d_i\}|\{\text{D}_i\})$. We define the model parameters as

$$\theta_M = \{M_0, \sigma_1, \sigma_2, \text{f}_{\text{eqb}}, \text{a}_{\text{T}}, \text{a}_{\text{T}_2}, \text{a}_{\text{logg}}, \text{a}_{\text{FeH}}\}, \tag{2.4}$$

and the data as

$$\{\text{D}_i\} = \{\varpi_i, \sigma_{\varpi_i}, m_i, \sigma_{m_i}, T_{\text{eff},i}, \log g_i, [\text{Fe/H}]_i\}. \tag{2.5}$$

We denote the distances to each star as d_i. Using Bayes theorem we can write the posterior probability of our proposed model as:

$$\text{p}(\theta_M, \{d_i\}|\{\text{D}_i\}) \propto \text{p}(\theta_M) \prod_i \text{p}(\text{D}_i|\theta_M, d_i) \, \text{p}(d_i|\theta_M), \tag{2.6}$$

Where $\text{p}(\theta_M)$ are the model priors, $\text{p}(\text{D}_i|\theta_M, d_i)$ is the likelihood function and $\text{p}(d_i|\theta_M)$ is the prior distance for each star. We use the exponentially decreasing volume density prior from Bailer-Jones (2015) for the distances, while we consider flat priors for the rest of the model parameters.

---

[1]This sequence is more precisely a caustic of the binaries' track in the T_{eff}-luminosity plane (El-Badry et al., 2018).

2.3.3   A note on extinction

In principle, we would need to correct the apparent magnitudes for dust extinction before applying our model. To gauge the importance of extinction for our sample we check the reddening values for each star from the 3-dimensional dust map by Green et al. (2015). This map provides the best-fit for $E(B - V)$ in each distance slice, for which we use $1/\varpi$. We note that this is an acceptable approximation given the fact that our sample extends just up to $d \lesssim 200$ pc, and the errors in the parallaxes are small ($\sigma_\varpi/\varpi < 0.1$). We take the extinction coefficient for the K–band (R_K) from (Yuan, Liu, & Xiang, 2013) to convert from reddening to the K band extinction, as $A_K = R_K \times E(B - V)$. However, we find that $A_K < 0.1$ and its mean value is $\approx 6 \times 10^{-2}$, therefore reddening is not important in this sub-sample. We emphasize that this procedure is done to test that the sample we use in building our model is dust free. But in order to obtain reliable distances using a different sample, a correction for extinction must be performed to the apparent magnitude.

2.4   Spectrophotometric distances

2.4.1   Finding the best-fit model and distances to stars with good parallaxes

2.4.1.1   Including measurement uncertainties in the likelihood

We assume that the observed apparent magnitude m_i of each star is the outcome from measuring the brightness of a star with true absolute magnitude $M_{Ki}(T_{\text{eff}}, \log g, [\text{Fe/H}] \mid \theta_M)$ at a true distance d_i with some measurement uncertainty σ_{m_i}. The observed parallax ϖ_i is assumed to be drawn from a Normal distribution described by the true parallax $1/d_i$ and the observational uncertainty σ_{ϖ_i}. The joint likelihood for each star is therefore:

$$\text{p}(\text{D}_i|\theta_M, d_i) = \text{p}(\varpi_i, \sigma_{\varpi_i}|d_i)\text{p}(m_i, \sigma_{m_i}|\theta_M, d_i), \qquad (2.7)$$

where $\text{p}(\varpi_i, \sigma_{\varpi_i}|d_i) = \mathcal{N}(\varpi_i|1/d_i, \sigma_{\varpi_i})$ is the parallax likelihood, defined as a Gaussian evaluated at ϖ_i and centered around a mean $1/d_i$ with dispersion σ_{ϖ_i}. We proceed analogously with the apparent magnitude: $\text{p}(m_i, \sigma_{m_i}|\theta_M, d_i) = \mathcal{N}(m_i|m_{i,\text{true}}, \sigma_{m_i})$, where $m_{i,\text{true}}$ is the predicted apparent magnitude,

$$m_{i,\text{true}} = \overline{M_{Ki}} + A_K + 5\log_{10}(d_i) - 5, \qquad (2.8)$$
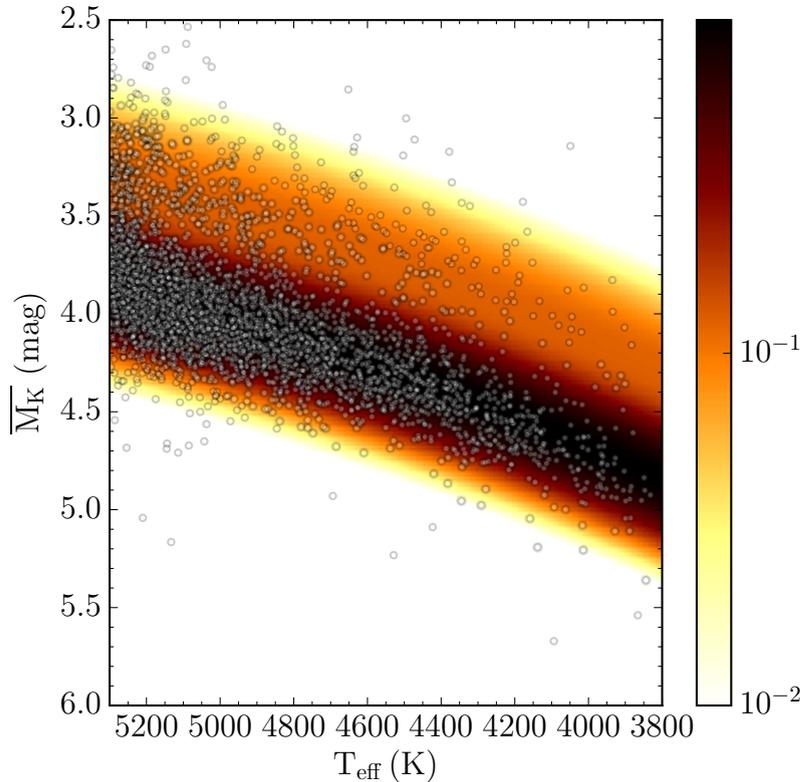
FIGURE 2.4: Mean absolute magnitude model fit to MS stars in the L⊗T 10% sample. Importantly, the model incorporates the binary sequence. The color represents the density of the model *pdf* for the mean absolute magnitude. This is the best-fit model (parameters from Table 2.1), convoluted with a Gaussian of 0.15 mag reflecting the typical parallax uncertainty, for direct comparison with the data.

and $\overline{M_{Ki}}$ has a possible range of values predicted by the distribution in Equation (2.3) given the spectroscopic parameters of the i-th star and the model parameters θ_M. The extinction term A_K is only included if necessary (see discussion in Sections 2.3.3 and 2.4.2.2). The likelihood function in Equation (2.7) is used in the posterior probability distribution in Equation (2.6).

2.4.1.2 Exploring the space of model parameters

We use emcee (Foreman-Mackey et al., 2013), a python implementation of Goodman & Weare's Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler to draw samples from the posterior distribution in Equation (2.6) for our model parameters and distances. The parameter space has $(8 + N)$ dimensions, the 8 parameters θ_M and the N distances d_i when fitting N stars. To reduce the dimensionality of this optimization problem, we start by considering a subset of 100 stars in the MS L⊗T 10% sample from
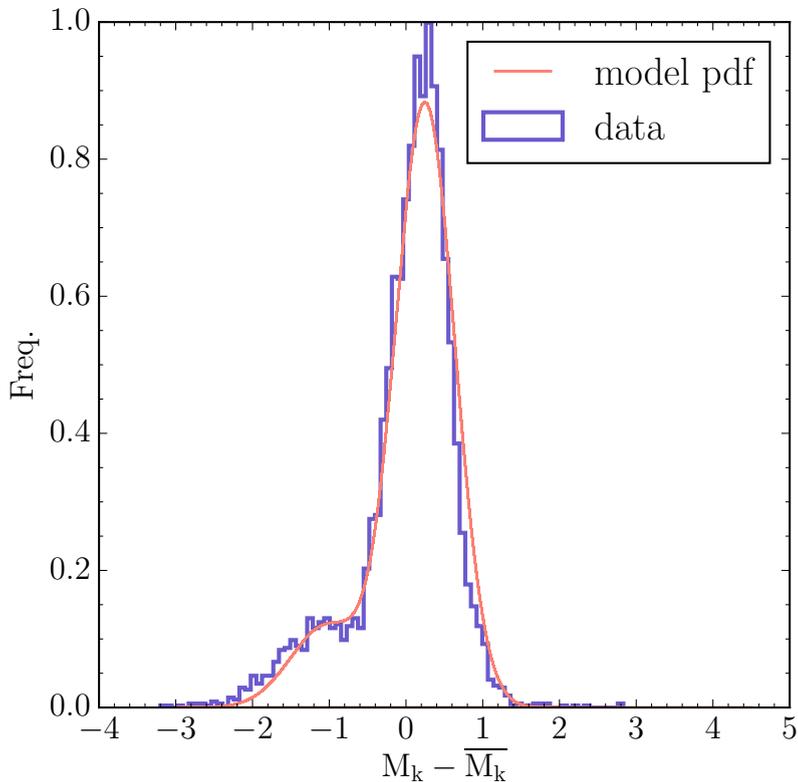
FIGURE 2.5: Normalized number of stars in red and probability for M_K (at a fixed T_{eff}) minus the mean absolute magnitude $\overline{M}_K$ at that same T_{eff} in blue. The data set and model *pdf* correspond to the ones represented in Fig. 2.4.

Fig. 2.3 to establish the parameter values of our model; these stars are randomly drawn. We sample the posterior distribution to find the best parameters that fit our model. In a second step, we use these values as the initial guess for sampling all of the stars within the MS L⊗T 10% sample but now marginalizing over the distances. At the end of this step we then have obtained the best parameters for the entire MS L⊗T 10% sample which we illustrate in Fig. 2.4. The best fit model parameters do not depend on the exact choice of 100 stars from the L⊗T 10% sample.

Fig. 2.4 illustrates the *pdf* density of the modeled $\overline{M}_K$ defined in Equation (2.2). For this purpose we use the best fit parameters presented in Table 2.1. Note again that we did not model the number of stars in the ($\overline{M}_K$, T_{eff}, $\log g$, [Fe/H]) plane, but rather the value of $\overline{M}_K$ given (T_{eff}, $\log g$, [Fe/H]). In order to correctly represent the model *pdf* of the mean absolute magnitude in Fig. 2.4 we also have to incorporate the error in the magnitude. We are considering stars with 10% error in parallax to construct the model which roughly translates into an error of 0.15 mag. Because we are already modeling the mean absolute magnitude as a Gaussian, we incorporate this error in Equation (2.3) by performing a convolution. In Fig. 2.5 we show a comparison between the data set and the model *pdf*
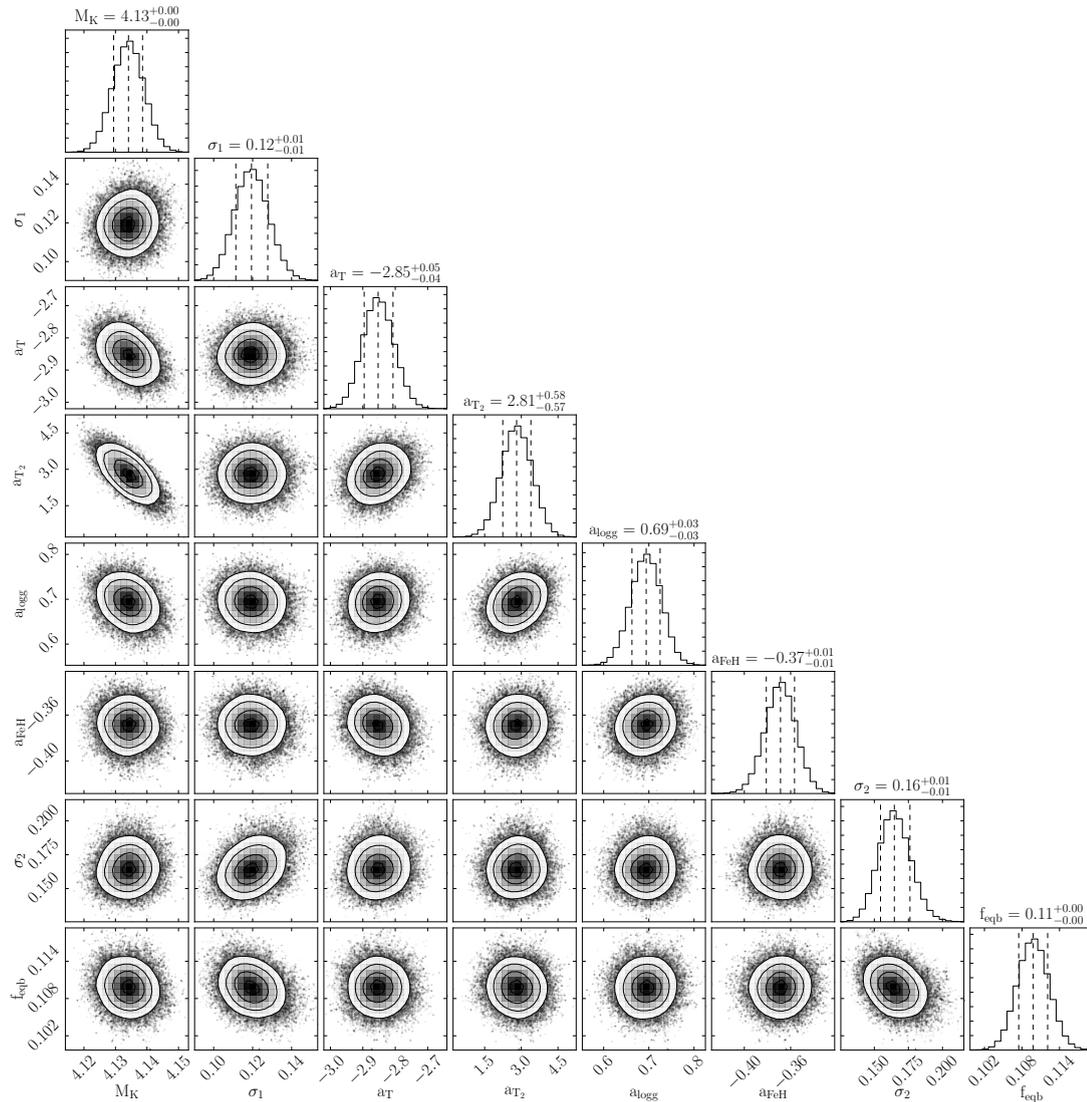
FIGURE 2.6: Corner plot showing the samples from the posterior probability for each parameter of our model for the main sequence considering binary stars. For building the model we use stars in the MS L⊗T 10% sample.

of the mean absolute magnitude also shown in Fig. 2.4. We observe that our model for the mean absolute magnitude captures the essence of the data, including binarity.

In Fig. 2.6 we show the samples from the posterior probability distribution for each of the parameters of our model. All parameters are well constrained with relative uncertainties of between 1% and 6%, and only weak covariances.

Having established the best fit for our model parameters, we can now proceed to obtain the distances for the entire MS L⊗T 10% sample.

One advantage of the type of modeling we have applied is that after finding the best fit for our parameters, we can treat the model as fixed and apply it to many more stars that have very bad parallax or no parallax information at all.

TABLE 2.1: Results obtained with emcee for the parameters of our model. In Fig. 2.6 we show the posterior probability for each of these parameters.

Model parameter		Best fit
Peak abs. magnitude in K-band	M_0 (mag)	$4.134^{+0.004}_{-0.005}$
Width of abs. magnitude distribution	σ_1 (mag)	0.119 ± 0.008
Prefactor of $(T_{\text{eff}} - \overline{T_{\text{eff}}})$ term	a_T	$-2.853^{+0.046}_{-0.043}$
Prefactor of $(T_{\text{eff}} - \overline{T_{\text{eff}}})^2$ term	a_{T_2}	$2.809^{+0.583}_{-0.567}$
Prefactor of $(\log g - \overline{\log g})$ term	$a_{\log g}$	$0.694^{+0.031}_{-0.032}$
Prefactor of $([\text{Fe/H}] - \overline{[\text{Fe/H}]})$ term	a_{FeH}	-0.369 ± 0.012
Width of binary sequence	σ_2 (mag)	$0.165^{+0.011}_{-0.010}$
Binary fraction (equal mass binaries)	f_{eqb}	0.110 ± 0.002

2.4.2   Distances to stars with no useful parallaxes

2.4.2.1   Applying the best-fit model for $\overline{M_K}$

With this model at hand, we can determine distances to entire LAMOST DR5 MS sample, most of which has currently no parallax information. We combine it with the GAIA $\times$ PS1 $\times$ SDSS (GPS1) catalog (Tian et al., 2017). From this cross-match we obtain $\sim 150\,000$ stars (hereafter, L$\otimes$G) with proper motions from GPS1 and spectroscopic information $\log g$, T_{eff}, [Fe/H] and line-of-sight velocities from LAMOST also apparent magnitudes in the K band from *2MASS*, and in the G–band from Gaia.

With the final parameters shown in Fig. 2.6 we proceed to apply our model to this sample, while the L$\otimes$T 10% data and model in Sec. 2.3 and Sec. 2.4.1 could be treated as dust-free, this is no longer true for the whole LAMOST sample. So we first correct the apparent K–band magnitude for extinction using the method described in detail below in Sec. 2.4.2.2. After this step we proceed to calculate their corresponding spectrophotometric distances using the parameters obtained from the best-fit model illustrated in Fig. 2.6 and Table 2.1.

The posterior distribution presented in Equation (2.6) provides a complete description of the distance. However, we want to obtain a single value of the distance along with its uncertainty. We do this by taking the median value of the distribution as the single value for the distance to each star, and for the uncertainties we consider the 16th and 84th percentile.

With the new calculated distances we illustrate the distribution of these stars in the galactic X–Y–Z plane in Fig. 2.7, where we assume that $R_\odot = 8$ kpc and $z_\odot = 0.025$ kpc. From these distributions we see that we have a sample more or less confined to the solar neighborhood.
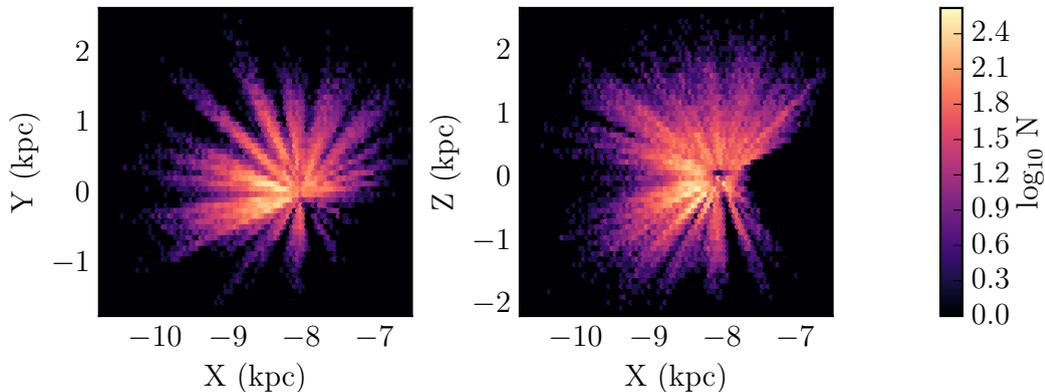
FIGURE 2.7: Color coded density distribution in logarithmic bins of the number of stars N in the Galactic X-Y-Z plane for the stars in the L⊗G sample, corrected by extinction and with distances calculated from our model.

2.4.2.2 Correcting for Extinction

For the case where we have no parallax information, we cannot retrieve the reddening values from the dust map by Green et al. (2015). Instead, we use the color G–K as an estimator for reddening, following the relationship between infrared and optical extinction proposed by Cardelli, Clayton, & Mathis (1989). We can write the extinction coefficient, A_K as a function of color and T_{eff} as $A_K = f(G\text{–}K\,|\,T_{\text{eff}})$. Because the value for the T_{eff} comes from spectroscopy, it is independent of reddening. In Fig. 2.8 we plot G–K vs. T_{eff} for the MS L⊗T 10% sample which has good parallax information and the L⊗G sample with no parallax information. We observe that the MS L⊗T 10% sample is tightly constrained in G–K as a function of T_{eff} and therefore it is not strongly affected by dust as we already noted in Sec. 2.3. The dashed line in Fig. 2.8 represents this empirical relation, and we use it to obtain the extinction coefficients in the K–band.

The L⊗G sample has a large spread in G–K, indicating that it is affected by extinction. We quantify the excess in color with respect to the dashed line in Fig. 2.8 as the amount of extinction.

We proceed to write an empirical relation to obtain A_K, where we follow Cardelli, Clayton, & Mathis (1989) and use their Eq.1 that expresses the mean extinction law as $< A(\lambda)/A(V) >= a(x) + b(x)/R_V$ and Table 3 for the value of a(x). Here, we have ignored the slight color dependence in the transformation between the G and V magnitudes (Jordi et al., 2010) and have simply treated the G as V band. We consider that all the points that lay above the dashed line in Fig. 2.8 are affected by dust, and therefore must be corrected for extinction. The points below are not corrected but still remain in our sample.
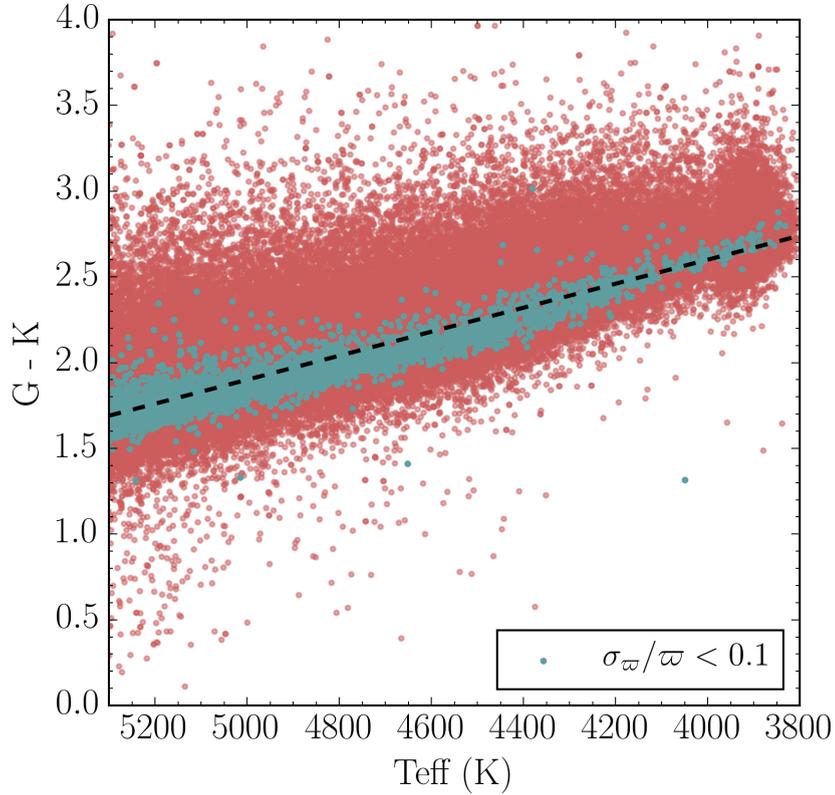
FIGURE 2.8: Color G–K vs T_{eff} for MS stars including those without parallax information from the L⊗G sample in pink dots. Blue dots show the MS L⊗T 10% sample, they are nearby and slightly reddened. The dashed line represents the mean location of the blue dots in this diagram and we use it to write an empirical relation to represent the extinction coefficient in the K–band.

We calculate the extinction coefficient A_K for each star, and then we can correct the apparent magnitude and therefore proceed to calculate the distances for each star using our model.

2.5 Discussion

In the previous sections we have established a model that in the first step relies on a subset of stars with precise parallax measurements. These stars are used to find the best-fit parameters for a mean absolute magnitude model for the MS that depends on spectroscopic information. In a second step, with the established and now fixed model we can obtain spectrophotometric distances. This allows us to obtain improved distances even for the stars that have poor or no parallax information. In the following we further illustrate this.
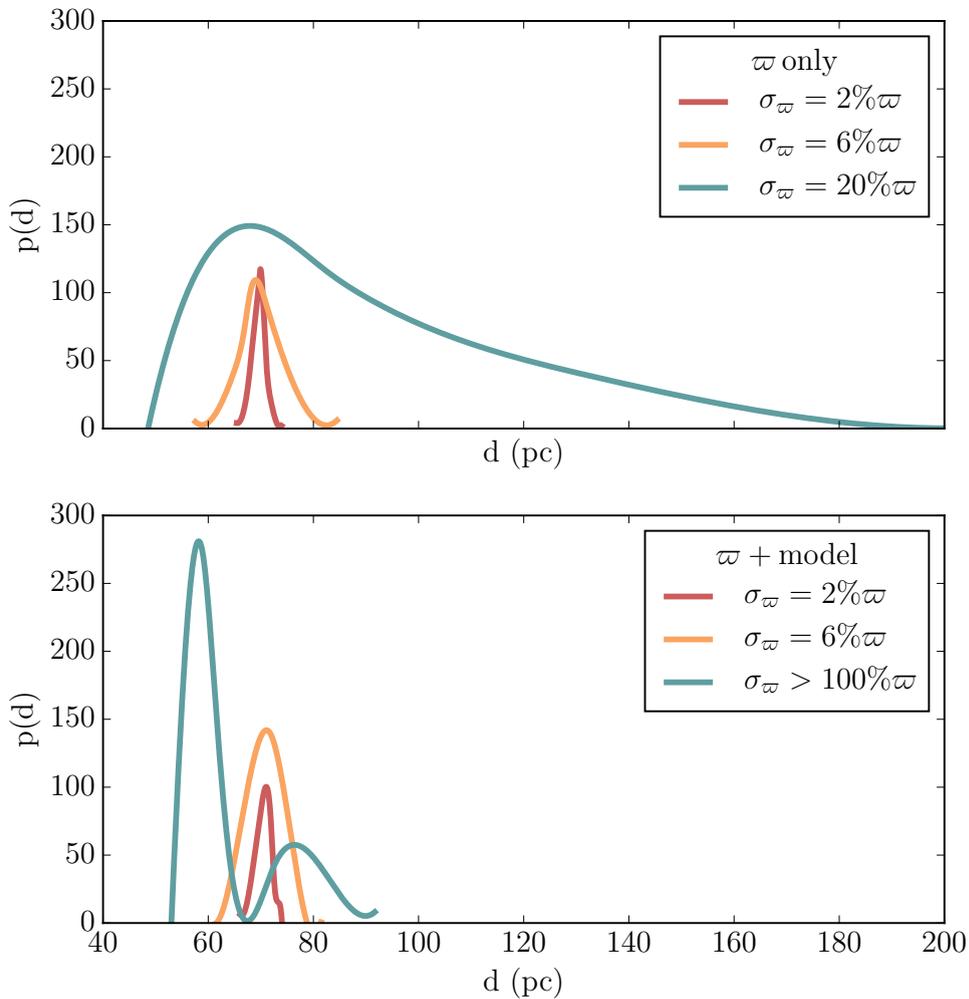
FIGURE 2.9: In both panels of this figure we show the result of the posterior distribution of the distance for the same star in different regimes of parallax error. The median value for distance of this star is $d =70$ pc. The upper panel shows the results considering only ϖ and the lower panel the results when applying our model. We show the results in the regime of a very good parallax with $\sigma_\varpi < 5\%$ in light red and $\sigma_\varpi < 10\%$ in yellow. In the bad parallax regime the upper panel shows the result for $\sigma_\varpi = 20\%$ in light blue and the lower panel for $\sigma_\varpi > 100\%$.

2.5.1 The effect of parallax uncertainty on the recovered distances

In this section we will explore how a star's best fit distance changes when we consider different regimes of parallax error. In this experiment, we consider the same star, but vary its σ_ϖ. In Fig. 2.9 we show the posterior distribution of the distances that we obtain if we apply our model with the best fit parameters from Table 2.1, for very precise parallaxes and for a very extreme case of poor parallax: $\sigma_\varpi = 2\%, 6\%, > 100\%\varpi$. We observe that in the limit of a very good parallax for a star, we obtain a single narrow Gaussian. However,

with bad parallax information, we get a bimodal distance distribution, as the model relies entirely on the (also bimodal) absolute magnitude probabilities to obtain the distances.

Using the same star, we now explore the posterior distribution of the distance that we obtain if we rely only on parallax information. This can be seen in the top panel of Fig. 2.9. Here we also show results for the very good parallax regime, with $\sigma_\varpi = 2\%, 6\%\varpi$. As an example of the regime of "poor parallaxes", we show the results for $\sigma_\varpi = 20\%\varpi$. Even larger parallax uncertainties yield divergent uncertainties on the distance estimates (Bailer-Jones, 2015).

2.5.2 Estimating orbital actions from data with observational uncertainties

In this subsection we will describe a direct application of our improved distances using the L⊗T sample that contains stars with poor parallax estimates. From the distribution of stellar orbits, we can learn about both the dynamics and formation of the Galaxy. The movement of a star on an orbit can be easily described by the canonical action-angle coordinates (J, θ). If we consider an axisymmetric gravitational potential, then these orbits can be fully determined by three integrals of motion $\mathbf{J} = (J_R, J_\phi, J_z)$, and they are defined as:

$$J_i = \frac{1}{2\pi} \oint p_i dx_i, \tag{2.9}$$

where the integral is evaluated along the orbit with position $x(t)$ and momentum $p(t)$. The actions $\mathbf{J}$ label orbits and each angle variable θ increases linearly with time and indicate the position of the star along the orbit. J_R quantifies the oscillations inwards and outwards in the radial direction, J_z quantifies the oscillations in bthe vertical direction and J_ϕ is the component of angular momentum. We redirect the reader to Binney & Tremaine (2008) Sec. 3.5 for a detailed description of actions.

To compute the actions we need the full 6D phase information i.e., velocities ($\mu_{R.A}$, μ_{dec}, v_{los}) and positions (ra, dec, distance). For this calculation, we make use of the *galpy* package, which is a python implementation for galactic-dynamics calculations (Bovy, 2015). We consider the simple axisymmetric Milky Way potential with a Miyamoto-Nagai disc, NFW halo and power law bulge that it is implemented in *galpy* as `MWPotential2014` (Bovy, 2015). We transform the position and velocities of each star to Galactocentric coordinates, where we consider the position and velocity of the sun for the coordinate transformation to be at (X,Y,Z)= (8, 0, 0.025) kpc and (U, V, W)= (-11 ,230, 7) km/s, respectively. From the 6D phase information, the distance is the one with the largest impact on the action distribution uncertainties. We translate the uncertainties from the observations to action space via Monte Carlo sampling of an error ellipse. We convert
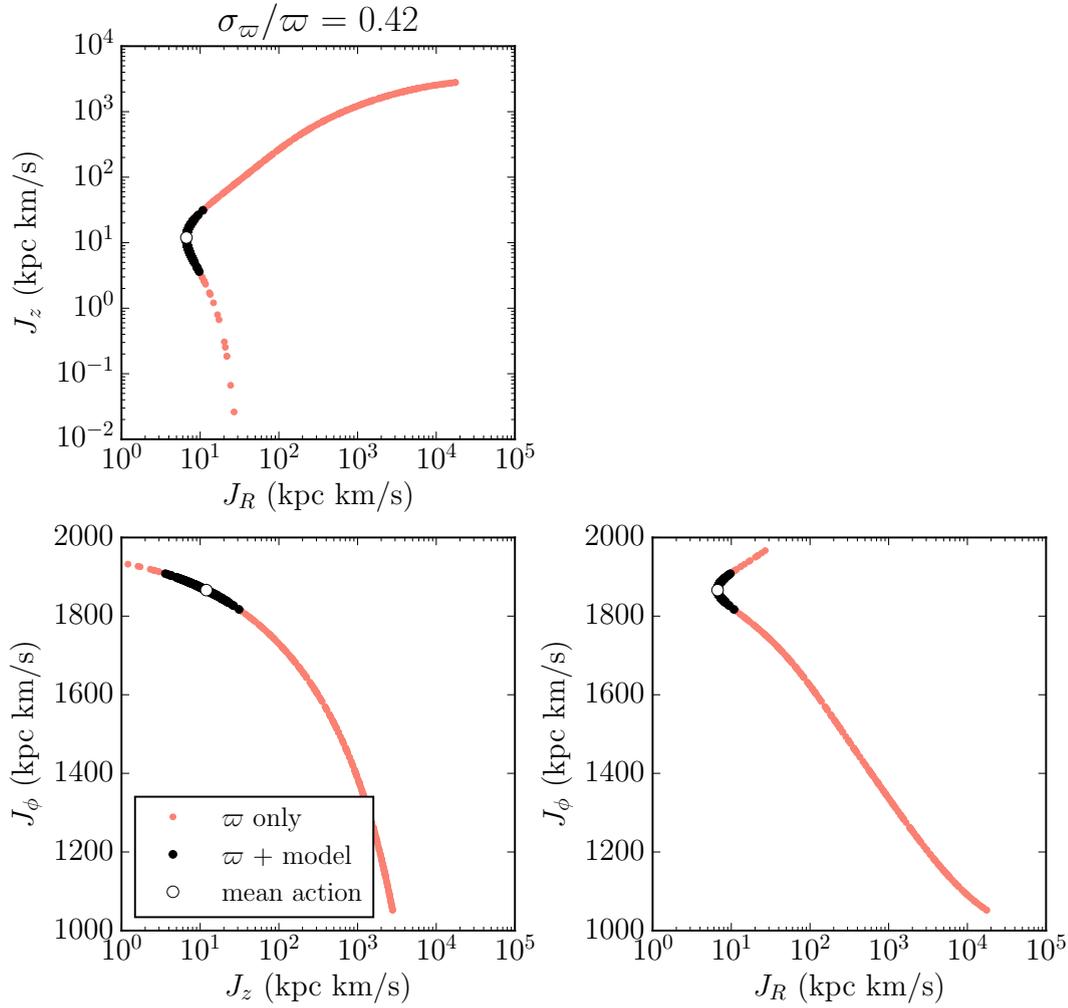
FIGURE 2.10: 1000 samples of measurement uncertainty ellipse transformed to action space (J_R, J_ϕ, J_z) performed via Monte Carlo sampling for a star in the bad parallax regime ($\sigma_\varpi = 42\%\varpi$) in the L⊗T sample. These show the extent of the uncertainties in action space when the parallax is very imprecise. We compare the results of our model in black dots and considering only ϖ in red dots. The white dot shows the action's mean measurement.

each sample of the error ellipse from the observable space to Galactocentric cylindrical coordinates and then to actions. We then run 1000 samples of the error ellipse to explore the extent of distance uncertainties.

2.5.2.1 The effect of improved distances

We compare the distance uncertainties obtained using only parallax information, and our model considering both parallax and spectroscopy. We have shown already in Sec. 2.5.1
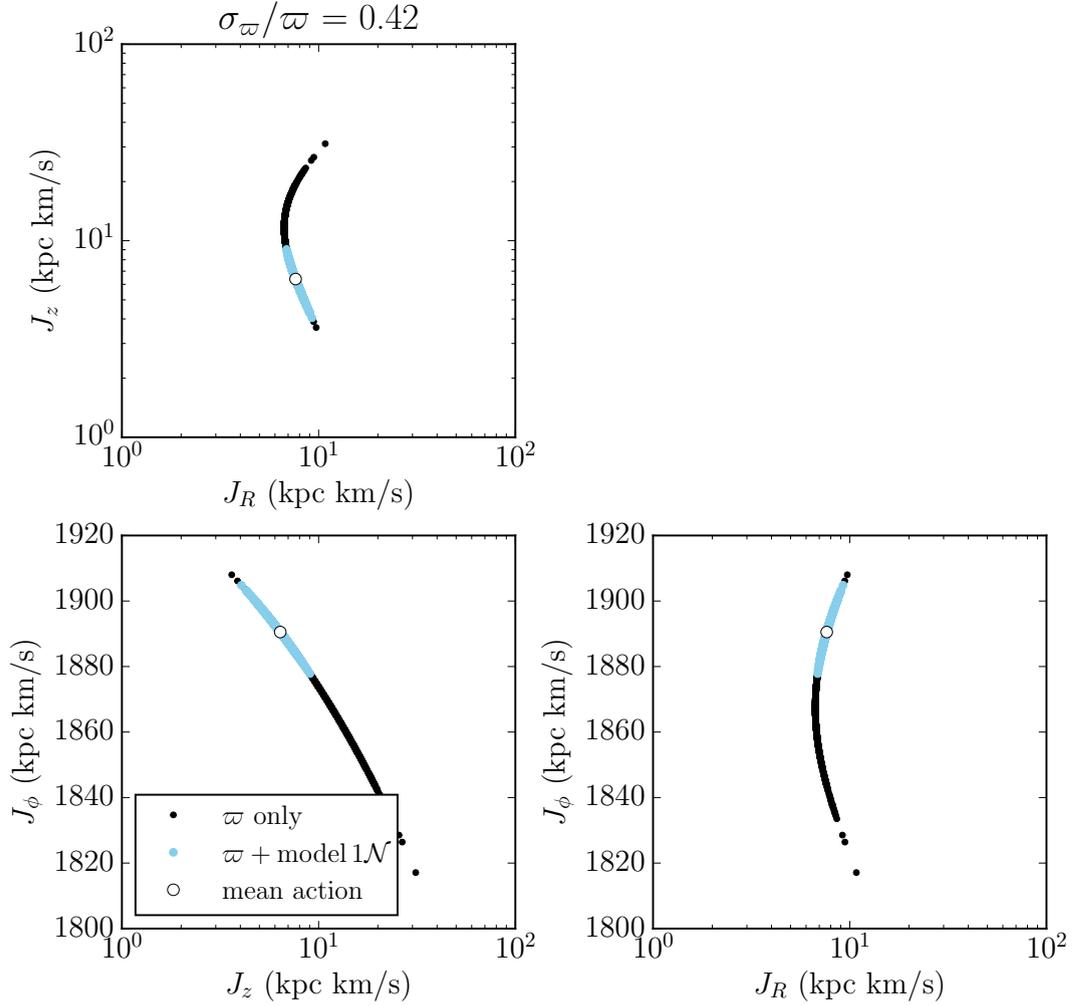
FIGURE 2.11: 1000 samples of measurement uncertainty ellipse transformed to action space (J_R, J_ϕ, J_z) performed via Monte Carlo sampling for the same star from Fig. 2.10. Here we compare the samples resulting from our model that considers binary stars in black dots again and a model that considers one gaussian i.e, only single stars in blue dots. Again, the white dot shows the action's mean measurement.

by exploring the distance's posterior distribution, that incorporating spectroscopic information becomes especially important in the regime with very bad parallaxes. When exploring the action space this effect is also visible. Relying just on parallaxes causes the measurement uncertainty distribution to spread out over a large portion of action space as can be seen in Fig. 2.10.

2.5.2.2   The effect of binarity

We can also explore the effect of binarity on the estimate of a stars orbits, seen in action space. We do this by modeling the absolute magnitude with only one Gaussian, not
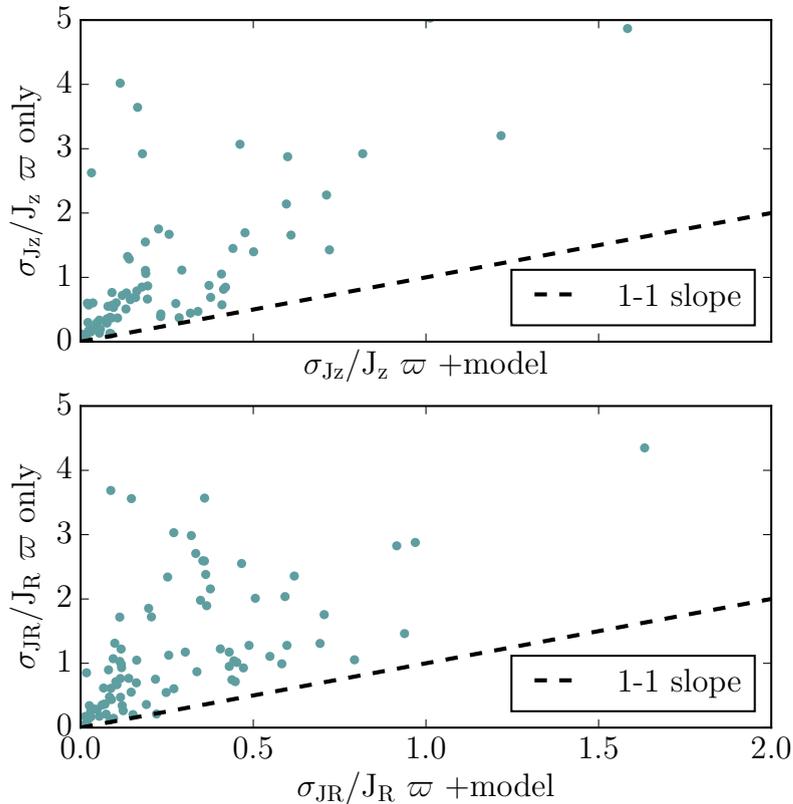
FIGURE 2.12: Comparison of the action estimate precision for ϖ only and the model + ϖ. We show how our model reduces the uncertainties also in action space by plotting σ_J/J for stars with $\sigma_\varpi/\varpi > 0.15$. We plot σ_{J_z}/J_z in the upper panel and σ_{J_R}/J_R in the lower panel. We observe that σ_J/J is always smaller when we use our model.

considering the contribution of binary stars. This can be seen in Fig. 2.11. Here we show the same star as Fig. 2.10. We note that the model with only just one Gaussian as expected, shows smaller uncertainties in action space. Nevertheless, the effect of considering binary stars does not translate into an important effect in action space.

We have shown that having large uncertainties in the distances translates also into large uncertainties in action space, especially when we consider only parallax information, as observed in Fig. 2.10. This effect becomes even more important when we consider stars with large uncertainties in parallax. But to quantify how much our model actually improves the action space in Fig. 2.12 we plot σ_J/J for the vertical and radial action. To obtain σ_J we average over the 1000 action samples per star. In this figure we plot stars with large uncertainties in parallax, $\sigma_\varpi/\varpi > 0.15$. We observe that in all of the cases, σ_J/J is smaller when we use our model to calculate the distances. Therefore, having precise distances clearly has an impact in the calculation of actions.

2.5.3   Spectophotometric distances and DR2

Now we want to explore in which regime the spectrophotometric distances will be more precise than the parallaxes in the second data release (DR2) of Gaia. We consider the sample L⊗G which contains long lived MS stars in the regime 3900 K $< T_{\rm eff} < 5300$ K, where the MS lifetime is comparable to the age of the disc, i.e, stellar masses lower than the turn-off mass. For this purpose, we define the fraction $\delta{\rm DM}_{\rm DR2}/\sigma_1$, which compares the uncertainty in distance modulus according to the Gaia DR2 parallax to the uncertainty in our model. In particular σ_1 is the result of our model with the best-fit parameters given in Table 2.1 and Fig. 2.6 and indicates the precision we achieve in distance modulus. The uncertainty in distance modulus from error propagation of the Gaia DR2 parallax is

$$\delta{\rm DM}_{\rm DR2} = \frac{5}{\ln 10} \frac{\delta\varpi(G)_{\rm DR2}}{\varpi({\rm DM}_{\rm phot})}, \tag{2.10}$$

where $\delta\varpi(G)_{\rm DR2}$ is the expected parallax uncertainty in DR2. We estimate this by using the projected end of mission uncertainty as a function of G magnitude (de Bruijne, 2012, their Fig. 10), as $\sqrt{3}\delta\varpi$. This takes into account the fact that we will have roughly 1/3 of the data after DR2. $\varpi({\rm DM}_{\rm phot})$ is the parallax that corresponds to the most likely photometric distance modulus to that star using our best fit model parameters given in Table 2.1.

We illustrate the results of this comparison in Fig 2.13, where stars that lie above the dashed line with $\delta{\rm DM}_{\rm DR2}/\sigma_1 > 1$ are stars for which our model would perform better than Gaia DR2. We also observe that these results correlate with the effective temperature, showing that for warmer MS stars in our sample the fraction of stars that still need spectrophotometric distances is quite large. For intrinsically low-luminous (cool) stars the advantage of Gaia will be greatest. Faint stars in this survey will tend to be nearby which translates into larger parallaxes with low uncertainties. On the other hand, luminous stars will be observed up to larger distances, therefore at smaller parallaxes and with large uncertainties.

2.6   Final remarks

In this Chapter we have presented a method to calculate spectrophotometric distances for main sequence stars in the Milky Way, where our model explicitly accounts for the parallax uncertainties, and for the common binarity of near equal mass binaries among main sequence stars. To build our model we make use of the parallax information from Gaia, and spectroscopic information from LMDR5.
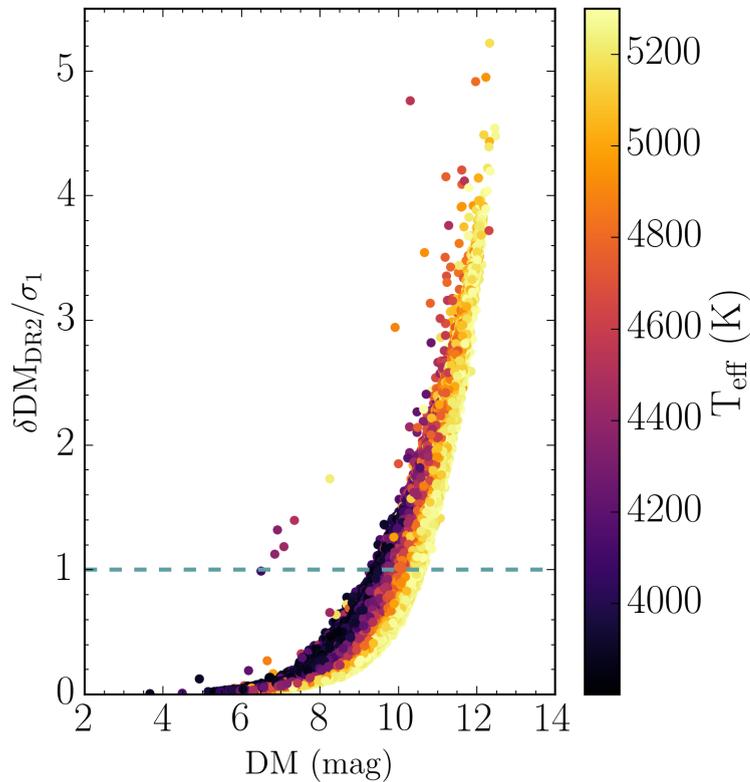
FIGURE 2.13: Distance modulus for stars in the regime 3800 K $< T_{\text{eff}} <$ 5300 K in the L⊗G sample compared to $\delta DM_{\text{DR2}}/\sigma$, color coded by effective temperature. $\delta DM_{\text{DR2}}/\sigma$ corresponds to the expected parallax uncertainty for DR2 divided by the parallax that corresponds to the most likely photometric distance to the star. Stars located above the dashed line are the stars for which our model performs better than DR2. We see a clear correlation with temperature.

We have built a model for the mean absolute magnitude of main sequence stars, which mostly draws on parallax information whenever parallaxes are useful ($\delta\varpi < 10\%$); for increasingly poorer parallax estimates, this model gradually draws on the spectrophotometric information to estimate the distance modulus. We obtain a value for the intrinsic dispersion in the absolute magnitude of single stars $\sigma = 0.12$ mag, which gives precisions in distance of $\sim 6\%$ for the fainter and more distant MS stars among current spectroscopic surveys.

As an application of precise distances we showed that they greatly improve the precision of orbital action estimates, as distance uncertainties dominate the orbit uncertainties.

3 | Orbit sub-structure in Gaia DR2

This chapter provides a summary of the article by Trick, Coronado, & Rix (2019), in which I am a co-author. My specific contributions, relevant to this thesis, are presented in Section 3.3.

In the previous Chapter we have established the importance of precise distances to calculate actions. Actions are a useful canonical coordinate system to describe stellar orbits, and provide valuable information about the the orbital distribution and sub-structure of our Galaxy. With the advent of Gaia DR2 we now have the opportunity to explore a larger sample with 7.2 million stars having not only parallaxes, proper motions and photometry information, but also radial velocities (Katz et al., 2019). This implies that we can now obtain 'good orbits' to ~ 1 kpc, which is roughly 1000 times the volume of Hipparcos.

Because we require precise distances to calculate reliable actions, we restrict this large dataset to stars within 1.5 kpc, which leave us with ~ 3.8 million stars. By doing this, the majority of stars in this sample have a distance error of $\sim 5\%$, with $\delta\varpi/\varpi < 0.05$. We therefore safely use $1/\varpi$ as a distance estimate. Nevertheless, we explore how the parallax uncertainties translate into action space in Section 3.3.

3.1 What signatures do we expect to observe in action space?

Before investigating the action distribution in the dataset we first explored what we expect to observe from a smooth distribution, perfectly phase-mixed. When plotting

the angular momentum J_ϕ (or L_z) vs. J_R, stars at $J_R \sim 0$ are moving on near circular orbits and stars with large J_R are moving on highly eccentric orbits. In such a figure, we also observe an envelope with a parabolic shape (see Fig. 3.1); showing the extent of the sample's selection function, and also the signatures of the survey volume. Stars that are on average further away from the Sun, cannot enter the survey volume if they have low J_R. Also, stars with $J_\phi/(8\,\mathrm{kpc} \times 220\,\mathrm{km/s}) < 1$ are close to apocenter passage, and stars with $J_\phi/(8\,\mathrm{kpc} \times 220\,\mathrm{km/s}) > 1$ are close to pericenter passage. For a smooth and phase-mixed distribution there should not be sub-structure present, instead we should only expect to see a smooth distribution in $J_\phi - J_R$.

3.2 The orbit distribution in GDR2 is highly structured

In contrast with a phase-mixed distribution of stars, we observe rich sub-structure in the actual data, even when considering stars up to 1.5 kpc from the Sun. To illustrate this, we separate our stars in three different distance regimes, which is presented in Fig. 3.1. The upper panel of this figure shows the action distribution in J_ϕ–J_R for stars in the solar neighborhood ($1/\varpi < 200$ pc), the middle panel for stars within $200 < 1/\varpi < 600$ pc and the lower panel the distribution of stars within 600 pc $< 1/\varpi < 1.5$ kpc. Here we also show to the left of each panel the distribution of stars in Galactocentric coordinates (R, z) and the $(U, V), (-v_R, v_T)$ velocities.

For stars within $1/\varpi < 200$ pc we recognise the well-known features of the moving groups in velocity space already shown in Gaia Collaboration et al. (2018a). Some of these groups are: Sirius, Coma Berenices, the Hyades, Pleiades and Hercules and their approximate location is illustrated with coloured ellipses in the upper panel of Fig 3.1. These groups also appear in action space, as extended or elongated overdensities in the radial action, J_R. Besides these moving groups, we also see more ridge-like features extending along higher J_R values, where they appear at almost constant L_z or J_ϕ. The middle panel of this figure shows stars within $200 < 1/\varpi < 600$ pc, and in (U, V) or $(-v_R, v_T)$ velocity space we notice that there is not much sub-structure left, however, action space still shows significant features in the form of extended ridges along J_R. Finally, at $600 < 1/\varpi < 1500$ pc in velocity space we see a smooth distribution, with no overdensities, whereas action space still shows some extended ridges. When going towards larger distances, beyond $1/\varpi = 200$ pc almost no feature at all can be easily identified in velocity space, this illustrates the power of action space.
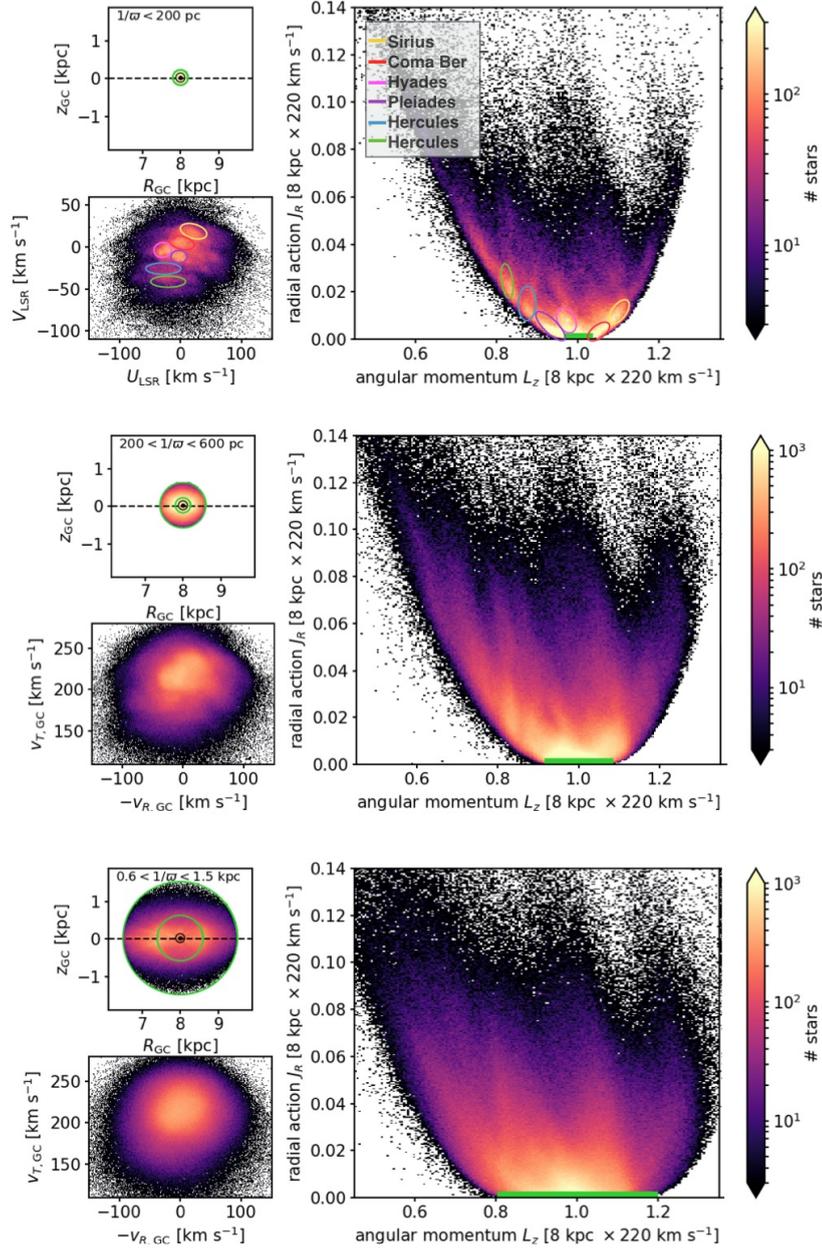
FIGURE 3.1: *Right panels:* Action distribution in L_z (or J_ϕ) vs J_R for ~ 3.8 million stars with radial velocities in GDR2. The green horizontal bar indicates the radial extent in L_z. *Left panels:* Corresponding distribution of these stars in Galactocentric coordinates (R, z) and U, V velocities for $1/\varpi < 200$ pc and $-v_R, v_T$ for $1/\varpi > 200$ pc. In the R, z figures the location of the Sun is indicated with $\odot$, the dashed line represents the Galactic plane and the annuli in Solar distance is shown in green. The approximate location of different moving groups in the solar neighborhood is indicated by coloured ellipses in the velocity and action plots. This figure clearly illustrates the great amount of –and very extended– orbit structure present in Gaia DR2. Adapted from Fig. 1 in Trick, Coronado, & Rix (2019).
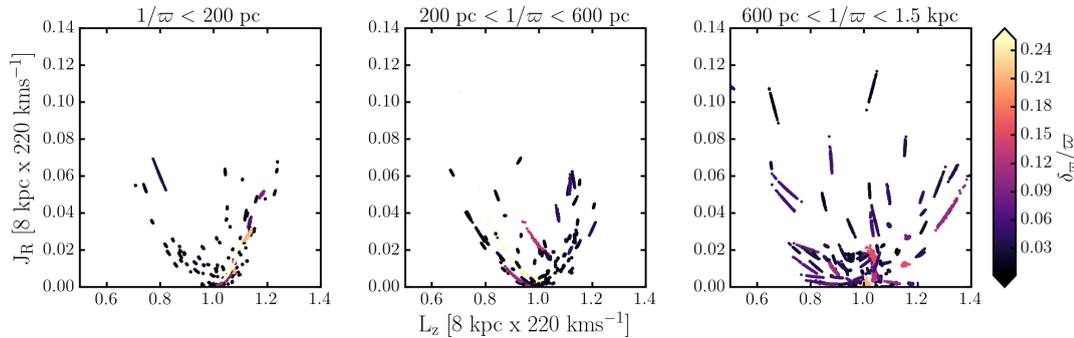
FIGURE 3.2: The effect of measurement uncertainties when calculating actions. Specifically this figure shows 100 samples of measurement uncertainty ellipse transformed to action space performed via Monte Carlo sampling, color coded by the fractional error in parallax (δ_ϖ/ϖ). At larger distances, specifically for stars at 600 pc$< 1/\varpi < 1.5$ kpc, the effect of the measurement uncertainties becomes important.

3.3 The effect of measurement uncertainties

At this point, it is important to analyse the extent of parallax, proper motion and velocity uncertainties, and how they translate into action space. As already mentioned, there are significant features present in action space, and thus we have to investigate if these are affected by the measurement uncertainties. We do this for the three distance regimes defined in the previous section: $1/\varpi < 200$ pc, 200 pc $< 1/\varpi < 600$ pc and 600 pc$< 1/\varpi < 1.5$ kpc. We choose 100 random stars in each of the distance bins, and we draw 100 Monte Carlo samples from the uncertainty ellipse in (ϖ, $\vec{\mu}$, v_{los}), including their respective covariances, and we translate them into action space. We show the result in Fig. 3.2.

At close distances ($1/\varpi < 200$ pc), the majority of stars have small errors in parallax, hence the ellipse distribution for these stars is also small. We notice that the uncertainty ellipses are narrower for stars at closer distances compared to larger distances, where they show a a more extended distribution in action space. As discussed in Chapter 3, Section 2.5.2, this could blur or smear out the sub-structure present in action space, particularly at larger distances.

We notice that stars in the solar neighbourhood ($1/\varpi < 200$ pc) have uncertainty ellipses typically smaller than the observed structures. Consequently, we expect that the features present in action space are real signatures in the Galaxy. However, when exploring larger distances, taking $1/\varpi$ as the distance may not be the best choice, as the parallax error has a great impact when translated into action space, with ellipses that are now comparable in size with the general features.

3.4   Final remarks

From Fig 3.1, we established the great amount of sub-structure present in action space for our Galaxy, where we observe several ridges at different locations in J_ϕ–J_R. For instance, the ridges observed at $1/\varpi < 200$ pc coincide with the position of known moving groups. These features are not only present at close distances, but they go beyond the moving groups and extend up to 1.5 kpc distances.

Additionally, several studies (Sellwood & Binney, 2002; McMillan, 2011b; Sellwood, 2012; Sellwood et al., 2019) have investigated the effect of non-axisymmetric perturbations (bar and spiral arms) and resonances in the Galactic disc. For example, resonances are found to cause narrow ridges at lines following $\sim$ constant $\Delta J_R / \Delta J_\phi$ in action space. Stars that are located near a resonance (in J_ϕ) move away from circular to orbits that become progressively more eccentric at larger J_R (Fouvry, Binney, & Pichon, 2015). Consequently, the rich structure we identified in action space contains much information on non-axisymmetric perturbations.

Furthermore, action space reveals more features than velocity space, which shows the capability of actions as a very powerful coordinate system. Finally, we stress the importance of having improved distances when going to large distances, for example, already at $1/\varpi > 600$ pc, some of the observed structure at that distance regime could be blurred out as a product of measurement uncertainties in the parallax.

4 | From birth associations to field stars: mapping the small-scale orbit distribution in the Galactic disc

Stars born at the same time in the same place should have formed from gas of the same element composition. But most stars subsequently disperse from their birth siblings, in orbit and orbital phase, becoming 'field stars'. Here we explore and provide direct observational evidence for this process in the Milky Way disc, by quantifying the probability that orbit-similarity among stars implies indistinguishable metallicity. We define the orbit similarity among stars through their distance in action-angle space, $\Delta(J, \theta)$, and their abundance similarity simply by $\Delta[\text{Fe/H}]$. Analyzing a sample of main sequence stars from Gaia DR2 and LAMOST, we find an excess of pairs with the same metallicity ($\Delta[\text{Fe/H}] < 0.1$) that extends to remarkably large separations in $\Delta(J, \theta)$ that correspond to nearly 1 kpc distances. We assess the significance of this effect through a mock sample, drawn from a smooth and phase-mixed orbit distribution. Through grouping such star pairs into associations with a friend-of-friends algorithm linked by $\Delta(J, \theta)$, we find 100s of mono-abundance groups with ≥ 3 (to $\gtrsim 20$) members; these groups – some clusters, some spread across the sky – are over an order-of-magnitude more abundant

than expected for a smooth phase-space distribution, suggesting that we are witnessing the 'dissolution' of stellar birth associations into the field.

4.1 Introduction

The ever increasing amount of stellar spectra collected by spectroscopic Milky Way surveys, such as APOGEE (Majewski et al., 2017a), GALAH (Buder et al., 2018), LAMOST (Cui et al., 2012; Zhao et al., 2012), RAVE (Casey et al., 2017) amongst others, provides precise information on the element abundances of millions of stars. Combining these surveys with the second release of the Gaia satellite (Gaia Collaboration et al., 2018a), that constrains the 6D phase space and orbit information for these stars, opens up the possibility to understand how our Galaxy has formed and evolved.

By studying the population properties of stars, their orbits, compositions, and ages we can learn about the assembly of different components of the Galaxy. Furthermore, clusters, either intact, dispersing or dissolving can teach us about the dynamical history of the Galaxy (e.g., Allison, 2012; Webb et al., 2013; Ting, Conroy, & Goodman, 2015).

One conceptual approach is the idea that stars that were born at the same time and in the same molecular cloud can reveal their common birth origin by their very similar chemical abundances (Freeman & Bland-Hawthorn, 2002), even when they could have been dispersed into different places afterwards (Ting, Conroy, & Goodman, 2015; Hogg et al., 2016). This is called 'chemical tagging'.

That stars disperse can mean that they are in different orbital phases (different locations) along nearly the same orbit. For example, when a star cluster gets disrupted, single-stellar population "streams" can be found extending for tens of kpc through the Galactic halo (Bovy, 2014; Contenta et al., 2017). Or this could mean that stars actually evolve to very different orbits, and this can happen through radial mixing or radial migration (e.g., Sellwood & Binney 2002; Roškar et al. 2008; Quillen et al. 2015). We now have clear and quantitave evidence that this migration is overall strong in the Galactic disc (e.g., Frankel et al. 2018).

At late times (last 8 Gyrs), stars in the Milky Way were presumably born on disc-like orbits. However, discs are susceptible to perturbations, and fluctuations in the gravitational field cause a star to change its original orbital actions, or diffuse in action space (e.g., Fouvry, Binney, & Pichon 2015). When radial migration is not at play, then we would expect that stars at a given radius have a clear relation between the age and metallicity of stars (Sellwood & Binney, 2002), reflecting the successive enrichment of the birth gas. However observations have not shown such correlation in the solar neighborhood and instead

have shown a large spread in metallicity, [Fe/H] (Edvardsson et al., 1993; Ibukiyama & Arimoto, 2002), implying that stars have in fact migrated over large radial distances during their lifetime (Bland-Hawthorn, Krumholz, & Freeman, 2010).

The task of identifying groups of stars from the same cluster purely by their chemical similarity, without information on velocity or distance, has been proposed (Freeman & Bland-Hawthorn, 2002) and put into practice by Hogg et al. (2016), Schiavon et al. (2017), and Garcia-Dias et al. (2019), amongst others. In this scenario, for chemical tagging to be successful, one of the conditions is that the progenitor cloud is uniformly mixed before the first stars are formed (Freeman & Bland-Hawthorn, 2002). In addition, birth clusters must have clear cluster-to-cluster abundance differences (Liu et al., 2016b). Open clusters are good laboratories for testing whether these conditions hold (Blanco-Cuaresma et al., 2015; Liu et al., 2016a; Bovy, 2016).

In recent years, data-driven methods have been used to extract high-precision abundances from spectra, even at moderate resolution and signal-to-noise (Ness et al., 2015; Rix et al., 2016; Ting et al., 2017; Ting et al., 2019). However, pure chemical tagging is still a challenging technique (Ting, Conroy, & Goodman, 2015; Blanco-Cuaresma & Soubiran, 2016), as shown by the presence of *doppelgangers* in field stars (Ness et al., 2018).

Once precise abundances have been determined, a procedure is needed to identify potentially co-natal "clumps" in abundance space. Some works make use of clustering algorithms such as *k-means* (e.g, Hogg et al. 2016) or the density-based spatial clustering of applications with noise (DBSCAN; Ester et al. 1996) (e.g., Shou-kun et al., 2019; Price-Jones & Bovy, 2019). While for *k-means* the number of clusters must be known in advance, and specified *a priori* in the algorithm, with DBSCAN the optimal number of clusters can be determined from the data in an automated way.

Most stars formed in a molecular cloud are expected to disperse quickly, in $\lesssim 100$ Myr (Lada & Lada, 2003), in orbit and consequently in orbital phase on a longer timescale. However, their observable chemical abundances are expected to remain largely unchanged. Including more dimensions than just chemical information (e.g. kinematics) increases the prospect of tracing back the origin of a dispersed cluster. Therefore exploring the extent to which stars with very similar abundances are also on similar or different orbits is a fundamental diagnostic. In the Galactic disc this tells us directly how strong radial migration was i.e., whether the present day orbit of normal disc stars has anything to do with their birth orbit.

Kamdar et al. (2019) has recently shown the existence of this dispersal by revealing that co-moving (in $\vec{x}, \vec{v}$) pairs in the solar neighborhood have a preference to have similar metallicities when compared to random field stars, even to distances beyond bound

pairs. This opens up the possibility to find disrupting star clusters. However, a Cartesian coordinate system, as the one used in that work, may not be optimally suited to identify such signatures beyond the sun's vicinity.

Action-angles (J,θ) are canonical coordinates to describe stellar orbits, and they may be a powerful coordinate system to find orbit-distribution sub-structure in our Galaxy. Whereas in configuration space ($\vec{x}, \vec{v}$) each of the coordinates has a complex time evolution, in action-angle space the three actions are integrals of motion and constant, and the three angles evolve linearly with time (Binney & Tremaine, 2008). Additionally, gradual changes in orbit may be described as a diffussion in action space (Sanders & Binney, 2015). Hence, if we want to study larger volumes in the Milky Way (e.g, $d \gtrsim 200$ pc, where the curvature of stellar orbits becomes pronounced), then this coordinate system may be better to identify stars that are on the same orbits, as compared to a Cartesian coordinate system (X, Y, Z, U, V, W). A cylindrical coordinate system (R, ϕ, z) could also be used as a better spatial alternative over a larger region of the Galaxy. Action-angles have already been used to study groups of stars on similar orbits, for example, Trick, Coronado, & Rix (2019) have revealed rich orbital sub-structure in Gaia DR2, that extends over several kpc. Action-angles are also convenient to study processes that might be responsible for orbit migration in the Galactic disc, like spiral arms (Sellwood et al., 2019) and bars (Hunt et al., 2019; Trick et al., 2019).

Here we combine the spectroscopic information from LAMOST's latest data release, LMDR5 (Xiang et al., 2019) with the astrometric information from Gaia DR2 to investigate the probability that star pairs that are close in action-angle space have exceptionally similar metallicities, through $p(\Delta[\mathrm{Fe/H}] \mid \Delta(J, \theta))$. We start by defining a metric in action-angle space, combined with chemical information, in a *generalised chemical tagging* approach. On this basis, we can show that the width of $p(\Delta[\mathrm{Fe/H}] \mid \Delta(J, \theta))$ grows continually with increasing $\Delta(J, \theta)$, from the regime of bound binaries to disc-halo pairs of stars, well beyond the distance regime probed in Kamdar et al. (2019). To see whether these ultra-wide pairs of stars trace the dispersal of birth associations, we then apply a friends-of-friends algorithm to stars of near-identical [Fe/H] to recover larger structures, recovering both known open clusters and widely dispersed groups. This method could constrain effects such as orbit diffusion in the Galactic disc.

This Chapter is organised as follows: in Section 2 we present the data used in this study, observational and a mock catalog, in Section 3 the method: pairwise distances between stars, in Section 4 we present the results and analysis of the *generalised chemical tagging*, in Section 5 the orbit clustering of stars with the same metallicities, Section 6 presents a comparison in ($\vec{r}, \vec{v}$) configuration space; and finally Section 7 presents the summary followed by the appendix.

4.2   Data

4.2.1   The Gaia DR2 ⊗ LAMOST DR5 Sample

The analysis of this Chapter draws on the combination of the second Gaia data release, GDR2 (Gaia Collaboration et al., 2018a), and the fifth data release (DR5) of the spectroscopic survey LAMOST (hereafter LMDR5) with stellar parameters derived from the Data-Driven Payne (DD-Payne, Xiang et al. (2019)), which is a data driven model that includes constrains from theoretical spectral models to derive abundances. We obtain the positions (ra, dec), proper motions (μ_{ra}, μ_{dec}) and the parallaxes ϖ from GDR2, where we impose selection criteria on the renormalized unit weight error ≤ 1.6 (Lindegren et al., 2018) and on the parallax $\varpi > 0$.

LAMOST provides spectra at a resolution of R ~ 1800. We consider only stars with $\mathrm{SNR}_G > 30$ in LMDR5 to decrease the uncertainties in [Fe/H]. For this subsample, the typical radial velocity precision is $(5-7)\,\mathrm{km\,s^{-1}}$, and the typical abundance precision is $\sim$ $0.05-0.07$ dex for [Fe/H]. We make use of the spectroscopic parameters $T_{\mathrm{eff}}, \log g$, [Fe/H], and also the radial velocities. For this work we make use of the recommended labels that combine results from the LAMOST-GALAH and LAMOST-APOGEE training sets, where we have selected stellar labels with no flags (Xiang et al., 2019).

Following the procedure described in Chapter 2, we calculate the spectro-photometric distances that combine the parallaxes and spectral information. We consider main sequence (MS) stars with the following criteria: $4800\mathrm{K} < T_{\mathrm{eff}} < 6000\mathrm{K}$, $\log g > 4.2$. This selection of MS stars in T_{eff} differs from the one adopted in Chapter 2, as estimates of [Fe/H] become less robust and accurate for $T_{\mathrm{eff}} < 4800\mathrm{K}$. We then combine the dataset with *2MASS* to obtain the *K*-band magnitude, needed to apply our spectrophotometric distance model. Otherwise, we essentially follow here the model of Chapter 2, with further slight changes explained in more detail in Appendix A.1. We are left with $\sim$ 550 000 MS stars after the GDR2⊗LMDR5 cross-match and selection criteria. The stars in the sample here have distances up to 3 kpc, however the majority of them are at $d <$ 1.5 kpc.

4.2.1.1   Wide Binaries in LAMOST as methodologial anchors

In addition to our primary analysis of all possible pairs within the GDR2⊗LMDR5 catalog, we also analyze a sample of 519 gravitationally bound wide binaries (WBs) for which both components have a high-quality spectrum from LAMOST. WBs represent the extreme low-$\Delta(J, \theta)$ limit for pairs close together in phase space: they not only have similar kinematics

and in most cases formed from the same gas cloud, but they are still gravitationally bound. Because WBs are generally chemically homogeneous (e.g., Hawkins et al., 2020), the distribution of $\Delta[\mathrm{Fe/H}]$ within the WB sample represents the highest degree of chemical homogeneity we can expect to measure for stars formed at the same time and place within the Milky Way, given the noise properties of the GDR2⊗LMDR5 sample.

We select wide binaries using the same general procedure described in El-Badry & Rix (2018): we identify pairs of stars with projected separations $s < 50,000\,\mathrm{AU}$ that have parallaxes and proper motions consistent with bound Keplerian orbits and both have high-quality LAMOST spectra. The measured $\Delta(J, \theta)$ for WBs is necessarily low, but it is nonzero because (a) the nonzero orbital velocities cause the total space velocities of the components of WBs to differ at the $\sim 1\,\mathrm{km\,s^{-1}}$ level, and (b) uncertainties in the parallaxes and proper motions of both components inflate their $\Delta(J, \theta)$ to the noise floor.

In contrast to the binary selection procedure of El-Badry & Rix (2018), which relied only on 5D Gaia astrometry, we also make use of LAMOST radial velocities in our selection, requiring the radial velocities of the two components to be consistent within 2σ (Fig. A.2 in Appendix A.2). This allows us to search for WBs out to a distance of $2\,\mathrm{kpc}$ while maintaining a low contamination rate. We refer to El-Badry et al. (2019) and Tian et al. (2020) for detailed discussion of the wide binary selection procedure, contamination rate, and effective selection function. In this work, we restrict our analysis to the highest-quality subsample of the WBs: those which both components have a LAMOST spectrum with $\mathrm{SNR}_G > 50$ and precise Gaia astrometry ($\varpi/\sigma_\varpi > 10$).

4.2.2 A Mock Catalog with a Smooth and Phase-Mixed Orbit Distribution

As a null hypothesis for our analysis, we need to understand the amount of clustering we expect to find in the case that all stars are in a smooth orbit distribution fully phase-mixed, where [Fe/H] only changes gradually with the "Galactic component", or radius.

We do this by creating a mock observation that matches our GDR2⊗LMDR5 selection in volume and depth, based on the Gaia DR2 mock stellar catalog by Rybizki et al. (2018). This catalog was created using a chemo-dynamical model based on Galaxia (Sharma et al., 2011), where the stars are sampled from the Besançon Galactic model (Robin et al., 2003). The 2003 Besançon model prescribes smooth distributions in phase and abundance-space to the four main Galactic components (thin, thick-disc,bulge and halo), with basic observational constraints, like the age-velocity-relation, age metallicity distribution and radial metallicity gradient, imprinted. It should be noted that the sampled version in GDR2 mock neither includes binaries nor spiral arms (any localised/clumpy star formation). We select stars in this GDR2 mock with criteria resembling those of our
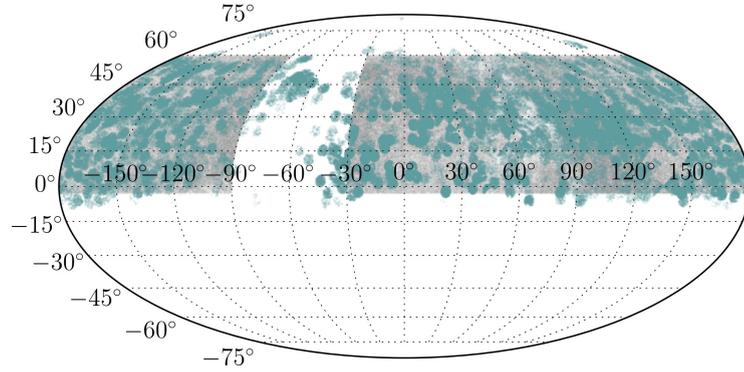
FIGURE 4.1: Footprint on the sky in equatorial coordinates of the LMDR5 sample in
blue dots, and the mock catalog in grey. Here we can see that the LAMOST survey
covers the northern hemisphere, and we select the stars in the mock catalog accordingly,
considering the areas of the sky that were mostly completely covered with LMDR5 stars.

dataset: in T_{eff} and $\log g$, with additional cuts in parallax and magnitude: $\sigma_\varpi/\varpi < 0.1$
and $10<$ `phot_g_mean_mag` < 14. The values provided in the catalog are noise-free,
hence parallaxes could be directly inverted to give exact model distances (Rybizki et al.,
2018). We proceed to add noise to the parallax by sampling from a Gaussian with the
true value of ϖ as mean and σ_ϖ as the standard deviation, as suggested in Rybizki et al.
(2018).

Then we match the sky coverage of the LAMOST survey, which covers much of the
northern sky (Fig. 4.1). After applying all of these cuts, we are left with $\sim 580,000$ stars
in the GDR2 mock, matching the sample size of our GDR2$\otimes$LMDR5 dataset.

4.3   Methodology: Pairwise Distances in Actions, Angles and [Fe/H]

In order to see if we can find signatures of stars that were born at the same time from the
same material, we investigate if we can quantify statistically how much closer pairs of
stars are in [Fe/H] if they are close in orbit space by studying $p(\Delta[\mathrm{Fe/H}] \mid \Delta(J, \theta))$.

4.3.1 Choice of variables: $\Delta(J, \theta)$ and Δ[Fe/H]

In the following sub-sections we describe operationally how to define and then calculate distances between pairs of stars, both in orbit space and in abundance space.

For two stars that are on nearly the same orbit and nearly the same orbital phase, their distance can be well defined in the Cartesian configuration space $(\vec{x}, \vec{v})$, as done by Kamdar et al. (2019). For wider separations, one could use classical integrals E and L_z (Jeans, 1916; Contopoulos, 1963; Ollongren, 1965; Binney & Spergel, 1984). But action-angle variables (J, θ) are arguably the best set of coordinates, as they form a 6D canonical coordinate system with several advantages. For an axisymmetric gravitational potential, all three actions (J_R, J_z, J_ϕ) are integrals of motion (Binney, 2012), where J_R quantifies the oscillations of the orbit inwards and outwards in the radial direction, J_z quantifies the oscillations in the vertical direction and J_ϕ (or L_z) is the azimuthal action that equals the angular momentum in the z direction. All actions have the same units, kpc×km/s. Actions are complemented by their three corresponding angles: θ_R, θ_z and θ_ϕ. These angles, reflecting the orbital phase in these coordinates, increase linearly with time, in practice modulo 2π. If a system is fully phased mixed, then the angles should be uniformly distributed between 0 and 2π.

In the presence of non-axisymmetric structures such as spiral arms or a bar, the three actions J_R, J_z, J_ϕ are not well defined, and are not exactly integrals of motion. However, (axisymmetric) approximations can still be made to compute them. For a thorough description of action angle variables, we redirect the reader to Section 3.5 of Binney & Tremaine (2008).

4.3.2 Action-angle computation

The calculation of actions and angles requires both phase-space coordinates, and an (assumed) gravitational potential. If we assume that the Galaxy's potential is close to an axisymmetric Stäckel potential, then the actions and angles can be easily calculated. We make use of the python package `galpy`, with its implementation of the action estimation algorithm *Stäckel fudge* (Binney, 2012) along with the `MWPotential2014` model. The latter considers a simple axisymmetric Milky Way potential model with a circular velocity of 220 km/s at the solar radius of 8 kpc (Bovy, 2015). Note that the absolute values of the actions never enter the subsequent analysis, just their differences. So, the choice of an updated circular velocity (e.g., Eilers et al. (2019)) would not significantly alter the results.

For the location and velocity of the Sun within the Galaxy we assume $(X,Y,Z) = (8,0,0.025)$ kpc and $(U,V,W)_\odot = (11.1,12.24,7.25)$ km/s (Schönrich, Binney, & Dehnen, 2010) to first calculate Galactocentric coordinates and then actions from the observed (ra, dec, d, v_{los}, μ_{ra}, μ_{dec}) of each star. As noted in Chapter 2, the largest contribution to the action uncertainties comes from the distances. However, by calculating the spectrophotometric distances as described in Appendix A.1, we obtain improved distances (at least for distant stars) with uncertainties of $\sim 7\%$ for single stars. We refer the reader to Section 2.5.2.1 in Chapter 2 to see the extent of the uncertainties in action space when applying this spectrophotometric distance model. Typical action uncertainties are $\sim 5 - 8\%$.

4.3.3   Defining a Metric in Action space

To calculate the pairwise distances between stars in action space, we first must define a metric that combines the three actions J_R, J_z and J_ϕ. For subsequent combination with the angle separation metric, we want this metric—or distance—to be unitless. Therefore we normalise each dimension by the ensemble variance in each quantity, defining the distance between a pair of stars (i,j) in action space as

$$\Delta J_{ij}^2 \equiv w_{J_R} \cdot (J_{R,i} - J_{R,j})^2 + w_{J_z} \cdot (J_{z,i} - J_{z,j})^2 + w_{J_\phi} \cdot (J_{\phi,i} - J_{\phi,j})^2 \qquad (4.1)$$

with

$$w_{J_k} \equiv \frac{1}{\mathrm{Var}(J_k)}, k \in \{R, \phi, z\} \qquad (4.2)$$

where the variance is defined as $\mathrm{Var} = \sum_{i=1}^{N}(x_i - \bar{x})^2/(N-1)$ for a sample size N.

4.3.3.1   Defining a Metric in Action-Angle space

Stars that drift apart in orbit space will then also drift apart in orbital phase, unless the orbital frequency stays identical. Therefore, the distance between two stars in orbital phase, or angle, matters. We define an analogous 6D metric that combines the actions J_R, J_z, J_ϕ with their respective angles θ_R, θ_z and θ_ϕ.

We start with the angle part:

$$\Delta \theta_{ij}^2 \equiv w_{\theta_R} \cdot \Delta\theta_{R,ij}^2 + w_{\theta_z} \cdot \Delta\theta_{z,ij}^2 + w_{\theta_\phi} \cdot \Delta\theta_{\phi,ij}^2. \qquad (4.3)$$

$w_{\theta_k} \equiv \frac{1}{\text{Var}(\theta_k)}, k \in \{R, \phi, z\}$, with

$$\Delta\theta_{k,ij} \equiv \min\left[|\theta_{k,i} - \theta_{k,j}|, (2\pi - |\theta_{k,i} - \theta_{k,j}|)\right], k \in \{R, \phi, z\} \tag{4.4}$$

By definition, the angles θ_k are in the range and periodic in $[0, 2\pi]$. Eq. 4.4 ensures that the correct (and smallest) angle distance is used. Again, we introduce a normalisation factor w_θ for each of the angles. We note that the variance in θ_R and θ_z has roughly the same value, we can see stars in basically all phases of their vertical and radial oscillation. For the azimuthal direction, defined to be 0 at the line from the Sun to the Galactic center, only a small fraction of angles will be within the sample volume. The values that we considered for the variance are the typical distance two stars can have in the angles. We use the same weights for the real data and the mock catalog. Combined, this yields a sensible action-angle distance metric:

$$\Delta(\text{J}, \theta)_{ij}^2 \equiv \Delta J_{ij}^2 + \Delta\theta_{ij}^2, \tag{4.5}$$

where both components of the metric are unitless.

4.3.4 Distance in Abundance Space: Δ[Fe/H]

We define a distance in abundance space by considering the differences in [Fe/H] exclusively. This is for several reasons: [Fe/H] has the largest variance compared to [X/Fe], it is robustly determined and it is available also in the mock catalog. Then, the pairwise distance for the metallicity is defined as:

$$\Delta_{ij}[\text{Fe/H}] \equiv |[\text{Fe/H}]_i - [\text{Fe/H}]_j| \tag{4.6}$$

For our dataset, the uncertainties in [Fe/H] are less than 0.1 dex.

With these definitions, we can proceed to explore the action-angle and metallicity distances between pairs of stars, as illustrated in Fig. 4.2, where we show the complete distribution of the pairwise distances $p(\Delta[\text{Fe/H}] \mid \Delta(J, \theta))$ that we obtain.
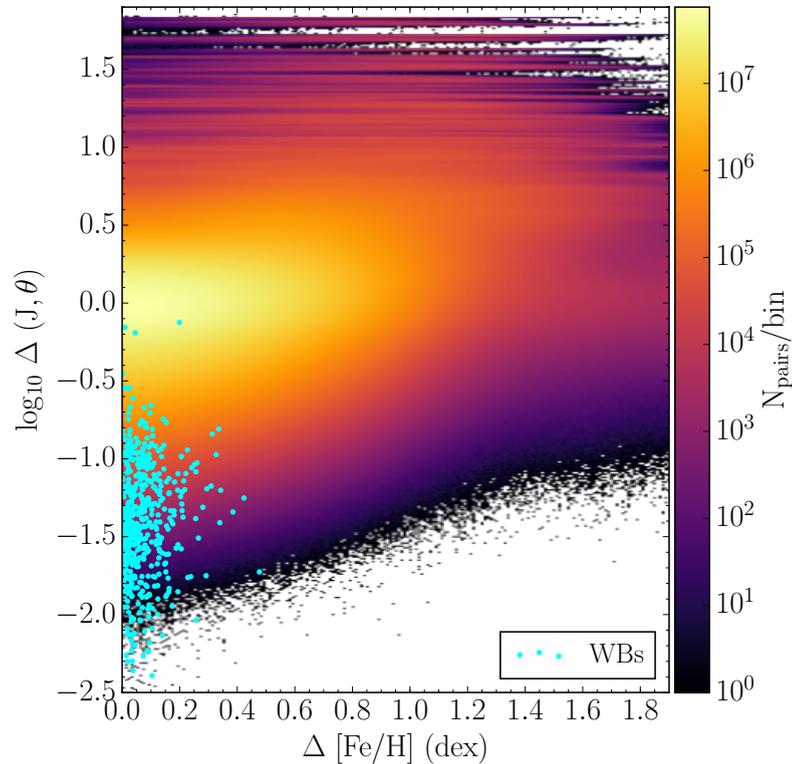
FIGURE 4.2: Distribution of pairwise distances in action-angle and metallicity space for MS stars in LAMOST $\times$ Gaia DR2, as defined in Eqs. 4.3, 4.4, 4.5 and 4.6. In cyan dots we show how Wide Binary pairs (WBs) are distributed in this same space. WBs are mostly concentrated at small distances in both action-angle, $\log_{10}\Delta(J,\theta)$ space and metallicity Δ[Fe/H], as expected from stars that were born together. There is a smooth transition from stars close in [Fe/H] - (J,θ) towards stars at larger [Fe/H] and (J,θ) distance. The typical uncertainties in $\Delta(J,\theta)$ are $\sim 5\%$. The bin size is 0.01 in this figure.

4.4 Generalised chemical tagging analysis: Orbit-similarity vs. Abundance-Similarity

In Sec. 4.3 we have defined pairwise distances between stars, both in action space only and in the full action-angle space. We will mainly analyse the results from a 6D phase-space metric, combined with chemical information. However, in Appendix A.3, we also show the results for the distance in action only space (i.e., a 3D coordinate system) also combined with [Fe/H].

4.4.1 Abundance differences of stars on similar orbits:
$$p\big(\Delta[\text{Fe/H}] \mid \log_{10}\Delta(J,\theta)\big)$$

In Fig. 4.2, we present the distribution of distances in action-angle space (using the metric defined in Eqs. 4.3, 4.4, 4.5) vs. $\Delta[\text{Fe/H}]$ for all $\sim 10^{11}$ stellar pairs in our sample. The peak of the distribution is reached at $\sim \log_{10}\Delta(J,\theta) = 0$ which by construction is the mean pair separation. We have additionally divided the distances by the number of dimensions, 6 in action-angle space and by 3 in the action only case. This figure already illustrates the broad trend that stars close in $\Delta(J,\theta)$ tend to be close in $\Delta[\text{Fe/H}]$ and *vice versa*. The extremes are wide binaries (bottom left of Fig. 4.2) and presumably disc-halo pairs (top right of Fig. 4.2). Those latter pairs would differ in both chemical composition and they would also be in completely different orbits. Overall, this shows that stars that are on similar orbits and close in the phase angles have also similar metallicities. In this figure we also show the distribution of WBs in action-angle and metallicity space. We discuss in more detail this sample in Appendix A.2.

To quantify this effect and put it into perspective, we compare the distribution of $p\big(\Delta[\text{Fe/H}] \mid \log_{10}\Delta(J,\theta)\big)$ for the observed in data to an idealized mock galaxy, that has broad population gradients, but no clustered star-formation. As mentioned in Sec. 4.2.2, we make use of the Gaia DR2 mock stellar catalog by Rybizki et al. (2018).

In Fig. 4.3, we present the cumulative distribution function (CDF) of stars as a function of $\Delta[\text{Fe/H}]$. The left side of this figure shows the GDR2⊗LMDR5 dataset (that we will now call the real MW pairs) where each coloured line represents the CDF for different orbit-similarity bins in Fig. 4.2 separated by 0.5 in $\log_{10}\Delta(J,\theta)$. The right panel in Fig. 4.3 shows the same but for the mock data pairs: Following the same procedure as for the real MW pairs, we calculated the metric in action-angle and [Fe/H] space for the mock data pairs, with Eqs. 4.3, 4.4, 4.5 and 4.6. We use the same values for the variance (Eq. 4.2) that we obtain from the real MW pairs, for the mock ones given that their values are similar. We obtain a histogram in $\log_{10}\Delta(J,\theta)$-$\Delta[\text{Fe/H}]$, and then using the same bins as for the real MW pairs, we produce the CDF. In both figures the cyan line shows the complete CDF of the WB pairs as presented by cyan dots in Fig. 4.2, where the WBs have $p(\Delta[\text{Fe/H}] = 0.1) \sim 0.8$. This clearly shows that most of the distribution of WBs is in fact close in [Fe/H].

For the real MW pairs, we find that in the smallest bin in $\log_{10}\Delta(J,\theta)$, $\sim 60\%$ of the pairs have metallicity differences within the measurement uncertainty of 0.1 dex. As for the mock data pairs, we find that the smallest bin in action-angle distance has $\sim 40\%$ of pairs at 0.1 dex in $\Delta[\text{Fe/H}]$. Given the large sample sizes, these differences are highly significant. Most importantly, the CDF's in the five closest orbit bins in the mock data
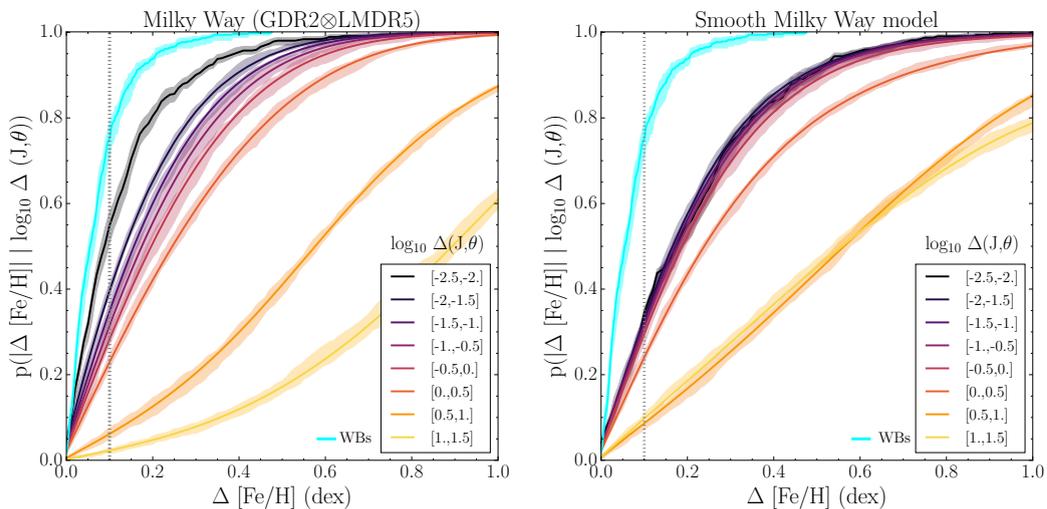
FIGURE 4.3: Correlation between $\Delta(J,\theta)$ and $\Delta[Fe/H]$ in the Milky Way vs. a mock Galaxy with no clustered star formation. Here we show the CDF of pairs in given distance bins $\Delta(J,\theta)$ as a function of Δ [Fe/H], for the LMDR5 × Gaia DR2 MS stars in the left, and the mock catalog in the right. The width of these lines show the 5th and the 95th percentile of a bootstrap re-sampling. The cyan line shows the WBs, for comparison. The dashed line is located at $\Delta[Fe/H] = 0.1$, that we consider as an upper limit for the uncertainties in [Fe/H]. We observe that for the first bin—with the smallest $\log_{10}\Delta(J,\theta)$ (black line)— $\sim$60% of pairs with action-angle distances $\log_{10}\Delta(J,\theta) < -2$ have metallicity within the uncertainty. As for the mock catalog, we see that for the first 5 bins, the lines are located at almost the same position, and we find a smaller value for $p(\Delta[Fe/H] \mid \log_{10}\Delta(J,\theta))$ for the smallest $\log_{10}\Delta(J,\theta)$ than the one shown by the data.

pairs are nearly identical, but the fraction of pairs with indistinguishable $\Delta[Fe/H]$ rises towards small $\Delta(J,\theta)$ in the real MW pairs. As for the large $\Delta(J,\theta)$ bins in the real MW pairs, the separation between them becomes wider, this is because in this regime we would expect to find more random pairs, that are not actually physically related. They are not only far apart in the $\Delta(J,\theta)$ metric, but also in [Fe/H]. The highest $\Delta(J,\theta)$ bin in the mock data pairs does not show a strong difference in [Fe/H] as the one observed in the real MW pairs, this is because of how the different components (thick disc, halo and bulge) are simulated in that catalog (Rybizki et al., 2018).

4.4.2 The fraction of stars with the same [Fe/H], as a function of orbit similarity: $f_{pairs}\big(\log_{10}\Delta(J,\theta) \mid \Delta[Fe/H] < 0.1\big)$

We now consider a statistic that perhaps speaks more immediately to the question of whether we see birth associations of stars disperse and transition of field stars. Specifically we consider the fraction of pairs at a given $\log_{10}\Delta(J,\theta)$ that have indistinguishable [Fe/H], $f_{pairs}\big(\log_{10}\Delta(J,\theta) \mid \Delta[Fe/H] < 0.1\big)$. As we consider larger orbit separations $\log_{10}\Delta(J,\theta)$ the chances of finding pairs of different birth origin should increase, and
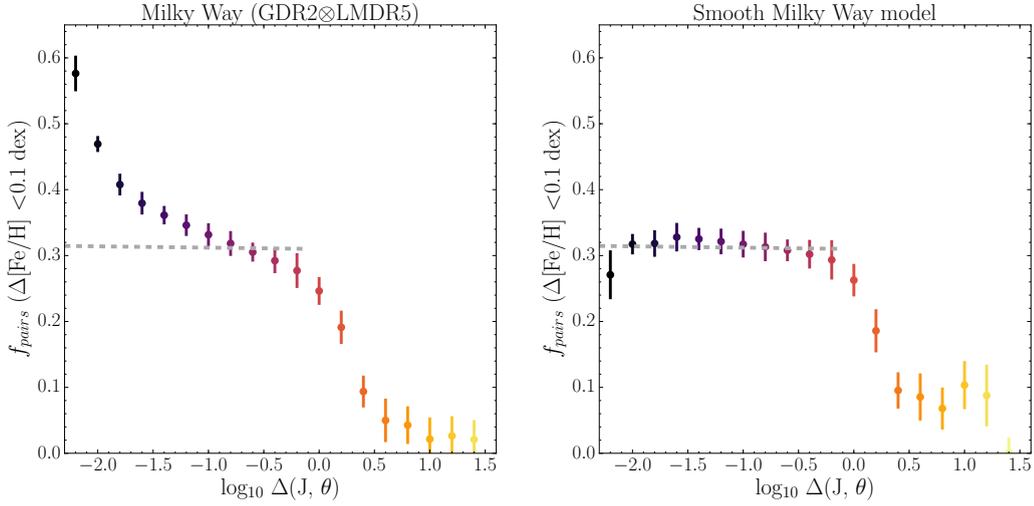
FIGURE 4.4: Fraction of pairs with indistinguishable metallicities ($\Delta[\text{Fe/H}] < 0.1$ dex) at different bins in $\log_{10}\Delta(J,\theta)$ for the MW pairs to the left, and the mock pairs to the right. Each colored dot corresponds to bins of 0.2 in $\log_{10}\Delta(J,\theta)$. The colors in this plot indicate different bin values, similarly to Fig. 4.3. This clearly illustrates the differences between mock and MW pairs, where we find that the fraction of mock pairs is roughly flat ($\sim 31\%$) for these 9 bins. Whereas for the MW pairs we see that for the first 2 bins the fraction of pairs is $\sim 60\%$ and $\sim 45\%$ respectively. The grey dashed line is a fit to 9 bins in the mock pairs –excluding the first bin– that we then over-plot also in the left panel with the MW pairs. Therefore, we find that there is a large fraction of real MW pairs at small $\Delta(J,\theta)$ with similar metallicities when compared to the mock pairs. In both mock and MW pairs the fraction of pairs decreases below 1% for the last two bins in $\log_{10}\Delta(J,\theta)$.

$f_{pair}(\Delta[\text{Fe/H}] < 0.1)$ should decrease. We choose $\Delta[\text{Fe/H}] < 0.1$ to denote indistinguishable [Fe/H] as our individual metallicity precision is about 0.07 dex. But of course, given the (local) metallicity dispersion of the low-α disk, the condition $\Delta[\text{Fe/H}] < 0.1$ may be satisfied for many star pairs born at different times in different parts of the disc. Such a test can therefore be only 'statistical', and we again put our findings into perspective by comparison with a mock catalog from a smooth galaxy model (with population gradients). The result of this analysis is quite striking, and is summarized in Fig. 4.4. The panels show $f_{pairs}(\Delta[\text{Fe/H}] < 0.1)$ as a function of $\log_{10}\Delta(J,\theta)$ (in bins of 0.2); the left panel shows the observations, the right panel the smooth mock catalog. The majority of real MW pairs in the closest $\log_{10}\Delta(J,\theta)$-bin have indistinguishable [Fe/H], which then decline to $\sim$30% at $\log_{10}\Delta(J,\theta) \sim 0$, and then quite precipitously fall to nearly 0 at $\log_{10}\Delta(J,\theta) > 0.8$ (presumed disc-halo pairs). The right panel, with the analogous analysis from the smooth galaxy model, shows a qualitatively similar behaviour at $\log_{10}\Delta(J,\theta) > 0$. But there is a striking difference for $\log_{10}\Delta(J,\theta) < -0.5$: the fraction of mono-abundance pairs is constant for all smaller $\log_{10}\Delta(J,\theta)$, while the fraction rises steeply for the actual observations, where we see that the fraction of MW pairs for the first two bins is $\sim$ 60% and 45% respectively. Even though the mock galaxy

is based on a chemo-dynamical model, we would not expect to find pairs of stars clumped in action-angle and [Fe/H]. Because the stars in GDR2 mock are distributed smoothly in phase-space there is no clustering (Rybizki et al., 2018). This means that there is a distinct excess of mono-abundance pairs at small orbit-separations in the real data, just as expected if there is a decreasing fraction of birth pairs as $\log_{10} \Delta(J, \theta)$ increases. On the scales of parsecs, this effect has been seen before (Oh et al., 2017; Kamdar et al., 2019). But we now see this effect in our data to far larger distances. This may not be apparent from the X-axis $\log_{10} \Delta(J, \theta)$; therefore we illustrate in Fig. 4.5 how a certain $\log_{10} \Delta(J, \theta)$ translates into typical spatial distances [in pc] or velocities [km/s]. Fig. 4.5 takes the same bins in $\log_{10} \Delta(J, \theta)$ (in the same color-coding as in Fig. 4.4) and calculates for these pairs the mean $\Delta \vec{r}$ and $\Delta \vec{v}$. The differences between the real and smooth mock data in Fig. 4.4 in the first 6 or 7 bins, in Fig. 4.5 now informs us that this corresponds to 10 km/s and nearly 500 pc. In the right panel of this figure we also show the projected distance of the MW pairs, $\Delta \vec{r_\perp}$, illustrating that even for the first bin in $\log_{10} \Delta(J, \theta)$ these pairs are well beyond the WB regime. It appears that by choosing action-angle coordinates, we can trace an excess of mono-abundance stars to quite enormous distances.

It is worth mentioning that we find only a very small fraction of pairs ($\sim 0.2\%$) having [Fe/H] < -0.5, and also we do not find pairs with both [α/Fe] > 0.15 and [Fe/H]< -0.5. Therefore the contribution from thick disc stars is very small.

4.5   Orbit Clustering of Stars with the Same [Fe/H]: Friends-of-Friends Analysis

If the ultra-wide pairs of stars with the same [Fe/H] are the descendants of dispersed birth associations, we would expect not only pairs, but triplets, quadruplets or larger associations of indistinguishable [Fe/H]. That such associations exist has been shown in the immediate vicinity of the Sun (Oh et al., 2017), using Gaia DR1.

In Sec. 4.3, we have defined a metric (Eq. 4.6) that allows us to find pairs of stars that are close in action-angle space (even at considerable distances in configuration-space). Certainly, we can use this metric to search for larger associations or structures than just pairs. One way to do this is to use the friends-of-friends (FoF), or percolation, algorithm that has been widely used in cosmology to identify features like clusters, halos or groups in density fields in N-body simulations and also in observations (White, Cohn, & Smit, 2010; Duarte & Mamon, 2014; Feng & Modi, 2017).

FoF algorithms can identify groups of sample members that can be linked by less than a certain threshold distance, or "linking length", which can be naturally defined for the
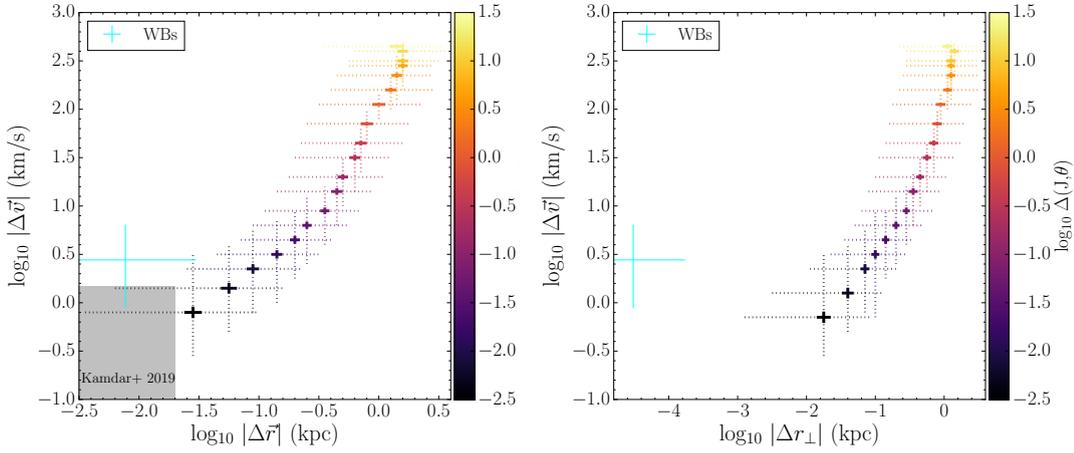
FIGURE 4.5: Differences in 3D velocities $\Delta\vec{v}$ and positions $\Delta\vec{r}$ to the left and transverse distance $\Delta r_\perp$ to the right; both in bins of 0.2 in $\log_{10}\Delta(J,\theta)$ for our MW pairs. In both panels we show the WBs in cyan, that are located at small $\Delta\vec{v}$, $\Delta\vec{r}$ and $\Delta r_\perp$, respectively. For the bins, the solid lines show the uncertainty of the mean value (calculated via bootstrapping), the dashed line –and solid line in the WBs– show the 5th and 95th percentile. To the left, the smallest difference in $\log_{10}\Delta(J,\theta)$ is located also at small differences in velocity-distance space, as expected, given that (J,θ) is only a different coordinate system for the same phase-space as $(\vec{r},\vec{v})$. To the right, the WBs have a projected –or transverse– distance of ~ 0.03 pc whereas the smallest bin in $\Delta(J,\theta)$ is well beyond the WB regime, showing a mean value of ~ 10 pc. Finally, the left side of this figure illustrates that using distances in action-angle space is analogous to position-velocities. The grey rectangle shows Kamdar et al. (2019) co-moving pair selection in $\Delta\vec{v}$ and $\Delta\vec{r}$.

case at hand by our metric in action-angle space. If larger associations are indeed present in the data, we may expect that the contamination by chance coincidences in phase-space (occurring in our mock catalog with smooth orbit and angle distributions) should be far smaller compared to pairs, when considering ensembles of $K = 3, 4, ...$ stars.

We will now briefly sketch the practical implementation, then show the properties of the $K > 2$ associations, and their statistics in the real and mock data. It turns out that the overabundance of such associations in the real data is quite dramatic (over an order of magnitude), compared to the spurious associations in the mock data.

4.5.1 Finding associations with FoF

To find associations based on the FoF algorithm we proceed as follows: after selecting an appropriate linking length, we limit our sample to (a) pairs that are closer than this distance, and (b) pairs that are indistinguishable in metallicity, for which we adopt $\Delta[\text{Fe/H}] < 0.1$. We then consecutively join all distinct pairs that have a star in common, which results in associations of $K \geq 3$ members for any linking length $l = \log_{10}\Delta(J,\theta)$; the

remaining isolated pairs ($K = 2$) are discarded from further consideration. In this procedure, the linking length is a free parameter, for which we will choose a range of values small enough to avoid linking vast numbers of "field stars". After some experimentation, we consider different linking lengths $\log_{10} \Delta(J, \theta)$: $l_i = [-1.8, -1.7, -1.6, -1.55]$.

4.5.2 Properties of the FoF-selected Associations

This FoF search yields a large number of associations with $K \geq 3$, among pairs with $\log_{10} \Delta(J, \theta) < l_i$ that are constrained to pairwise $\Delta[\text{Fe/H}] < 0.1$. We now illustrate the ensemble properties, both in action-angle space and in the space of direct observables, for a few particularly large ($K \geq 15$) candidate associations, for a linking length of $\log_{10} \Delta(J, \theta) = -1.7$: the upper panel of Fig. 4.6 shows the distribution of these associations in proper motion, velocity-distance and position space. Among these nine algorithmically-identified candidate associations, six turn out to be well-known open clusters: M67, Praesepe, the Pleiades, NGC 1662, NGC 1647 and NGC 2281, labelled in the top left panel. Most of the clusters we find within this linking length, with a minimum of 15 stars per group are located at a distance between 100 and 500 pc, while M67 is located at ~ 950 pc, and NGC 1647 and NGC 2281 are located at ~ 600 pc. Their distribution in action-angle space is illustrated in the lower panel of this figure, with the local standard of rest in these coordinates at $J_R, J_z = 0$ and $J_\phi \sim 1760$ (or $J_\phi = 1$ in the figure). Most of the groups show a more confined structure in action-angle space than in configuration space, presumably by construction through the condition $\log_{10} \Delta(J, \theta) < l_i$. Note that the "finite" extent of the known clusters in action space may well result from the individual distance errors, especially for the most distant group (M67): we did not assume that the line-of-sight extent of any association should not be much larger than the transverse, angular extent (see Fig. 4.7).

Remarkably, there are also three associations with $K \geq 15$ that are just as tight in action-angle space, but widely spread in proper motion or sky position. Especially the association with black points spreads hundreds of degrees in the sky; yet it is very confined in action-angle space. The extent, distances and radial velocities of these stars seem to reveal that this group is Pisces Eridanus: the newly discovered stellar stream in Gaia DR2, which could be the remnant core of a tidally disrupted cluster or OB association (Meingast, Alves, & Fürnkranz, 2019). As a reference, in Table 4.1 we present the ages and metallicities for some of the groups we find in Fig. 4.6. These clusters have solar metallicity and are mostly young, except for M67. The age of the newly discovered Pisces Eridanus is still under debate. While Meingast, Alves, & Fürnkranz (2019) claim that the age of this cluster is ~ 1 Gyr, Curtis et al. (2019) find this structure to be only 120 Myr using TESS data.

Cluster	Age (Gyr)	[Fe/H] (dex)	Ref.
M67	$3.5 - 4.8$	0.03 ± 0.01	*a,b*
Praesepe	0.65 ± 0.70	0.12 ± 0.04	*c*
Pleiades	0.013 ± 0.005	0.03 ± 0.05	*c*
NGC 1662	0.42	-0.09	*d,e*
NGC 2281	0.609 ± 0.013	0.13 ± 0.11	*f*
Pisces Eridanus	$0.12 - 1$	-0.04 ± 0.15	*g*

TABLE 4.1: Ages and metallicities of clusters in Fig. 4.6.
a,b: Yadav et al. (2008) and Randich et al. (2006)
c: Gossage et al. (2018)
d,e: Twarog, Ashman, & Anthony-Twarog (1997) and Reddy, Giridhar, & Lambert (2015)
f: Kharchenko et al. (2013)
g: Meingast, Alves, & Fürnkranz (2019) and Curtis et al. (2019)

We have not been able to identify these associations with known groups or clusters: they may well be newly found associations. Note from the bottom set of panels in Fig. 4.6 that part of the association marked with red symbols may be closely associated to the Pleiades. This only makes the point that parsing star groups into distinct entities has its limitations.

We notice that Pisces Eridanus and the group with blue points are nearly split at (θ_R, θ_z) = 0, respectively. The angles in `galpy` with the Stäckel approximation are defined such as $\theta_R = 0$ at pericenter and increasing going towards apocenter and θ_z starts at zero at $z = 0$ increasing towards positive z_{max} (Bovy, 2015). Therefore, the group with blue points for example is currently crossing the disc.

Additionally, all of these groups have low vertical action ($J_z < 9$ kpc km/s, Fig. 4.6), and thus the harmonic oscillator approximation applies. In this regime, the frequencies are independent of the amplitudes. Consequently, our estimates for (J_z, θ_z) are not strongly affected by our choice of the Galactic gravitational potential (MWPotential2014).

As we will show below, for any linking length, the number of associations grows rapidly with decreasing membership K. And the set of resulting associations depends of course both qualitatively (is an association found) and quantitatively (e.g. how many pairs are linked to, say, the Pleiades) on the choice of linking length.

For the moment, we just note that our FoF approach with this GDR2⊗LMDR5 sample not only recovers algorithmically know clusters as 'action-angle associations of indistinguishable [Fe/H]' and finds new ones, but also finds dispersed clusters.

Clearly, extensive follow-up of these associations is warranted.

4.5.3 Statistics of the FoF-selected Associations

We now consider the basic statistics of the associations that our FoF approach identifies. If the true action-angle distribution indeed has a clustered component, while the smooth mock catalog has not, we can expect that the contrast between the real and mock data is larger for groups than for pairs alone: if 'chance-pairs', drawn from a smooth orbit distribution at a given [Fe/H] are an important contaminant, then 'chance-triplets', etc. should be less so.

We quantify the statistics by asking what fraction of all pairs are involved in association of ultimate size K, at a given linking length l_i; this is shown in Fig. 4.8. This figure shows that for all linking lengths associations of at least 10 members are found; for $l_i = -1.6$ even 100 of them. The figure also shows that at very small linking length (e.g. $l_i = -1.8$) even the well-known clusters are not completely identified (see Fig. 4.6), presumably because measurement errors push pair separations beyond this linking length. Most dramatic in this figure, is the large difference between the fraction of real (solid line) and mock (dashed line) pairs that are in associations of $K \geq 3$: for $l_i < -1.6$ there is a magnitude or more associations in the real data than in the smooth mock data. This shows quite dramatically the clustering of stars with the same [Fe/H] in action angle space, not just pairs but clearly larger ensembles or associations. This is seen more clearly in Fig. 4.9 where we show the number of FoF groups as a function of $N_{members}$ at different linking lengths. We find that the number of groups at a given $N_{members}$ is always at least one order of magnitude larger in the real data, compared to the mock data (where they are "spurious", by construction). For the largest linking length $l_i = -1.55$ the number of groups found in the mock become comparable to the real data. However, at that linking length, the largest group in the mock catalog has only $N_{members} = 38$, compared to 10^3 in the real data.

4.6 Comparison to star pairs in $(\vec{r}, \vec{v})$ configuration space

We have presented a method to calculate pairwise distances in action-angle space between MS stars in the GDR2⊗LMDR5 cross-match, and we have used it to quantify the level orbit-space and abundance space clustering of the stellar distribution in our Milky Way. We have found an excess of pairs—i.e. *clumping* of stars—at small phase-space distances $\Delta(J, \theta)$ and small abundance differences $\Delta[\text{Fe/H}]$ when compared to a mock catalog that has a smooth and phased-mixed orbital distribution (Fig. 4.3). In addition, we could show that extensive sets of star associations can be found by their orbit similarity, if they have indistinguishable [Fe/H]; we implemented the identification of these associations
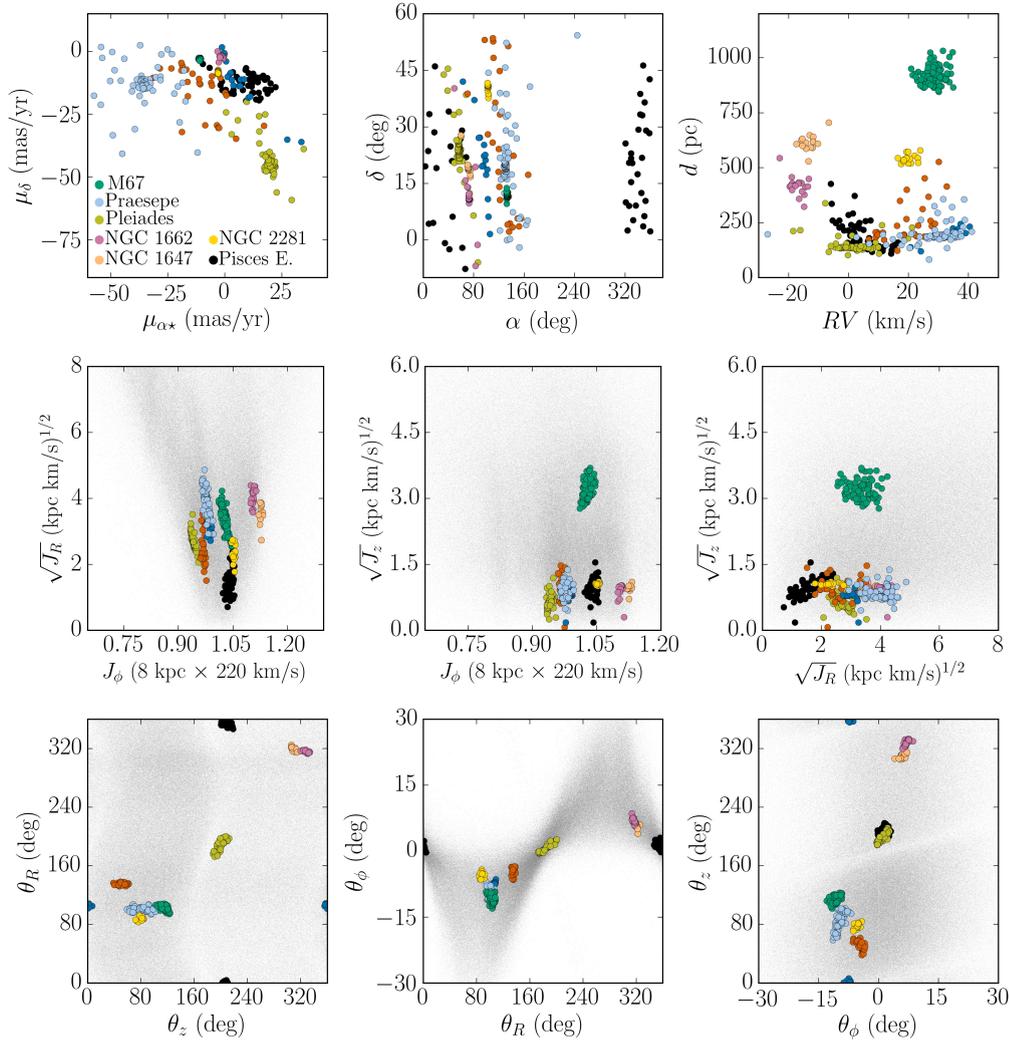
FIGURE 4.6: Result of our implementation of the friends-of-friends algorithm for a linking length of $\log_{10}\Delta(J,\theta) = -1.7$. We only show the largest groups we find for this specific linking length: 9 groups with a minimum of 15 members per group. *Upper panels*: We recover 7 known associations: the open cluster M67, Praesepe, the Pleiades, Pisces Eridanus, NGC 1662, NGC 1647 and NGC 2281. M67, NGC 2281, NGC 1662 and NGC 1647 (hidden behind NGC 1662) appear as concentrated clusters in μ_α and μ_δ while Praesepe and the Pleiades have a well defined center and then their structure extends further out. Most of the clusters appear as large extensions in the sky in *(ra,dec)* and some of the clusters extend up to several pc in distance. *Middle and lower panels:* Here we show the three actions J_z, J_R, J_ϕ and their three respective angles, θ_R, θ_ϕ and θ_z for all the groups we find in this specific linking length. The grey dots in the background correspond to the complete dataset. As expected by construction, all of the groups appear clustered in action and angle space. Most of the groups shown here are confined to $J_z < 3$ (kpc km/s), only M67 reaches up to $\sim$ 9 (kpc km/s) and extends up to 12 (kpc km/s). In the radial action J_R none of the found clusters extend beyond 25 (kpc km/s) and they are tightly constrained in J_ϕ. As a reference, a star near the solar position would be located at $J_\phi = 1$ in this figure. From all of these groups, the black cluster (i.e., Pisces Eridanus) is the most intriguing, being very constrained in action space, but having members completely spread in *ra*, separated by 240 deg in the sky. In angle space the associations are also very confined, with Pisces Eridanus having members located at $\theta_R \sim 0$ and $\theta_R \sim 360$ deg, showing the periodicity of the angles. The same behaviour is observed for the dark blue group but in θ_z.
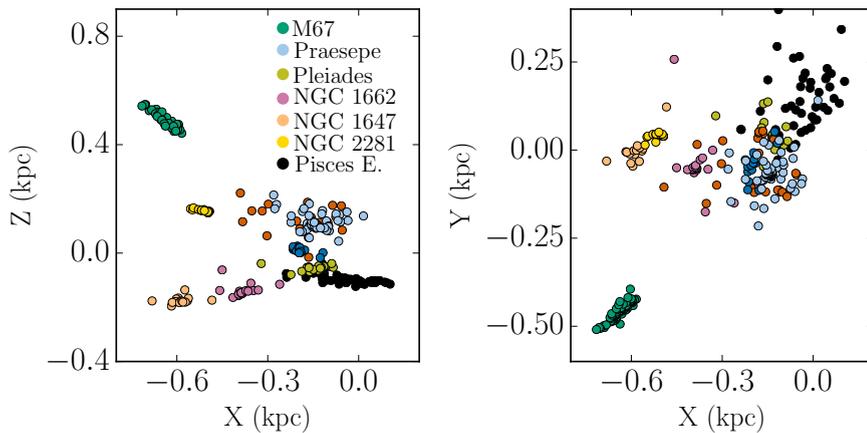
FIGURE 4.7: Position in rectangular Galactic coordinates X,Y,Z of the associations we find with the FoF algorithm. These are the same as the ones presented in Fig. 4.6. The Z coordinate is positive pointing towards the North Galactic pole, X increases in the direction of the Galactic center and the sun is located at (0,0,0). M67 and NGC 2281 appear mostly confined in the XZ and ZY plane, whereas Praesepe, the black, dark green and NGC 1662 associations have members spread in the XY plane.

by a friend-of-friends algorithm. We now provide and discuss some context for these findings. This is of course not the first time, that orbit-[Fe/H] clustering has been explored with contemporary data sets. The recent studies by Oh et al. (2017) and in particular by Kamdar et al. (2019) have shown that co-moving pairs, identified by their physical separation and velocity difference, were most likely born together, as these pairs showed a strong preference for having similar metallicities. Kamdar et al. (2019) defined a primary metric in metallicity difference $|\Delta[\text{Fe/H}]|$ to determine if a pair is co-natal, and they also include the velocity and position differences of these pairs Δr and Δv. However, the work by Kamdar et al. (2019) focused on pairs that are close in $(\vec{r}, \vec{v})$ with $2 < \Delta r < 20$ pc and with $\Delta v < 1.5$ km/s, and the present work – in part by choosing action-angle coordinates – extends to far greater distances, as we illustrate in Fig. 4.5. We take the same bins as presented in Fig. 4.4, i.e., bins of 0.2 in $\log_{10}\Delta(J, \theta)$, but now we map them into position and velocity space to illustrate to what "typical" distances in configuration space $(\vec{r}, \vec{v})$ a certain $\log_{10}\Delta(J, \theta)$ corresponds to. For example $\log_{10}\Delta(J, \theta) = -1.6$, corresponds to a mean $\Delta r \sim 150$ pc and $\Delta v \sim 3$ km/s, with many pairs encompassing considerably greater distances in configuration space. The selection by Kamdar et al. (2019) is shown as a grey rectangle in Fig. 4.5, and we see that it is closer to the properties of our wide-binary reference sample than even our smallest bin in $\log_{10}\Delta(J, \theta)$, or our smallest linking length.
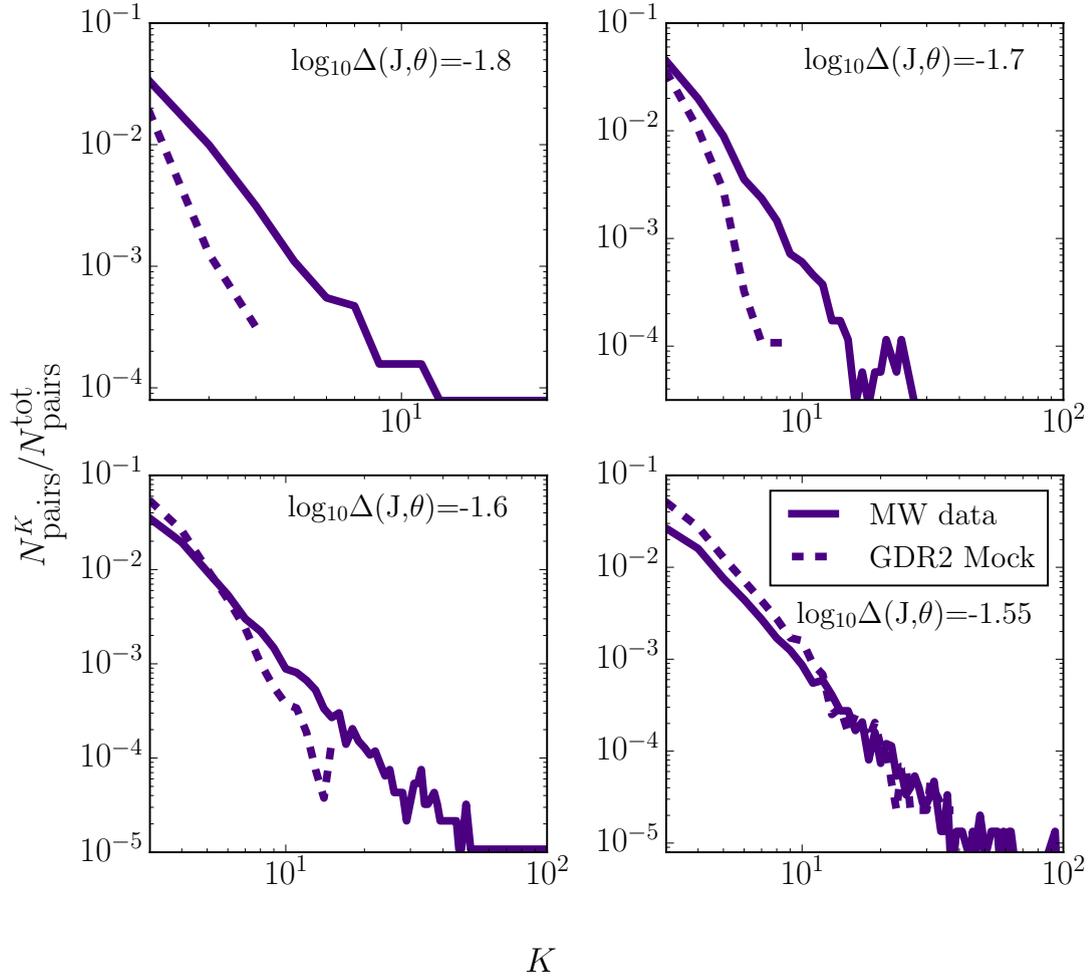
FIGURE 4.8: Fraction of pairs at a given linking length that end up in FoF groups vs. the number of members in that group, K. The solid line reflects this fraction for the GDR2⊗LMDR5 sample, and the dashed line to the GDR2 mock with a smooth orbit distribution. The different panels show different linking lengths, from the smallest on the upper left to the largest on the lower right. The differences between the data and the mock catalog are dramatic: associations with $K \geq 3$ are proportionally more common in the real data by an order of magnitude, except for the largest linking length. In the largest linking length bin (and any larger ones), the real data barely show more associations than a smooth distribution.
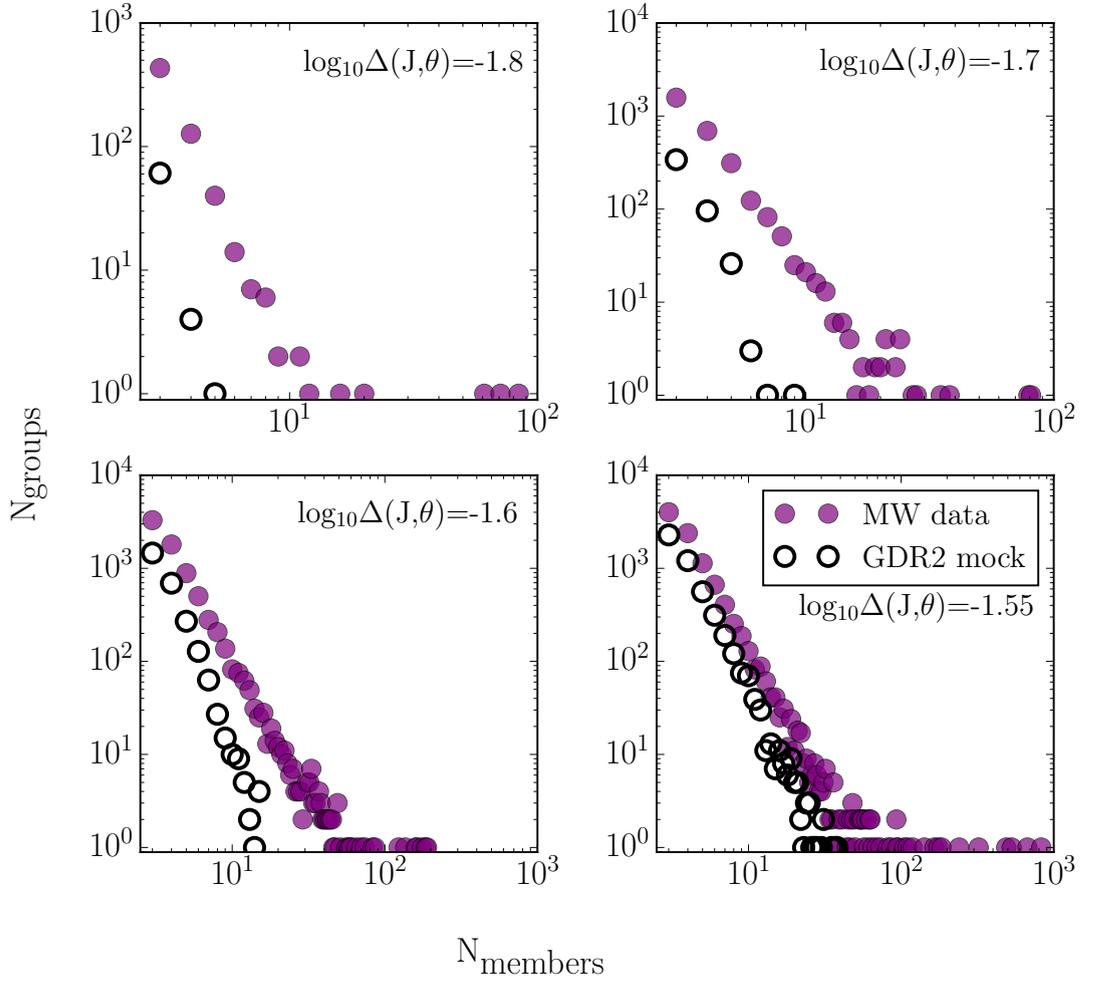
FIGURE 4.9: Number of FoF groups at a given $N_{members}$, for the GDR2⊗LMDR5 sample(solid circles) and GDR2 mock (open circles). The different panels show different linking lengths, from the smallest on the upper left to the largest on the lower right. For the shortest two linking lengths the number of FoF groups is over an order of magnitude larger, compared to GDR2 mock. At larger linking lengths these differences are less prominent. At $\log_{10} \Delta(J, \theta) = -1.55$, the number of groups that we find for the MW data and GDR2 mock become comparable, for groups with less than 40 members; yet the largest group in the real data has $\sim 10^3$ members, the largest mock group only 38.

4.7 Final remarks

In this Chapter we have explored and quantified the orbit-space clustering of stars in the Galactic disc, as a function of their metallicity differences. We have defined the orbit similarity between pairs of stars as the distance in action-angle space $\Delta(J,\theta)$, and the abundance similarity as Δ[Fe/H].

We determined the pairwise $\Delta(J,\theta)$ and Δ[Fe/H] for a sample of over half a million main sequence stars, with radial velocities and [Fe/H] from LAMOST and astrometric information from Gaia. Among these $\geq 10^{10}$, we found an excess of mono-abundance pairs(Δ[Fe/H] < 0.1), extending to large separations. We then use these action-angle distances as an input for a friends-of-friends (FoF) algorithm and we recover a number of known clusters: e.g. M67, Praesepe and the Pleaides. We also recovered the Pisces Eridanus stream, an association with a very extended distribution in configuration space: extending hundreds of parsecs, and covering 120 degrees in the sky.

5 | Exploration: Do the major action-angle groups appear in the GDR2–RVS sample?

In Chapter 4 we found hundreds of mono-abundance associations using a FoF algorithm, and we focused our attention on the 9 largest ones. However, this analysis was restricted to a sample of $\sim 600{,}000$ main sequence stars in the LAMOST DR5 and Gaia DR2 cross-match. In this Chapter, we broaden this investigation to the complete radial velocity sample (RVS) in Gaia DR2 (Katz et al., 2019) that is comprised of ~ 7.2 million stars with radial velocities (see Section 1.4.1.2 for details). This new sample is not restricted to MS stars, however we have no metallicity information. Here we start our analysis with 6 of the 9 largest groups identified in Chapter 4, but our search is open to other groups that might show up in the vicinity of their action-angle space.

The motivation for exploring GDR2–RVS is that a) this sample offers an independent confirmation of the action-angle groups found; b) it is a considerably larger sample ($\times 10$) and c) with this sample we have fewer restrictions in the CMD (T_{eff}) space, which can offer age constraints on the groups; where we will also have an all-sky coverage, as we are not limited to a spectroscopic survey (the LAMOST survey covers the northern hemisphere). However, with no metallicity information we can't apply a mono-abundance criteria.

In Chapter 3 we have also used the GDR2–RVS sample, however we were limited to stars within 1.5 kpc as we were using $1/\varpi$ as the distance. In that case, considering larger distances was not possible, since that would have introduced stars with larger uncertainties in the parallax.

5.1 The GDR2–RVS sample

Throughout this thesis, we have stressed the importance of having precise distances to calculate actions. This has limited the datasets we have been able to use, mainly because we needed to cross-match with spectroscopic surveys, such as LAMOST, to obtain spectro-photometric distances, and keep uncertainties at $\sim 6\%$ for single stars. Additionally, in Chapter 4 we used one of these surveys to obtain abundance information, in order to relate stellar metallicity ([Fe/H]) and orbits.

Now, in this Chapter we make use of the Bayesian distances that Schönrich, McMillan, & Eyer (2019) provide for the RVS sample. A summary of the cuts they impose is: $\varpi/\sigma_\varpi > 3$, $(G, G_{BP}, G_{RP}) > 0$, $v_{\mathrm{los\,err}} < 10$ km/s, n_{vis} (number of observation periods) > 5, reasonable measured radial velocities $v_{\mathrm{los}} < 5550$ km/s, d and $1/\varpi < 10$ kpc. We also add a selection criteria on RUWE ≤ 1.6 (see Section 1.4.1.2). This leaves us with a final sample composed of 6.2 million stars.

5.2 Gaia Mock DR2 to match the GDR2–RVS sample

We have already compared our results to a smooth and phase-mixed orbit distribution, by making use of Gaia Mock (Rybizki et al., 2018) in Chapter 4, Sec. 4.2.2. However, we now need to match GDR2 mock to the GDR2–RVS sample. For this, we query stars with $\sigma_\varpi/\varpi < 0.1$, $4<$ `phot_g_mean_mag` < 13 and $3500< T_{\mathrm{eff}} < 6900$, as suggested by Katz et al. (2019). We obtain ~ 6.8 million stars with this query. We invert the parallaxes that give us exact model distances in GDR2 mock (Rybizki et al., 2018)

5.3 Finding action-angle groups in GDR2–RVS

In the previous Chapter we have defined a metric in action-angle space that has allowed us to recover known associations and also find many mono-abundance groups. Since this is a slightly different and larger dataset, we first want to examine if we can recover some of the associations we found in Chapter 4, Sec. 4.5.2.

Because we know the location of these groups in action space, we proceed to select them in the three different action combination: J_ϕ–J_R, J_ϕ–J_z and J_R–J_z by drawing an ellipse with mean action and 2σ (standard deviation) encircling the group. We show an example of this selection in Fig. 5.1. This method will also select stars that are not necessarily associated to each particular group. We then proceed to plot each selected ellipse in
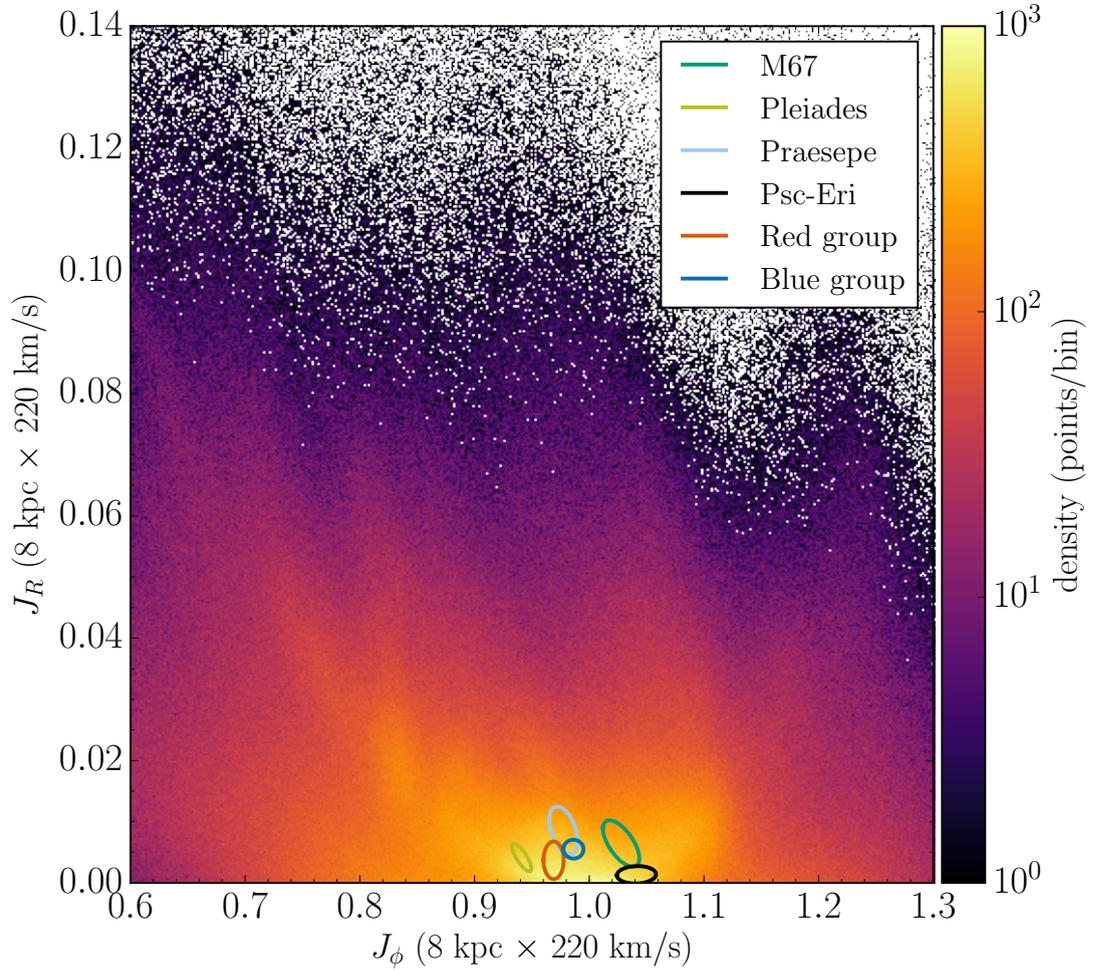
FIGURE 5.1: Action distribution in J_ϕ and J_R for the RVS sample with Schönrich, McMillan, & Eyer (2019) distances. The different colored ellipses show where the 9 groups identified in Chapter 4 are located. The sun's location is at $J_\phi = 1$ in this figure.

angle and position-velocity space. We find that for M67, Praesepe and the Pleiades, when selecting them in action-space only, we recover their locations in both angle-space and position-velocity.

Fig. 5.2 shows the angle distribution of the ellipse selection in J_ϕ–J_R, J_ϕ–J_z and J_R–J_z for M67. We see that the cluster shows up clearly in the three angle combination, θ_z–θ_R, θ_R–$\theta\phi$ and θ_ϕ–θz. Our ellipse selection in action space is not exact, therefore we are selecting more stars than just the ones belonging to the cluster. This is why we see that the angle distribution covers a wide range (especially in θ_z–θ_R). This figure also shows the comparison with a smooth and phase-mixed orbit distribution, where the mock catalog reproduces the general features seen in the data but without the cluster. In Fig. 5.3 we also show the position, proper motion and distance–velocity distribution
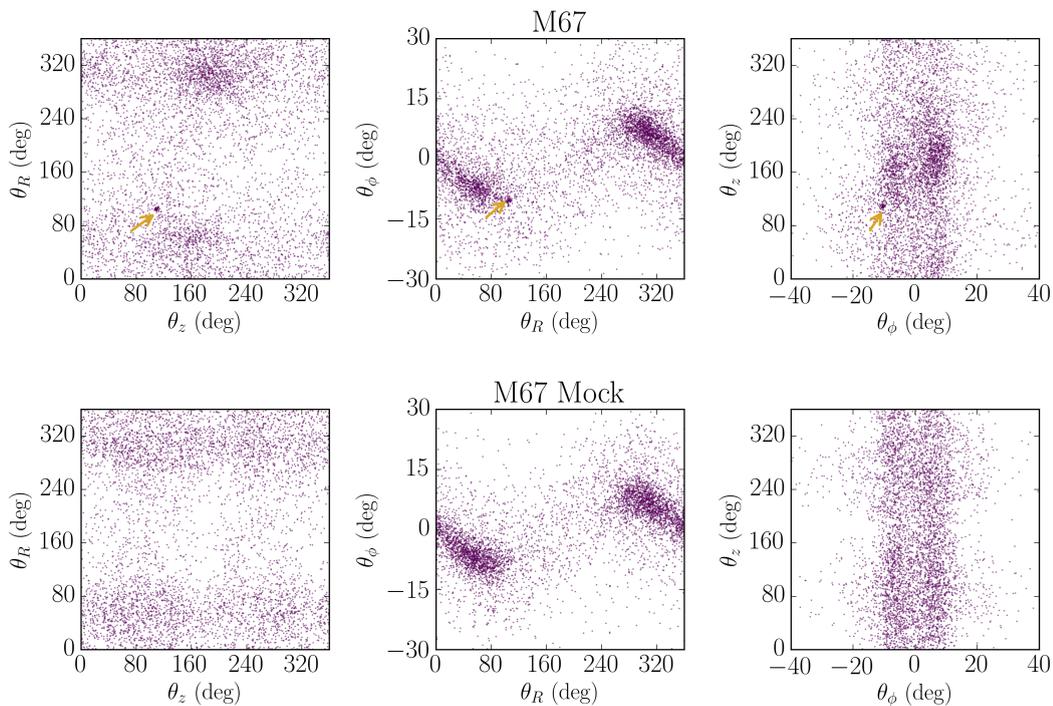
FIGURE 5.2: M67 distribution in angle space ($\theta_R, \theta z, \theta_\phi$). The top panel shows the real data and the bottom panel GDR2 mock. The approximate location of M67 is indicated with a yellow arrow. The mock catalog shows the same general features of the data, however the cluster is not visible in this smooth distribution in angle space, as expected.

of M67. We again notice that the cluster is clearly visible in the data, whereas in GDR2 mock we see exclusively the general features being very well reproduced, for example, the arch like structure that shows up in the velocity–distance plot.

We observe similar trends for the Pleiades (Figs. 5.4 and 5.5) and Praesepe (Figs. 5.6 and 5.7). Both associations clearly show up in angle, position, distance and velocity space as very confined clusters, but only in the data. Once again, GDR2 mock reproduces the general structure of the data, while showing none of these clusters.

In the vicinity of Praesepe, however, we notice a second cluster, that we highlight with a red ellipse in the upper panels of Fig. 5.6. This additional feature is clearly visible in θ_z–θ_R and θ_ϕ–θ_z. In θ_R–θ_ϕ (central panel), the cluster overlaps with Praesepe. This second cluster is also clearly present when looking at the positions, distances and velocities of stars (Fig. 5.7), again highlighted with a red ellipse. In proper motion, this cluster is very extended, and therefore we show it in red points for clarity.

Notice that the proper motions from GDR2 mock do not seem to show a particular overdensity at the location of this cluster, implying that it is likely a physically related stellar association. Because of this group's proper motion, position, distance (~ 46 pc)
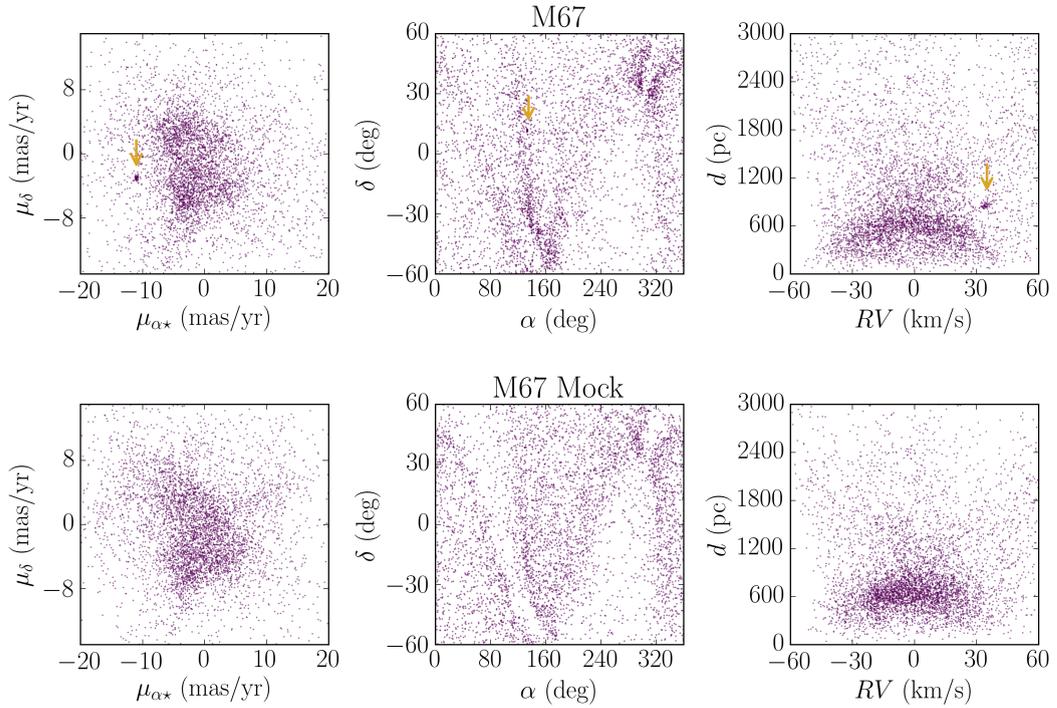
FIGURE 5.3: M67 distribution in position (ra, dec), proper motion and distance–velocity space. Top and bottom panels show real and mock data respectively. Again, the approximate location of the cluster is indicated by the yellow arrow. We see that the general distribution in position, distance and velocity space in the data is well reproduced by the mock catalog, e.g., the same arch in velocity–distance is also present.

and velocity (~ 37 km/s) we are able to identify it as the Hyades. We continue analysing this group in the following section.

5.3.1 A closer look to the Hyades

Since we did not find the Hyades cluster in Chapter 4 with our FoF algorithm, in this section we are able to take a closer look at it.

The Hyades and Praesepe are part of the Hyades supercluster (Brandt & Huang, 2015). It has long been thought that these clusters are coeval, having formed from a single molecular cloud, or cloud complex (Boesgaard, Roper, & Lum, 2013). Both clusters seem to have similar ages, as shown by isochrone fitting (Perryman et al., 1998; Salaris, Weiss, & Percival, 2004) and gyrochronology (Douglas et al., 2014; Kovács et al., 2014). Moreover, their metallicities are very similar at around [Fe/H]$\sim +0.13$ (Boesgaard, Roper, & Lum, 2013). It is also long known that their kinematic properties are similar (Eggen, 1959; Eggen, 1960). By looking into the action-angle coordinates (Fig. 5.6), we find that Praesepe and the Hyades move in fact on similar orbits.
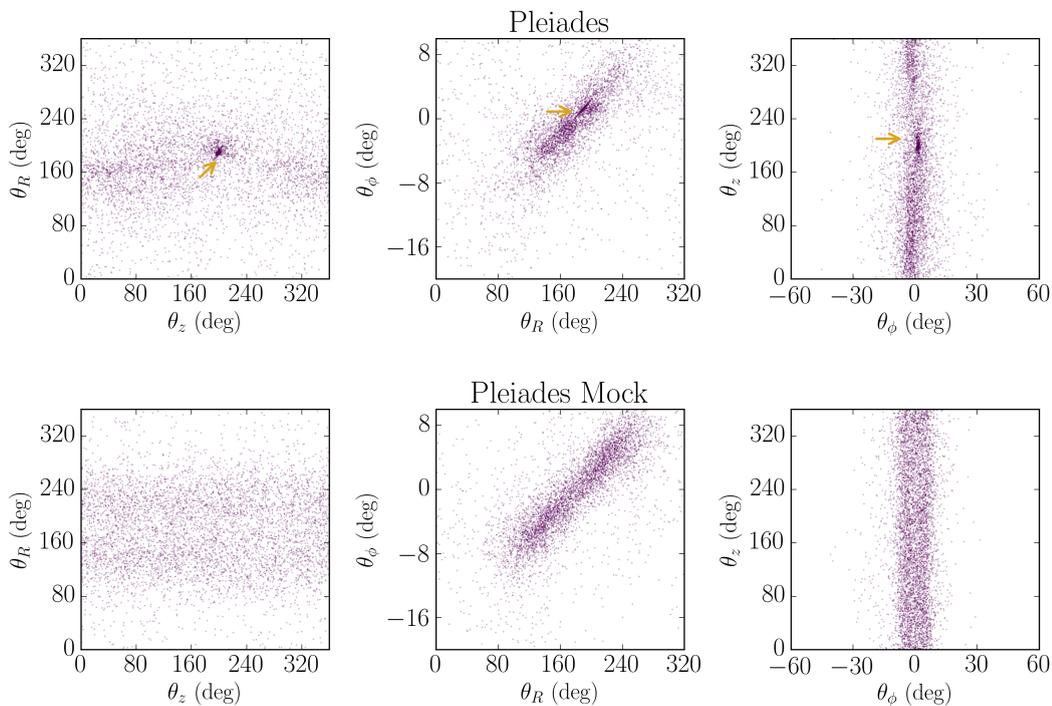
FIGURE 5.4: Pleiades distribution in angle space ($\theta_R, \theta_z, \theta_\phi$). The top panel shows the real data and the bottom panel GDR2 mock. The approximate location of the Pleiades is indicated with a yellow arrow. The mock catalog shows the same overall features of the data, however as also seen with M67, the cluster is not visible in this smooth distribution in angle space, as expected.

We identify the Hyades members by simply encircling the prominent cluster seen in the three angles ($\theta_R, \theta_z, \theta_\phi$), using appropriate ellipses, as shown in Fig. 5.6. We find 192 stars applying this selection criteria.

In Fig. 5.8 we present the sky position (ra, dec), proper motion and velocity–distance of the selected stars. By looking at the latter, we find that the majority of stars are concentrated at ~ 50 pc, which is consistent with the measure cluster distance of 46.75 $\pm$ 0.46 pc (Gaia Collaboration et al., 2017).

Although we find a very confined center, the cluster seems to spread widely in proper motion and position space. For instance, we find members extending up to ~ 100–200 pc in distance. In radial velocity the cluster also appears very extended, ranging from -20 km/s to ~ 45 km/s, with the center located at approximately 40 km/s which is consistent to observed members of this cluster (Debernardi et al., 2000). In Fig. 5.9 we show the all-sky distribution of the Hyades in galactic and equatorial coordinates. We see that the entire structure extends more than 100 degrees in the sky.

Meingast & Alves (2019) and Röser, Schilbach, & Goldman (2019) have recently reported the discovery of two tidal tails emerging from the Hyades cluster using Gaia DR2 data.
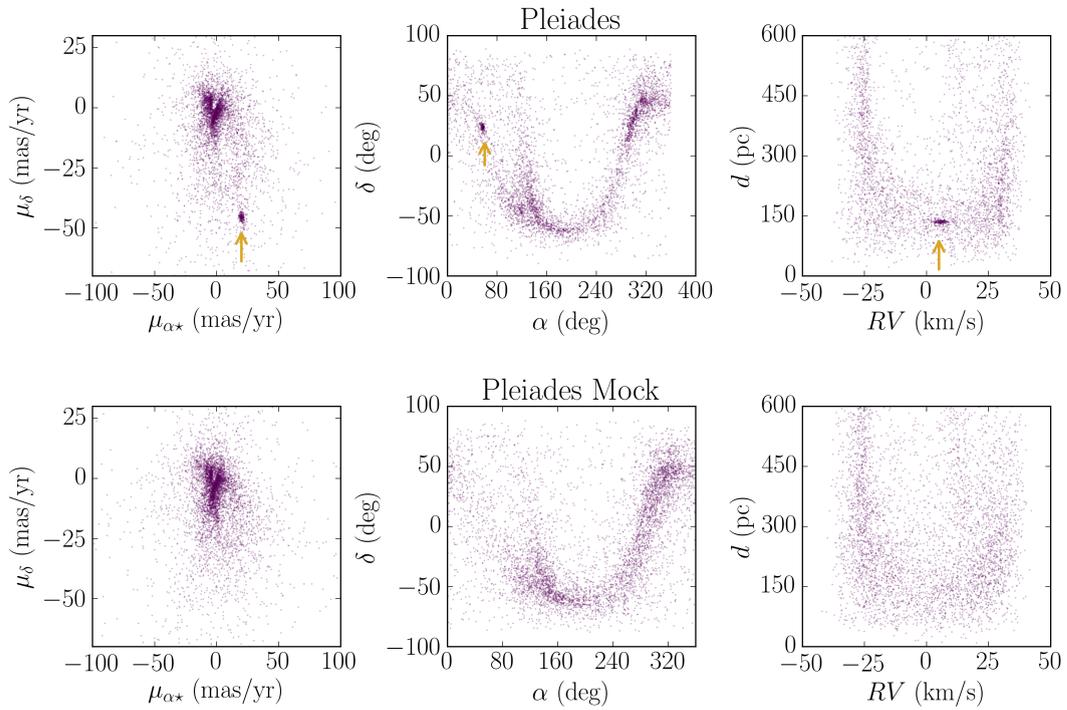
FIGURE 5.5: Pleiades distribution in position (ra, dec), proper motion and distance–velocity space. Top and bottom panels show real and mock data respectively. The cluster pops up clearly in the data, where its approximate location is indicated by the yellow arrow. We see that the overall distribution in position, distance and velocity space in the data is well reproduced by the mock catalog.

Both works search candidates in 3D velocity space within a 200 pc sphere around the Sun. In the case of Meingast & Alves (2019) they move from a cartesian to a cylindrical coordinate system (r, ϕ, z) to search for cluster members. They then apply a spatial density filter eliminating sources with fewer than 3 neighbours within 20 pc. They find 238 potential members for the Hyades, of which an important fraction is part of the tails. On the other hand, Röser, Schilbach, & Goldman (2019) find candidates in a cartesian coordinate system (X,Y,Z), and afterwards they make some cuts in (X,Z) to constrain the volume around the Hyades when searching for tidal tails. They find a leading tail extending up to 170 pc from the cluster center containing 292 stars (with 36 contaminants), and a trailing tail up to 70 pc with 237 members (with 32 contaminants).

In order to further explore if we are also seeing the presence of tidal tails, we compare our dataset to Meingast & Alves (2019). In Fig. 5.10 we show the cartesian coordinates XY of the Hyades members as found by us (magenta circles) and by Meingast & Alves (2019) (grey stars). Here we notice that both datasets cover the same range in XY. Furthermore, we find that the stars in the cluster seem to follow a recognizable S pattern, with a central cluster core and curved extended tails, similar to what Meingast & Alves (2019) find. This picture is also consistent with predictions from numerical simulations.
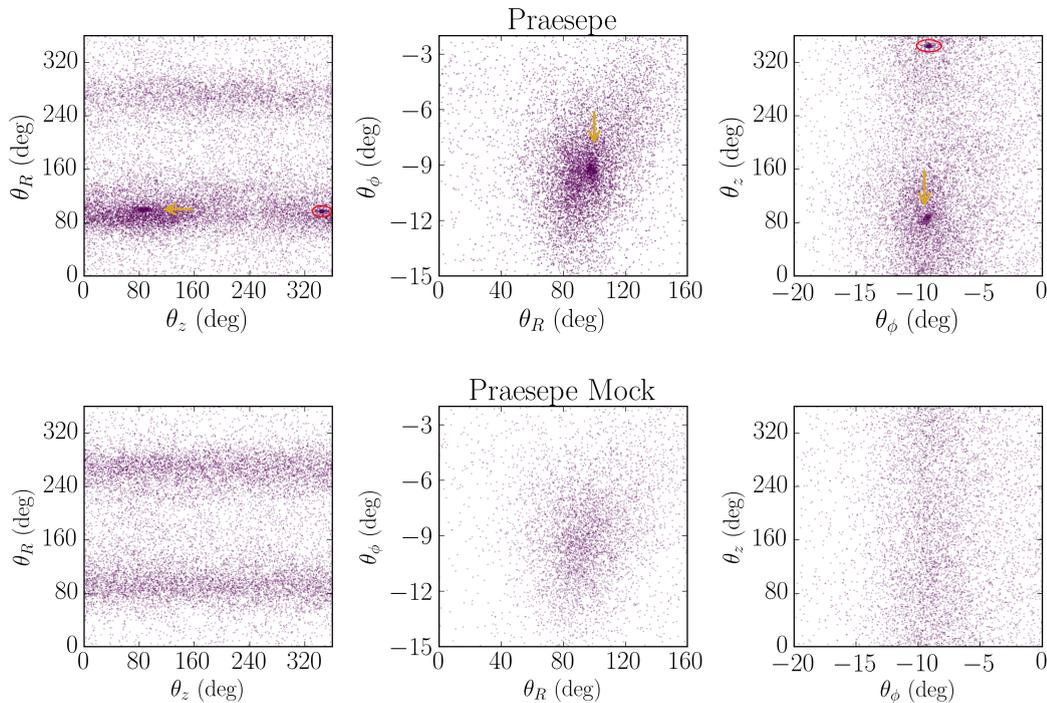
FIGURE 5.6: Praesepe distribution in angle space ($\theta_R, \theta z, \theta_\phi$). The top panel shows the real data and the bottom panel GDR2 mock. The approximate location of Praesepe is indicated with a yellow arrow. Here we notice that there is a second cluster that is highlighted with a red ellipse, that we identify as the Hyades. In $\theta_R - \theta_\phi$ the Hyades is almost overlapping with Praesepe. In the mock data, neither of these clusters show up.

Chumak, Rastorguev, & Aarseth (2005) studied the dynamical evolution of the Hyades with a numerical algorithm that included tidal forces and the orbit of the cluster under a Miyamoto-Nagai gravitational potential. In these simulations they find a centrally concentrated region with two stellar tails elongated along the Y axis and slightly bent. This description matches what we observe in Fig. 5.10.

In this section we have presented a qualitative analysis of the Hyades cluster. We have found members of this cluster by first selecting stars in action space nearby Praesepe. We then plotted these stars in angle space and found that both Praesepe and the Hyades naturally show up as very distinctive clumps. Based on their proximity in action and angle space, we conclude that Praesepe and the Hyades are moving on similar orbits around the Galaxy, which could imply (as already long thought) that both clusters were born in the same cloud complex or molecular cloud. Furthermore, with our simple selection in angle space, we not only found the Hyades cluster with 192 members, but also the presence of its tidal tails. The tails come out naturally from the combined selection in action and angle space, without any need of further selection criteria or cuts to especially find the tails, as previous works. This highlights the power of the action-angle coordinate system to find these type of structures.
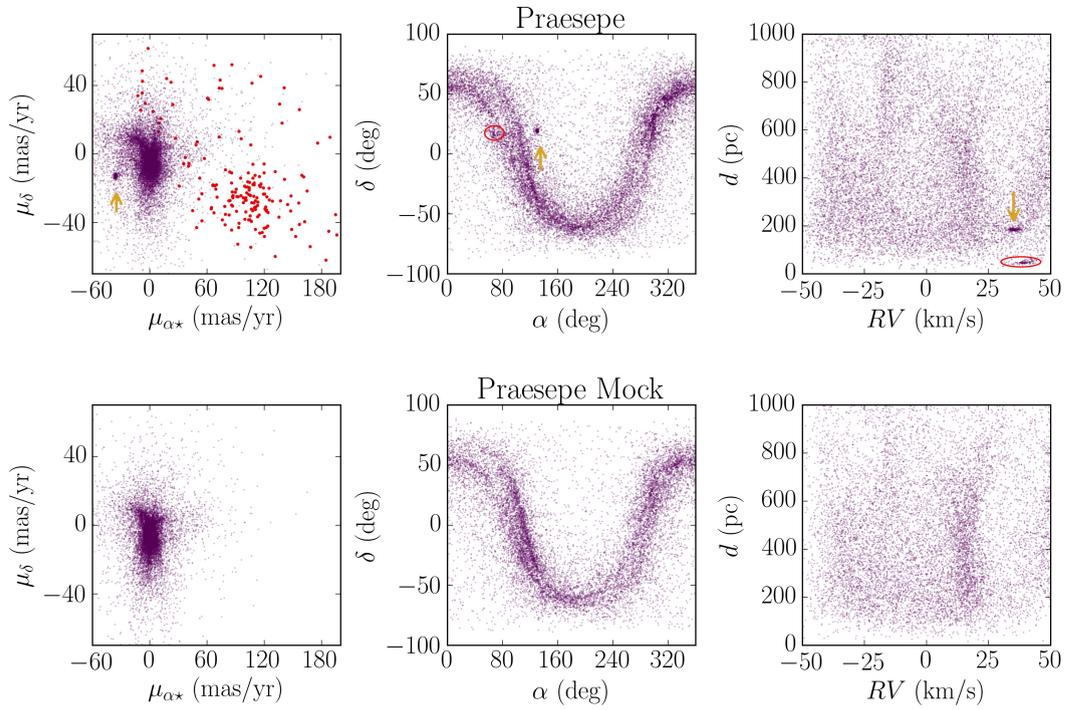
FIGURE 5.7: Praesepe distribution in position (ra, dec), proper motion and distance–velocity space. The second cluster (Hyades) also appears clearly in (ra, dec) and in velocity–distance (highlighted with a red ellipse). In proper motion, the Hyades is plotted with red points. We notice once more that GDR2 mock reproduces the general properties of the data, however none of the clusters show up.
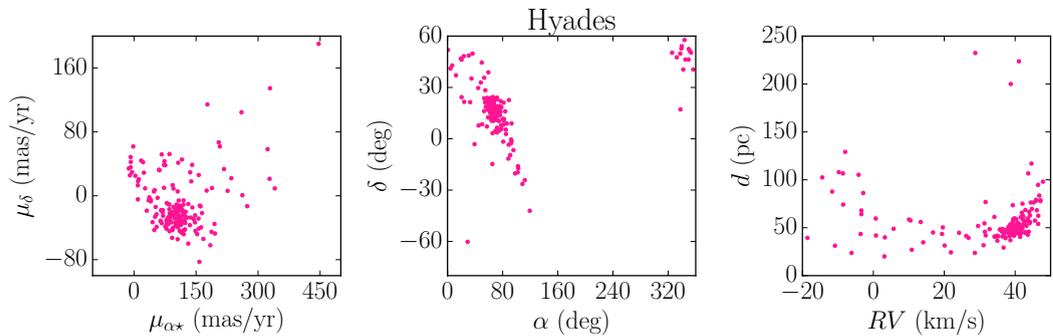

FIGURE 5.8: Hyades distribution in position (ra, dec), proper motion and distance–velocity space. We notice that the cluster has a concentrated center, and then extends broadly in distance, velocity and position.
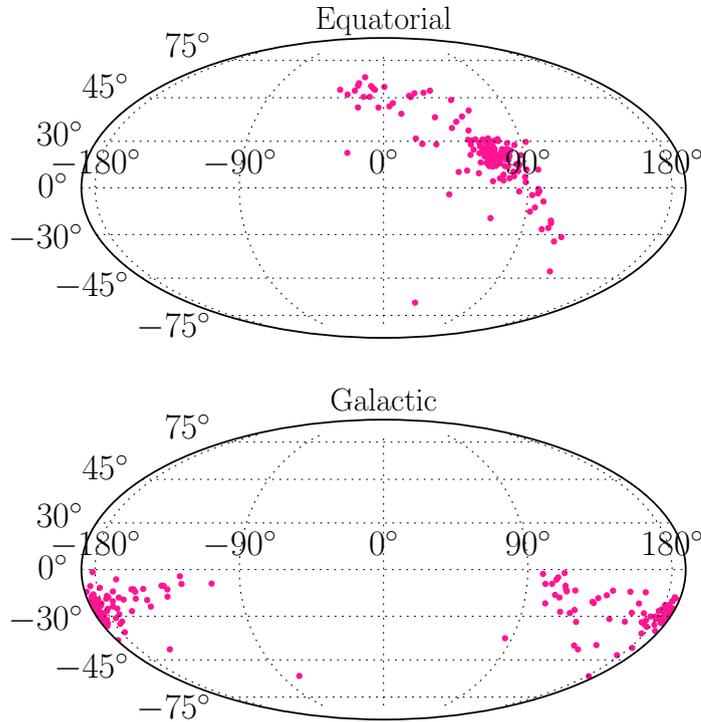
FIGURE 5.9: Equatorial (*top*) and Galactic (*bottom*) distribution of the Hyades. This figure shows more clearly the extent (>100 deg. in the sky) of the cluster.

5.3.2 Pisces Eridanus stream

In Chapter 4 we also recovered the Pisces Eridanus (Psc–Eri) stream. This association was recently found in Gaia DR2 by Meingast, Alves, & Fürnkranz (2019), with an estimated distance range of $\simeq$ 80–226 pc, and extending across $\sim$ 120 degrees in the sky. The stream was identified by searching overdensities in the 3D velocity space within 300 pc of the Sun. They move from a cartesian to a cylindrical coordinate system (r, ϕ, z), and apply a spatial density filter with a threshold of 7 neighbours within a radius of 30 pc. After this, they are left with 256 potential members.

We make a simple selection of possible members of Psc–Eri, as with the previous clusters. We start by identifying the approximate location of the stream in action space, as described in Sec. 5.3. The different angles of the selected stars are shown in Fig. 5.11. Here, we can easily observe that there are two prominent clusters in angle-space. The approximate location of Pisces Eridanus is indicated with a yellow arrow, while the second cluster is encircled with a red ellipse.
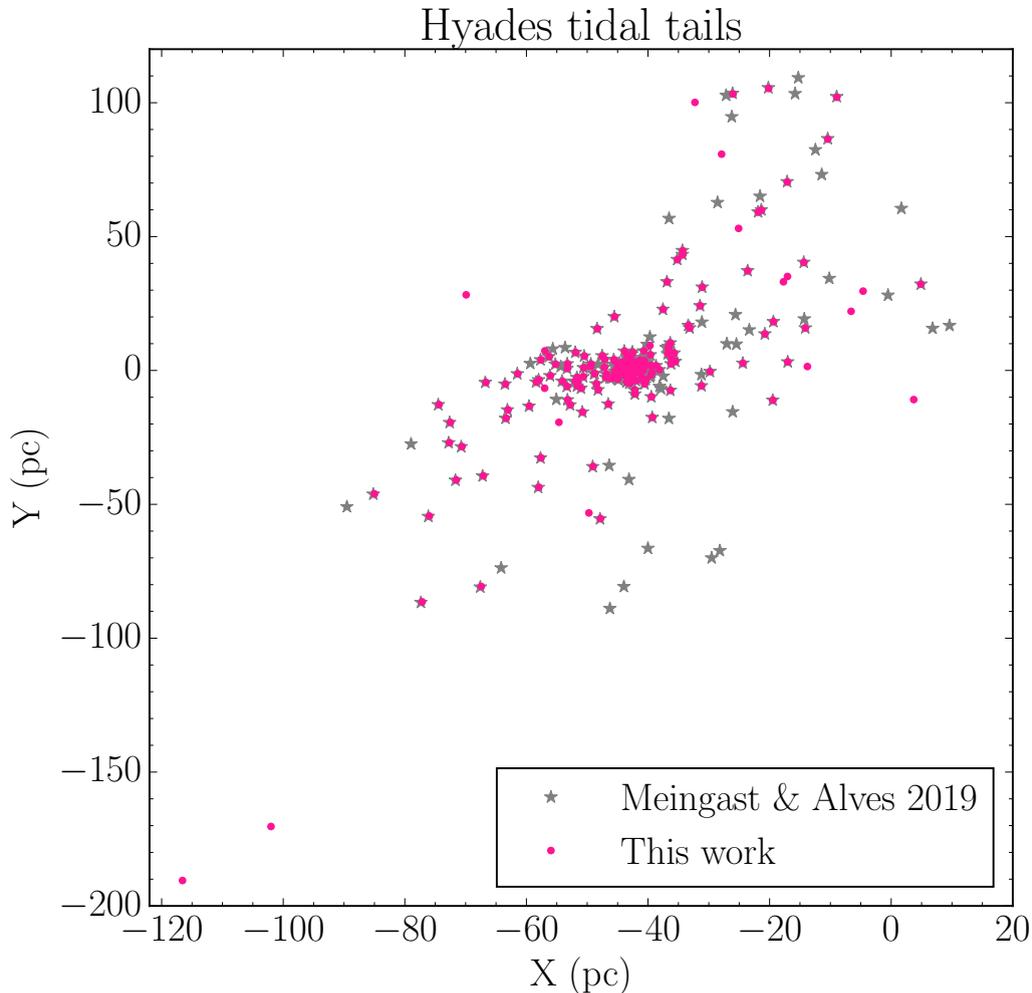
FIGURE 5.10: Hyades cluster members (pink dots) in cartesian XY coordinates. In grey points we overplot the data from Meingast & Alves (2019). We can see that the cluster has a central core, with the presence of tidal tails extending hundreds of pc.

We follow the same procedure to select possible Pisces Eridanus members as with the Hyades cluster. We simply encircle the prominent cluster seen in the three angles $(\theta_R, \theta_z, \theta_\phi$, Fig. 5.11). After applying this method, we are left with 376 stars. We proceed to do the same with the second cluster encircled in a red ellipse, and we are left with 45 stars in this cluster. We show the position, distance, velocity and proper motions of both associations in Fig. 5.12. The value of these parameters strongly suggests that the second cluster is Coma Berenices. This implies that Coma Berenices is moving on similar orbits as Psc–Eri. We find that the Pisces Eridanus stream appears as a very extended structure in the sky, covering hundreds of degrees, as we already noticed in Sec. 4.5.2 of Chapter 4.

Psc–Eri is a recently discovered stream, and it has been investigated only by a few studies, for example: Meingast, Alves, & Fürnkranz (2019), Curtis et al. (2019) and more

recently by Hawkins, Lucey, & Curtis (2020). Meingast, Alves, & Fürnkranz (2019) first discovered the stream, and besides a detailed kinematic study, they also estimate its age to be ~ 1 Gyr through isochrone fitting, using the PARSEC isochrone library (Bressan et al., 2012). However, there is no visible main sequence turn off in this stream, making the age identification a difficult task. If 1 Gyr is truly the age of this stream, it would make it a benchmark cluster, as the oldest coeval stellar population within 300 pc. However, Curtis et al. (2019) revisited this stream to determine its age through gyrochronology, using TESS photometry data. They measure $P_{\rm rot}$ for ~ 100 members, and find that they overlap with some of the Pleiades members. This might indicate that Psc–Eri is coeval with the Pleiades, and consequently young, with an age of ≈ 120 Myr. They also find members (34 candidates) to track the upper main sequence of the Pleiades and compare it to Psc–Eri, reinforcing their conclusions of a young age for this stream. They lastly discuss that in the Meingast, Alves, & Fürnkranz (2019) sample there is one star, 42 Ceti, which is an evolved multiple system (G8IV subgiant; Houk & Swift, 1999) that points to an older age for Psc–Eri, but they conclude that it could be an interloper. Finally, Hawkins, Lucey, & Curtis (2020) studied the detailed chemical nature of the stream, where they find that the stream has near solar metallicity, with [Fe/H] $= -0.03$ dex and a metallicity spread of 0.07 dex (0.04 when excluding outliers). They also find an abundance of Li indicating that Psc–Eri is ~ 120 Myr old, consistent with the gyrochronology result.

Here we have sufficient information to investigate the CMD for this stream, since we are not limited to MS stars, as in Chapter 4. Because coeval stars form well defined sequences in a CMD, then this is a good way of establishing if the stars we find in this group have in fact the same age.

In Fig. 5.13 we present the CMD of the Psc–Eri members that we find with our method. We overplot 3 PARSEC isochrones[1] of 1 Gyr, 500 and 100 Myr with [Fe/H] $= -0.05$ dex (which is the mean metallicity we find for this stream with LMDR5 data in Chapter 4). We notice that we also find 42 Ceti with our method, and two more evolved stars (HD 203382 and HD 206294). Meingast, Alves, & Fürnkranz (2019) also finds HD 203382, however their spatial density filter removes this star. Although our findings are consistent with an older age for this stream, here we can only put an upper limit of 1 Gyr on Psc–Eri. Further investigation of this stream is needed to solve the age tension. For example, finding stars to populate the upper main sequence is of paramount importance. However, this is currently not possible with the GDR2–RVS sample, as it is not complete at the bright end, hence we do not see the main sequence turn off of Psc–Eri. Additionally, investigating the detailed chemical abundances of the stream members, as Hawkins, Lucey, & Curtis (2020) explored, is also an important avenue to follow. Elements such as Li or Eu can put constrains on the age and formation of Psc–Eri.

---

[1]`http://stev.oapd.inaf.it/cgi-bin/cmd_3.3`
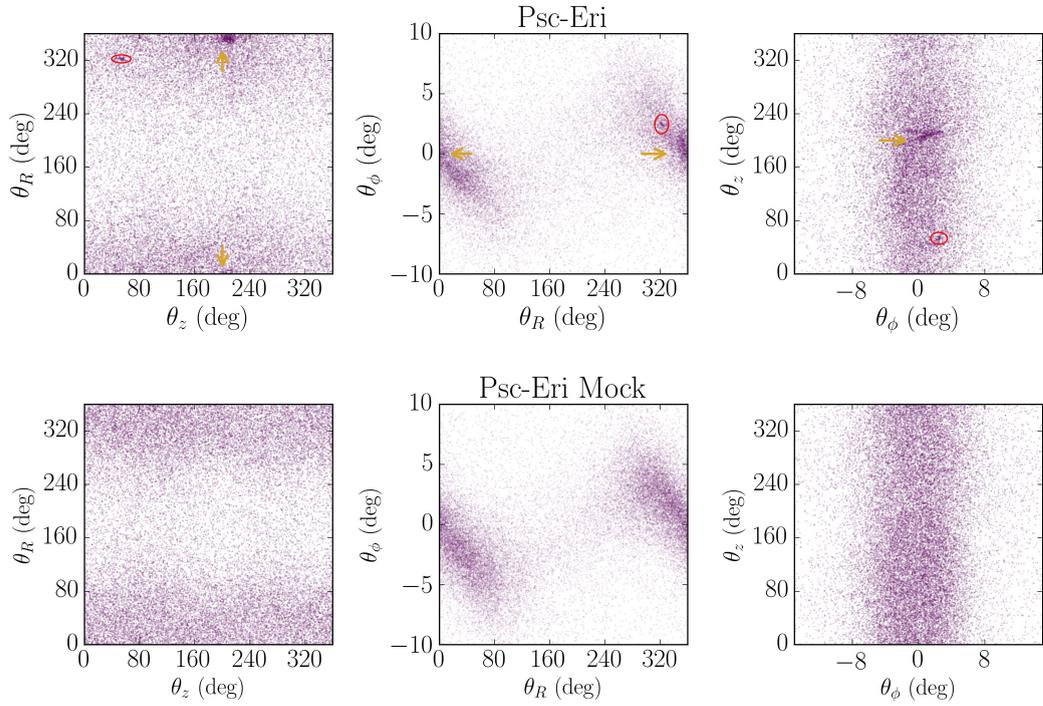
FIGURE 5.11: Pisces Eridanus distribution in angle space $(\theta_R, \theta z, \theta_\phi)$. The top panel shows the real data and the bottom panel GDR2 mock. The approximate location of Pisces Eridanus is indicated with a yellow arrow. Here we also notice that there is a second cluster (Coma Berenices) that we highlight with a red ellipse. The mock catalog reproduces the general features of the real data, with no localized clustering, as expected.

Notice that we are finding 120 more stars than Meingast, Alves, & Fürnkranz (2019), however this method is probably introducing some interlopers. Assessing the true membership of stars in this stream is very relevant, as the evolved 42 Ceti system clearly shows, given that its inclusion or removal makes a significant difference on the age of Pis–Eri. Recall that potential Psc–Eri members were identified by simply selecting stars around a clump in action-angle space. Even though this procedure could be improved (e.g., using a clustering algorithm), we were able to recover clusters and also very extended structures, which demonstrates the strength of using their location in a 6D coordinate system. Furthermore, we were able to find an additional group, Coma Berenices, due to its proximity to Psc–Eri in action space.

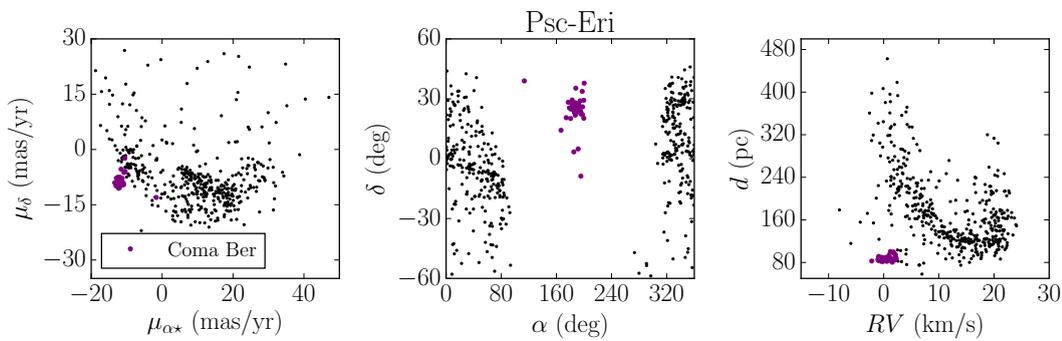FIGURE 5.12: Pisces Eridanus and Coma Berenices distribution in position (ra, dec), proper motion and distance–velocity space. The former is presented with black circles, while the latter is purple. Coma Berenices appears as a very confined cluster, whereas Pisces Eridanus is very extended structure in proper motion, but most dramatically in (ra, dec) covering more than 100 degrees in the sky.

5.3.3 Properties of the *unknown* action-angle groups in GDR2–RVS

In this section we focus on the remaining two groups shown in Fig. 5.1, and for simplicity we refer to them as Red and Blue. Similarly as with the previous groups, we now proceed to investigate their different properties.

Fig. 5.14 shows the angle distribution of the Red group. The clumping produced by the presence of the association is more clearly seen in θ_z–θ_R and θ_ϕ–θ_z (left and right panels, respectively). Following the same procedure as before, we select its possible members. We find 56 stars belonging to the Red group. In this case the task of finding members was a bit harder, as Fig. 5.14 shows, this group appears much more dispersed compared to Psc–Eri, for example.

In Fig. 5.16 we show the positions, distances and velocities of stars in this group. Even though it has a very extended structure, we can see some clustering in proper motion and (ra, dec). Here we find 28 more stars belonging to this group compared to the results presented in Chapter 4. With more stars, now we also investigate the CMD for this cluster, as shown in Fig. 5.17. We also plot three PARSEC isochrones of 1, 1.5, 2.5 Gyr with [Fe/H] = 0.05, which is the mean metallicity we found for this group using LMDR5 data (Chapter 4). The best fit seems to be for the 2.5 Gyr isochrone, mainly because we have one evolved star. However, the 1.5 Gyr isochrone also seems like a good fit if we ignore this star. With the information that we have available here, we can only give an upper limit to the age of this cluster of 2.5 Gyr.

Finally, we apply this same method for the Blue group. In Fig. 5.18 we show positions, distances and velocities of the stars selected from action space. We find 15 more stars than the original Blue group presented in Chapter 4, Sec. 4.5.2. Here we also observe
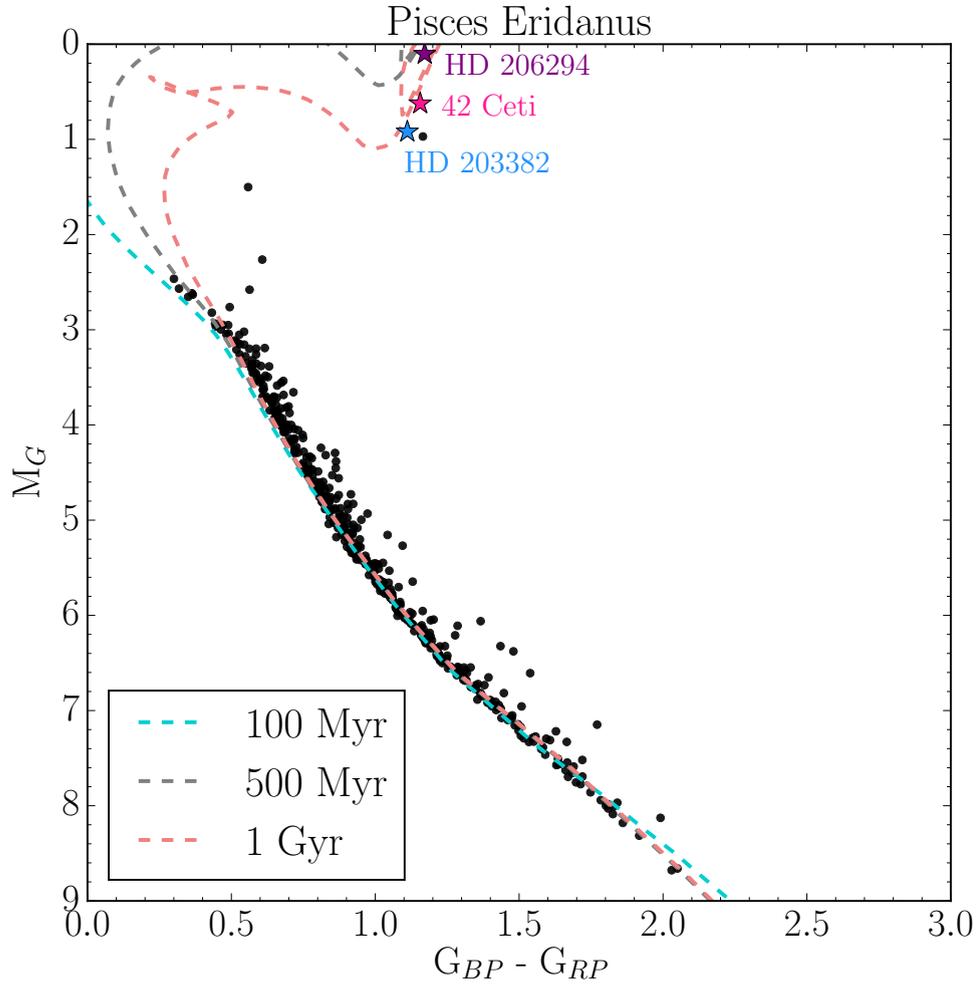
FIGURE 5.13: CMD for members of the Pisces Eridanus stream. The dashed lines are isochrones of 1Gyr, 500 Myr and 100 Myr, as indicated by the legend, all with [Fe/H] $= -0.05$ dex. We highlight 3 evolved stars that we find with our method including 42 Ceti, found also by Meingast, Alves, & Fürnkranz (2019) that points to an age of 1 Gyr for this stream.

that this group is mostly an extended structure. In Fig. 5.19 we show the CMD, along with 1, 1.5 and 2.5 Gyr isochrones with [Fe/H] $= 0.1$, which is the mean metallicity that we find for this group with LMDR5 data in Chapter 4. We do not have enough stars to say which of the three isochrones is the best fit, however we can give an upper limit of 2.5 Gyr to the age of this cluster.
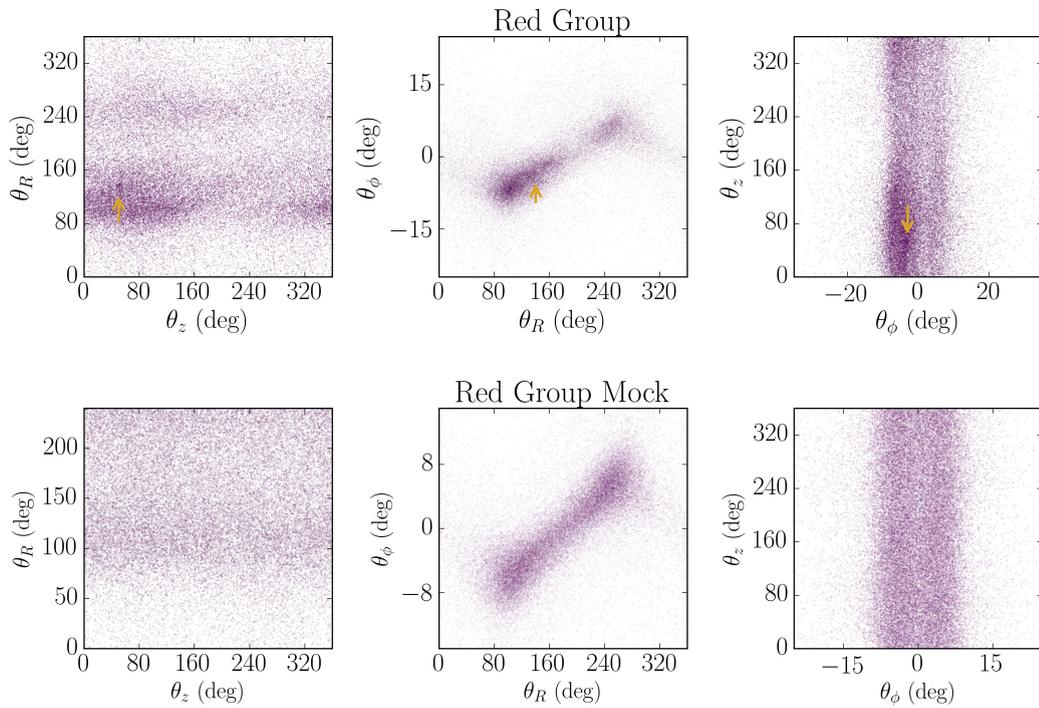
FIGURE 5.14: Red group distribution in angle space ($\theta_R, \theta_z, \theta_\phi$). The top panel shows the real data and the bottom panel GDR2 mock. The approximate location this group is indicated with a yellow arrow. This group is more clearly visible in $\theta_R - \theta_z$ and $\theta_\phi - \theta_z$. The mock catalog reproduces the overall features of the real data.
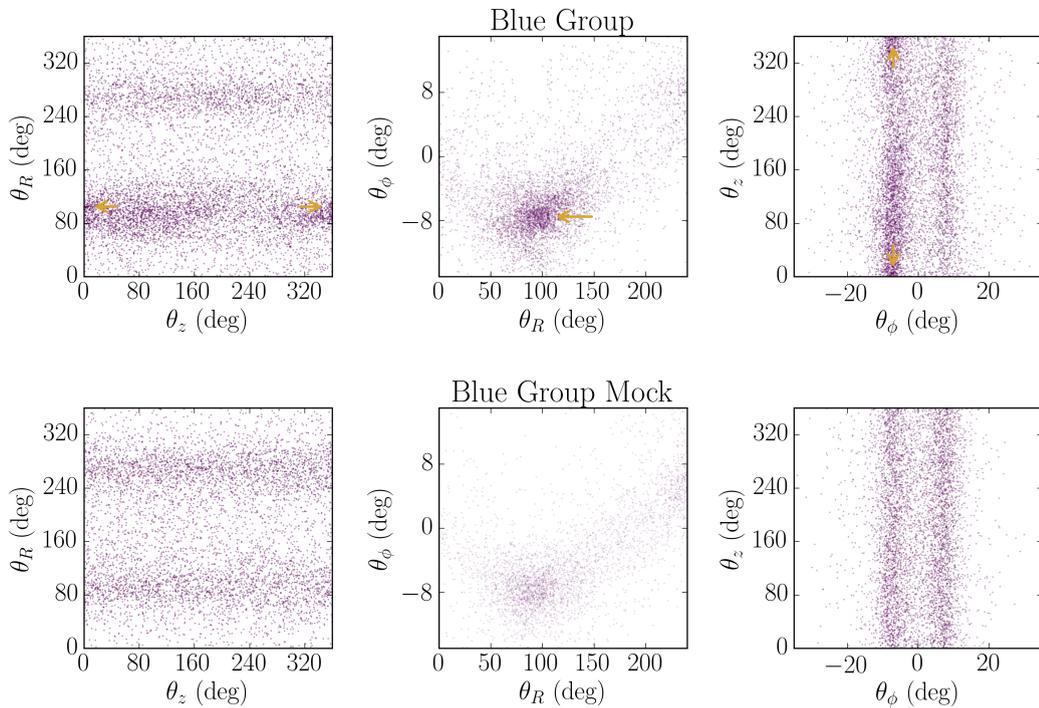
FIGURE 5.15: Blue group distribution in angle space ($\theta_R, \theta_z, \theta_\phi$). The top panel shows the real data and the bottom panel GDR2 mock. The approximate location this group is indicated with a yellow arrow. The mock catalog reproduces the overall features of the real data.
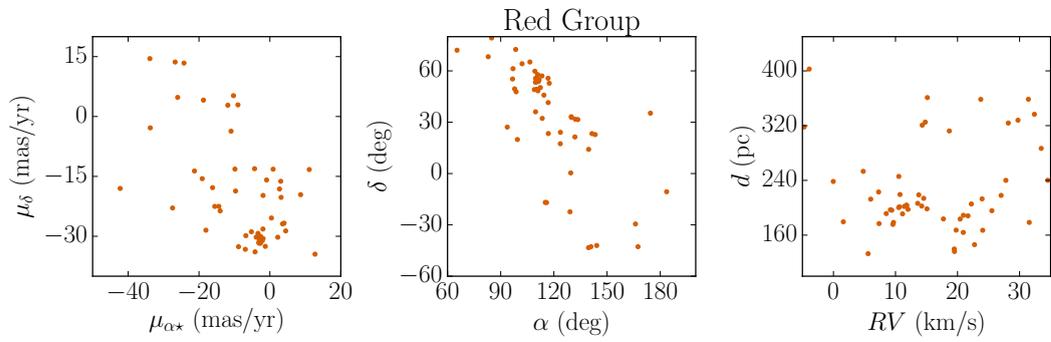
FIGURE 5.16: Distribution in position, velocity and distance for the Red group. In proper motion and (ra, dec) we see some clustering, but the remaining members are completely spread, especially in distance–velocity space.
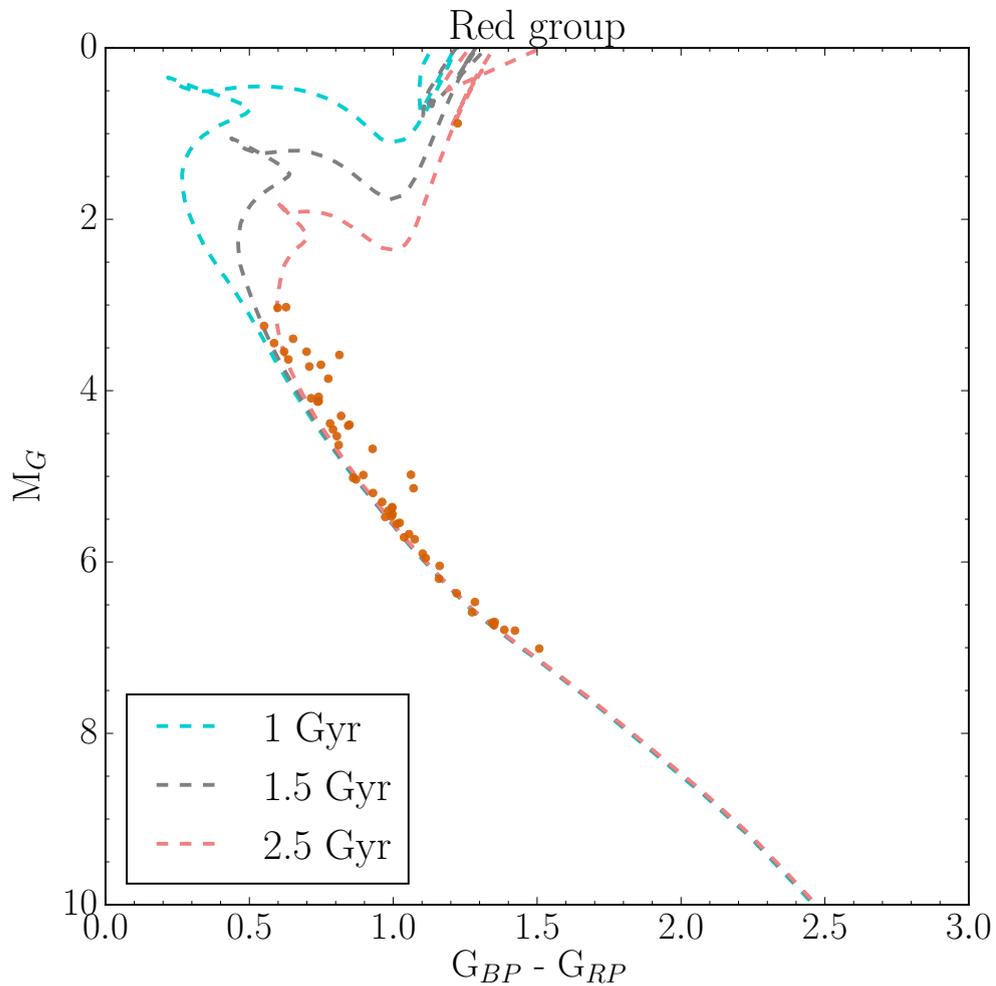
FIGURE 5.17: CMD for members of the Red group. The dashed lines are isochrones of 1, 1.5 and 2.5 Gyr with [Fe/H] $= 0.05$ dex. The 2.5 (but also the 1.5) Gyr isochrone seems like a good fit. We can only provide an upper limit of 2.5 Gyr for the age of this cluster.
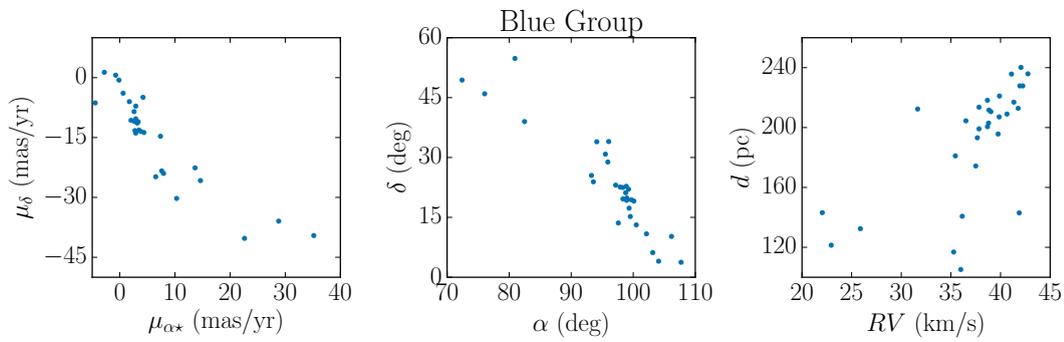
FIGURE 5.18: Distribution in position, velocity and distance for the Blue group. In (ra, dec) we see some clustering, but the remaining members are completely spread, especially in distance–velocity space
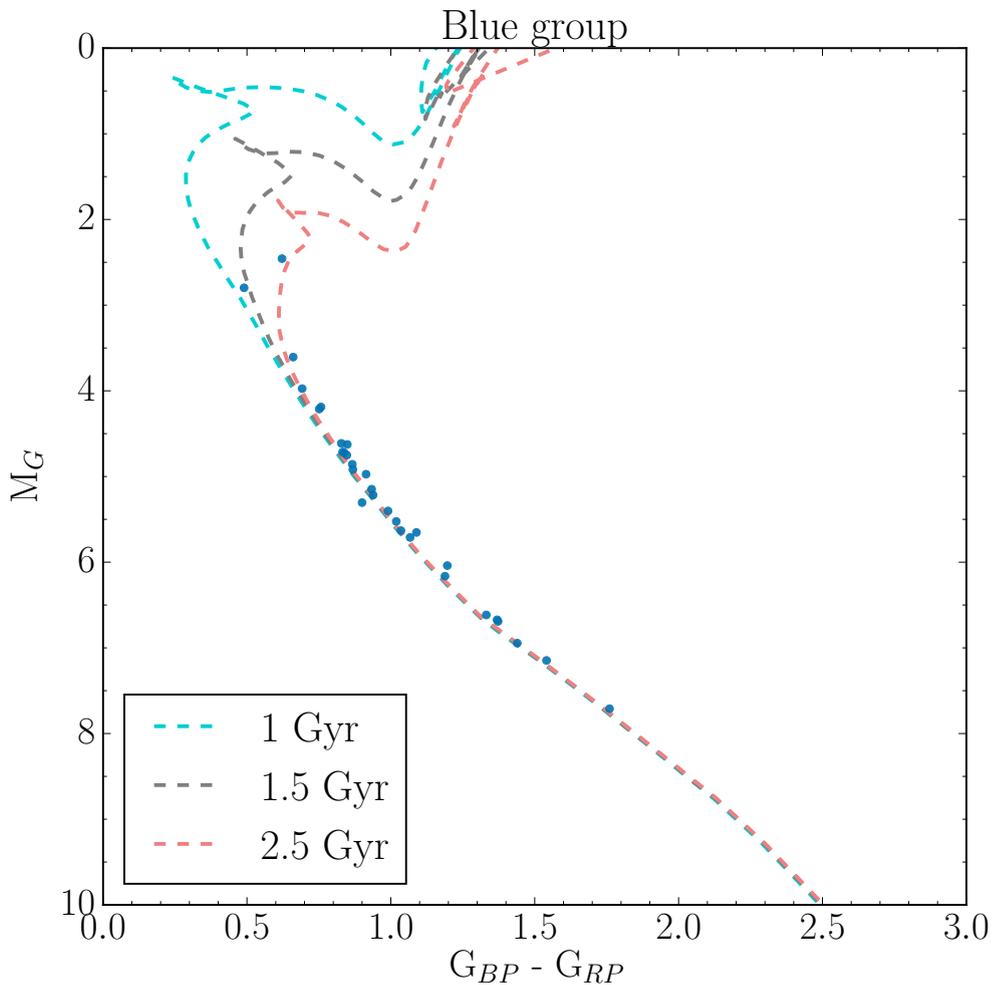
FIGURE 5.19: CMD for members of the Blue group. The dashed lines are isochrones of 1, 1.5 and 2.5 Gyr with [Fe/H] = 0.1 dex. The 2.5 (but also the 1.5) Gyr isochrone seems like a good fit. We can only provide an upper limit of 2.5 Gyr for the age of this cluster.

5.4   Final remarks

In this Chapter we have done an exploratory analysis in action and angle space of a significantly larger sample (6.2 million stars), not limited to MS stars as before. We have successfully recovered the 6 major groups found in Chapter 4, Sec. 4.5.2. This confirms not only the known clusters we found with an action-angle metric in the previous Chapter, but also the unknown associations (Blue and Red groups) and the Psc–Eri stream, with an independent dataset. However, this is just a first step as we have yet to exploit such large dataset. The prospects of this method become very relevant for associations such as the Red and Blue group. These groups appear as very extended associations in proper motion, position and distance–velocity, however they show up as confined clusters in action-angle space. In a different coordinate system they would not be selected as groups. This makes action-angles an ideal coordinate system to find such associations.

Additionally, we found the Hyades cluster and Coma Berenices in this new sample, both not recovered with the previous dataset. We found the Hyades when trying to identify Praesepe in action-angle space, thus uncovering that both clusters move on similar orbits. This is a confirmation of something that is already long thought: that both clusters were born in the same cloud complex or molecular cloud. We took one more step, and we investigated if we could see the presence of the Hyades tidal tails as a result of our selection in action-angle space. We identified a recognizable S pattern among the members of the Hyades, and we additionally compare our results to Meingast & Alves (2019). We have not applied any special selection or filter to find these tidal tails, they naturally appear when plotting the selection done in action-angle of the Hyades members.

Coma Berenices was identified in the vicinity of Psc–Eri in action-angle space. These clusters also move on similar orbits. The implications of this could be that both systems were born from the same molecular cloud. If this is the case, it would also put an age constraint on Psc–Eri, as Coma Berenices is a relatively young open cluster ($\sim$ 500 Myr, Melnikov & Eislöffel, 2012). In this Chapter, we have put an upper limit of 1 Gyr to Pisces Eridanus using isochrone fitting. This method, however, heavily relies on three evolved stars that could be interlopers. We need a more robust method than the one we have implemented in this Chapter to safely include or rule out these stars.

One of the caveats of this method is that it will probably introduce some interlopers in the groups. However as a first approach this method works, since we are clearly recovering associations with similar features as reported by previous studies.

Future work includes taking advantage of such a large and statistically more significant sample ($\times 10$ stars). By following the procedure shown in Chapter 4, we could calculate the distances between pairs of stars. We would then be in a position to run our FoF algorithm and link different members by $\Delta(J, \theta)$. This would allow us to find groups in a more robust way, reducing contamination. However, this method is significantly more computationally expensive, since it requires computing the distances of each star with respect to the entire sample.

6 | Summary and Outlook

6.1 Summary

In this thesis we investigated the small-scale structure of the MW's orbit distribution. We started in Chapter 2 by presenting a method to calculate spectrophotometric distances for main sequence stars in the Milky Way, building a model from parallax and spectroscopic information, combining GDR1 and LMDR5. One of the direct applications of precise distances is to improve the determination of stellar orbits in the Milky Way, which can be characterized by the actions (J_R, J_z, J_ϕ).

This model explicitly accounts for the parallax uncertainties and for the common binarity of near equal mass binaries among main sequence stars. Specifically, we constructed a model for the mean absolute magnitude in the K–band, $\overline{M_K}(\mathrm{T_{eff}}, \log g, \mathrm{[Fe/H]})$, from 4000 MS stars with small parallax errors ($\sigma_\varpi < 0.1$) in LAMOST $\times$ TGAS.

We then applied this model to stars with very imprecise, or even no parallax information, obtaining good spectrophotometric distances for 150 000 MS stars in this sample. This work explicitly takes into account possible binarity of stars, which has not been considered in analogous models. Ignoring binaries could result in biased distances.

We have built a model for the mean absolute magnitude of main sequence stars, which mostly draws on parallax information whenever parallaxes are very informative ($\delta\varpi < 10\%$); for increasingly poorer parallax estimates, this model gradually draws on the spectrophotometric information to estimate the distance modulus. We have shown that even in the regime of uninformative or missing parallaxes the model performs well, exploiting the information in the spectra: we obtain a value for the intrinsic dispersion in the absolute magnitude of single stars $\sigma = 0.12$ mag, which gives precisions in distance of $\sim 6\%$ for the fainter and more distant MS stars among current spectroscopic surveys. We show that if we compare the distance moduli for the 150 000 stars in LAMOST $\times$ Gaia sample to the expected parallax uncertainties at the end of the Gaia mission presented

in de Bruijne (2012) spectrophotometric distances are still needed. Especially for more luminous (and more distant) MS stars.

As an application of precise distances we showed that they greatly improve the precision of orbital action estimates, as distance uncertainties dominate the orbit uncertainties with proper motions from Gaia.

In Chapter 3 we investigated the orbital actions in the Galactic disc, using the GDR2–RVS sample that provides the full 6D phase space and orbit information for 7.2 million stars. As we require precise distances to calculate reliable actions, we restricted this large dataset to stars within 1.5 kpc, which left us with ~ 3.8 million stars. We found a large amount of coherent features present in the Milky Way disc, where we observe several overdensities or ridges at different locations in J_ϕ–J_R. Ridges observed in the solar vicinity ($1/\varpi < 200$ pc) coincide with the position of known moving groups. These observed features are not only present at close distances, but they extend up to 1.5 kpc distances.

Moreover, action space revealed more features than the usual phase-space (U, V), which shows the capability of actions as a very powerful coordinate system. These features have been most likely created by dynamical processes: non-axisymmetric perturbations (bar and spiral arms) and resonances in the Galactic disc.

We finally investigated the effect of measurement uncertainties when calculating actions. We showed that already at $1/\varpi > 600$ pc, some of the observed structure at that distance regime could be blurred out as a product of uncertainties in the parallax. This becomes important when exploring the orbit sub-structure to larger distances, beyond the solar neighborhood.

In Chapter 4 we explored and quantified the orbit-space clustering of stars in the Galactic disc, as a function of their metallicity differences. We have done this by defining the orbit similarity between pairs of stars as the normalised distance in action-angle space $\Delta(J, \theta)$, and their abundance similarity as $\Delta[\text{Fe/H}]$; we then considered both $p\left([\text{Fe/H}] \mid \Delta(J, \theta)\right)$ and $p\left(\Delta(J, \theta) \mid [\text{Fe/H}]\right)$. We expect the fraction of 'mono-abundance' pairs (with the same [Fe/H]) to be large for very small differences in actions *and* angles, $\Delta(J, \theta)$, as those stars are either wide binaries or stem from the same birth association. The fraction of mono-abundance pairs should then decrease towards larger $\Delta(J, \theta)$, as more of these star pairs on very different orbits were born at different times or at difference radii, and hence have different metallicities.

We determined the pairwise $\Delta(J, \theta)$ and $\Delta[\text{Fe/H}]$ for a sample of over half a million main sequence stars, with radial velocities and [Fe/H] from LAMOST and astrometric information from Gaia. Among these $\geq 10^{10}$, we found an excess of mono-abundance

pairs (Δ[Fe/H] < 0.1), extending to remarkably large separations. In configuration space ($\vec{r}, \vec{v}$) this $\Delta(J, \theta)$-selected excess of mono-abundance pairs extends to $\Delta r \sim 300$ pc; this is an order-of-magnitude larger than the 25 pc to which Kamdar et al. (2019) traced it with a configuration-space selection. We assess that this is a significant 'excess' through comparison with a mock sample, drawn from a smooth and phase-mixed orbit distribution with a similar selection function (Rybizki et al., 2018); in that smooth models such pairs just reflect chance similarities in action-angle space and in Δ[Fe/H] (given the modest metallicity dispersion of the disc).

We then use these action-angle distances as an input for a friends-of-friends (FoF) algorithm, to investigate which fraction of these mono-abundance pairs can be linked into larger groups (at a given linking length). Through this FoF approach, we recover a number of known clusters and associations: e.g. M67, Praesepe, the Pleiades, NGC 1662, NGC 1647 and NGC 2281. Whereas Praesepe and the Pleiades show a more extended structure in proper motion, position and distance-velocity space, the remaining known clusters are mostly confined in position and velocity space.

However, through this orbit-space FoF approach, we also find hundreds of mono-abundance associations with a very extended distribution in configuration space: extending hundreds of parsecs, and covering many degrees in the sky. For instance, we found the Pisces Eridanus stream which shows that our algorithm recovers not only clusters, but also these extended structures. Nevertheless these stars are on similar orbits and share the same chemical information, [Fe/H]. Many of these would not have been selected as associations in a different coordinate system.

Our analysis shows that the orbit distribution of Galactic disc stars reveals distinct small-scale clustering, among stars with indistinguishable metallicities, extending across distances of hundreds of parsec. At least qualitatively, this clustering has an obvious explanation: stars born in the same cluster, association, or even spiral-arm piece, will be born with the same [Fe/H]. Most of these birth associations will gradually disperse, as many of them may never have been gravitationally bound systems. This dispersal is driven both by orbit or action changes, which can be driven by cluster dynamics or radial migration (Sellwood & Binney, 2002; Frankel et al., 2018; Kamdar et al., 2019), and by the resulting orbit-phase mixing. All these effects plausibly reflect the transition from clustered star-formation to field stars.

In Chapter 5 we investigated if we could recover the major action-angle groups we found in Chapter 4 using the GDR2–RVS sample. We made use of Bayesian distances from Schönrich, McMillan, & Eyer (2019) that were available for the RVS sample, and we were left with 6.2 million stars. This is not only a larger dataset than the one we have previously investigated ($\times$ 10 stars), but it is also not limited to MS stars. However,

there is no chemical information ([Fe/H]) available, as we do not cross-match it to any spectroscopic survey. We searched the largest groups in action space first, since their location in J_ϕ–J_R, J_ϕ–J_z and J_R–J_z is known from Chapter 4. Then, we investigated the corresponding angles, finding that the groups naturally clustered in these coordinates. The selection of each association was made by simply encircling the stars that seem to cluster in action-angle space. Since this method is not exact, we are probably introducing interlopers. Nevertheless, as a first approach, this method works because we recovered 6 of the 9 largest groups from Chapter 4: M67, the Pleiades, Praesepe, Pisces Eridanus, the Red and Blue groups. Additionally, we recovered the Hyades and Coma Berenices, which did not show up in the sample presented in Chapter 4. We found the Hyades in the action-angle space vicinity of Praesepe, showing that both move in similar orbits; these clusters were known to have comparable kinematic properties. We found a similar situation with Coma Berenices and Pisces Eridanus. The latter is a recently discovered stream, and thus there are just a few studies that have investigated it. Only comparisons of basic astrophysical characteristics of the stream with those of the Pleiades have been done so far (e.g., Röser & Schilbach, 2020).

The last chapter was exploratory, and the first step towards applying a more robust method, such as the one presented in Chapter 4. We have shown that we can successfully recover not only the known clusters we have found in Chapter 4, but also more extended associations: Pisces Eridanus, the Blue and Red groups. The two latter are unknown groups, and we were unable to identify them within known clusters, however we have managed to put an upper limit on their ages (2.5 Gyr).

6.2 Outlook

The results presented in this thesis suggest follow-up in various directions. We can now use this information to quantify how effective orbit migration is in the Galactic disc, as we know that there are several established phenomena that can cause a star to move to a region of the disc different from its birth radius. In Chapter 4 we have seen the effects of stellar birth associations dispersing, as we have found small-scale clustering of stars with indistinguishable metallicities extending to several hundreds of parsecs.

Furthermore, the method presented in Chapter 4 has provided a large number of stellar association candidates. While undoubtedly some will be spurious, our FoF analysis should open a path to studying many groups of stars barely remembering their common birth origin. We have presented the statistics of the groups we found: hundreds of associations (with more than 10 members) that we have yet to investigate. However, in our analysis we focused on a specific linking length ($\Delta(J, \theta) = -1.7$) and also on the groups with

the largest amount of members. Immediate future work involves investigating the great amount of remaining associations that are left, for example at $\Delta(J, \theta) = -1.6$, where the overabundance of such groups in the real data is over an order of magnitude, compared to the spurious associations in the mock data. The importance of this, is that we will not only find clusters, but also extended associations such as Pisces Eridanus, or the Red and Blue groups. Continuing to explore the results presented in Chapter 4 also offers the opportunity to go beyond the metallicity, [Fe/H]. The LAMOST data will allow us to explore whether these associations are truly mono-abundance populations (not just of the same [Fe/H]), by looking at the other 5-10 abundances that LAMOST provides (Xiang et al., 2019). The latter is also of interest to further investigate the detailed chemical information on associations such as the Pisces-Eridanus stream as a way to further constrain its age and formation.

In Chapter 5 we have already started some follow-up, searching in orbit-space for more members in the Gaia 6D data, but where metallicities are missing. We have taken a first step with this dataset, only investigating if we could recover the largest groups from Chapter 4. We have found those associations and recovered two more, showing that even with a qualitative method we are able to detect such groups, revealing the valuable information present in such a dataset. Then, a natural next step would be to apply our FoF algorithm to this data. We would first need to calculate the distance between pairs of stars, and then use our $\Delta(J, \theta)$ metric to link members in groups. This however, will require an important amount of computing time, as our implementation of our metric is done by computing the distances in $\Delta(J, \theta)$ of each star with respect to the rest of the sample, and we would need to do this for 10^{12} pairs. Even without chemical information ([Fe/H] or element abundances) we can establish a significant statistical sample of possible groups clustered in action-angle space. We could even complement these results by searching if members of the groups have been observed in different spectroscopic surveys (LAMOST, GALAH, APOGEE) to obtain their chemical information. Trying to follow-up the candidate clusters with a high-resolution spectrograph to observe all (or most) of the members of different large associations is inconvenient, as we would need several observing runs. But observing some of the members is feasible. Either complementing these results with a spectroscopic survey, or applying for an observing proposal would allow us to establish if the members of these groups have in fact the same chemical information, and are therefore, birth associations. This is of much importance, as we would have many spurious stars in such groups.

6.3   Final remarks

Overall, the results of this thesis show that we are now in a great position to study the transition from clustered star formation to field stars in an unprecedented way. Thus far, strong chemical tagging has been most effective when used to characterize groups found by other methods, and to test for potential membership. In the future, it seems possible that the most powerful applications of chemical tagging will continue to be in conjunction with other methods (Quillen et al., 2015) rather than by itself. In this thesis we have shown that orbital actions are a powerful tool to discover these groups. A recent review by Krumholz, McKee, & Bland-Hawthorn (2019) discusses that, to date, no open clusters or globular clusters (or their streams) have been discovered using actions, but these are early days. Here in this thesis we have taken the first step and recovered not only known open clusters, but also a stream and two unknown associations using a combination of action-angles and chemical information. And, as we have shown, there are many groups in action-angle space that have yet to be followed up on.

A | Appendix

We divide the appendix in three sections. In the first one we brifly discuss the changes we apply to the model for the spectrophotometric distances presented in (Coronado, Rix, & Trick, 2018). In the second section we discuss in more detail the WB selection, and finally in the third one we show the results we obtain for a metric in action only, p($|\Delta$ [Fe/H]$|$ $\log_{10}\Delta J$).

A.1 Spectrophotometric distances

In this section we show in more detail the changes that we applied to the model in (Coronado, Rix, & Trick, 2018) to calculate the spectro-photometric distances with the LMDR5 $\times$ Gaia DR2 dataset used in this work. Here, we apply the same model for main sequence and binary stars defined in Sec. 3 of that work. We follow closely the same steps defined there, where the absolute magnitude of main sequence stars is a function of the spectroscopic parameters and we expand it up to first order in $\log g$, [Fe/H] and second order in $T_{\rm eff}$. However, in this case the normalization of each parameter by the mean value changes, because the dataset considers a different range in $T_{\rm eff}$. Hence, $\overline{T_{\rm eff}}$ = 5500 K, $\overline{[{\rm Fe/H}]}$ = -0.16 and $\overline{\log g}$ = 4.4. In TableA.1 we show the new parameters obtained with *emcee* of the best fit model for the dataset used in this work, and in Fig. A.1 we show the mean absolute magnitude model fit to MS stars in the LMDR5 sample.

$$
\begin{aligned}
\overline{M_K}(T_{\rm eff}, \log g, [{\rm Fe/H}] \mid \theta_K) = M_0 + {\rm a_T} \frac{T_{\rm eff} - \overline{T_{\rm eff}}}{\overline{T_{\rm eff}}} \\
+ {\rm a_{T_2}} \left(\frac{T_{\rm eff} - \overline{T_{\rm eff}}}{\overline{T_{\rm eff}}} \right)^2 + {\rm a_{logg}}(\log g - \overline{\log g}) \\
+ {\rm a_{FeH}} \left([{\rm Fe/H}] - \overline{[{\rm Fe/H}]} \right).
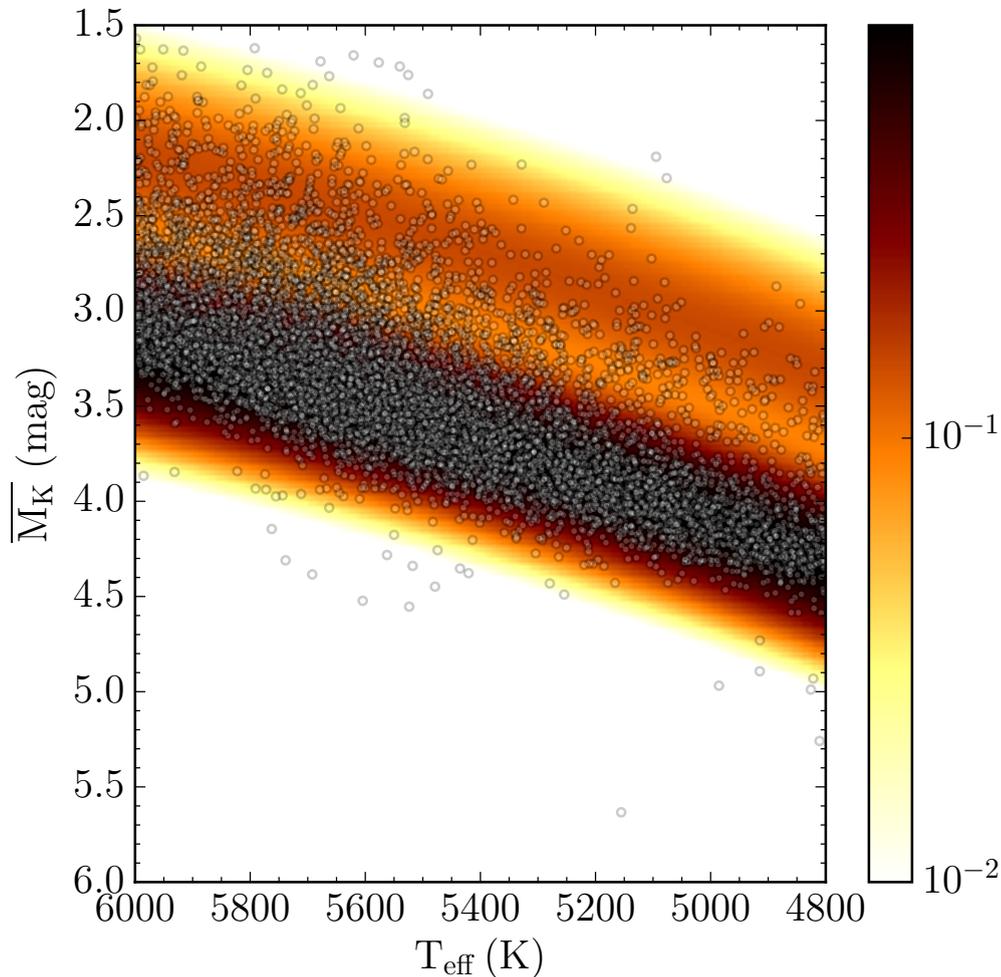\end{aligned}
\tag{A.1}
$$

FIGURE A.1: Mean absolute magnitude model fit to MS stars in the LMDR5 sample. Importantly, the model incorporates the binary sequence. The color represents the density of the model *pdf* for the mean absolute magnitude. This is the best-fit model (parameters from Table A.1), convoluted with a Gaussian of 0.15 mag reflecting the typical parallax uncertainty, for direct comparison with the data.

TABLE A.1: Results obtained with emcee for the parameters of our model.

| Model parameter		Best fit
Peak abs. magnitude in K-band	M_0 (mag)	$3.602^{+0.054}_{-0.100}$
Width of abs. magnitude distribution	σ_1 (mag)	0.145 ± 0.042
Prefactor of $(T_{\mathrm{eff}} - \overline{T_{\mathrm{eff}}})$ term	a_{T}	$-4.935^{+0.077}_{-0.337}$
Prefactor of $(T_{\mathrm{eff}} - \overline{T_{\mathrm{eff}}})^2$ term	a_{T_2}	$5.796^{+0.038}_{-0.150}$
Prefactor of $(\log g - \overline{\log g})$ term	a_{logg}	$1.009^{+0.132}_{-0.067}$
Prefactor of $([\mathrm{Fe/H}] - \overline{[\mathrm{Fe/H}]})$ term	a_{FeH}	$-0.358^{+0.133}_{-0.033}$
Width of binary sequence	σ_2 (mag)	$0.268^{+0.020}_{-0.010}$
Binary fraction (equal mass binaries)	f_{eqb}	0.152 ± 0.001
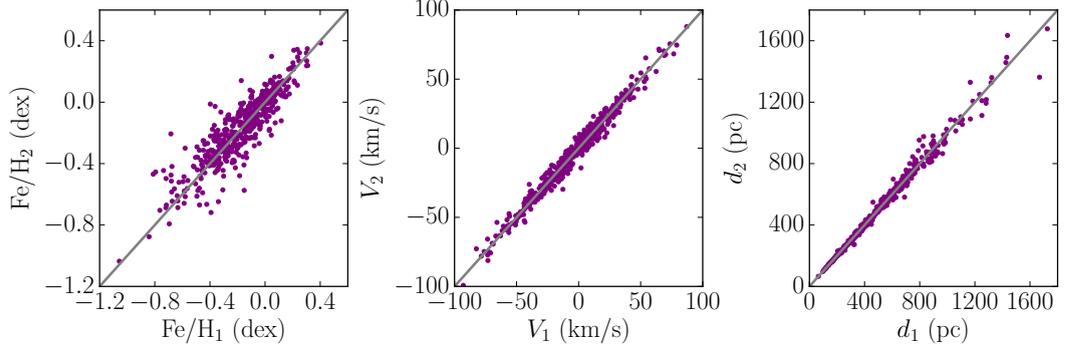
FIGURE A.2: Distribution of the WB sample in distance, velocity and metallicity, where both components show consistent velocities and distances. There is some spread in the metallicity distribution, but nonetheless, it seems mostly consistent for both components. Moreover, from Fig. 4.2 we already saw that most of the sample has differences in metallicity for these pairs of Δ[Fe/H] < 0.1. In grey we overplot a 1:1 line that show how well these pairs agree.

A.2 Cross-check with Wide Binaries

In Fig. 4.2 we also plot (in cyan dots) the sample of WB pairs. It is expected that WBs should have similar chemical composition if they formed from the same molecular cloud (Duchêne & Kraus, 2013), and they are also common proper motion pairs, so their phase-space coordinates should also be consistent. Therefore, this is a good sample to compare our results to. In Fig. A.2 we present the velocities, distances and metallicities for this sample, with all of them showing consistent values for stars in a binary system.

We calculate the pairwise distances for each WB as defined in Eqs. 4.3 and 4.5. This distribution in Fig. 4.2 falls in the area of small $\log_{10}\Delta(J, \theta) - \Delta$[Fe/H] as we would have expected: most of the WBs have $\log_{10}\Delta(J, \theta) <$ -1.5 and Δ[Fe/H] < 0.1 dex, with the latter corresponding to the measurement uncertainty in [Fe/H]. This also shows us that the features at small distances in (J, θ) in the histogram of pairwise distances that we have obtained are actually real.

A.3 Metric in action space only

In this section we present the results of the metric in action space only. These results are not intended as a comparison to the metric in action-angle space. As we move from a 6D to 3D coordinate system, then a direct comparison is not possible. However, with these results we want to highlight that with the actions metric we still see a signature of pairs close in ΔJ-Δ[Fe/H]. In the smallest bin this signal seems weaker than the one present when we include the angles, as illustrated in Fig. A.3, but again this is because we are
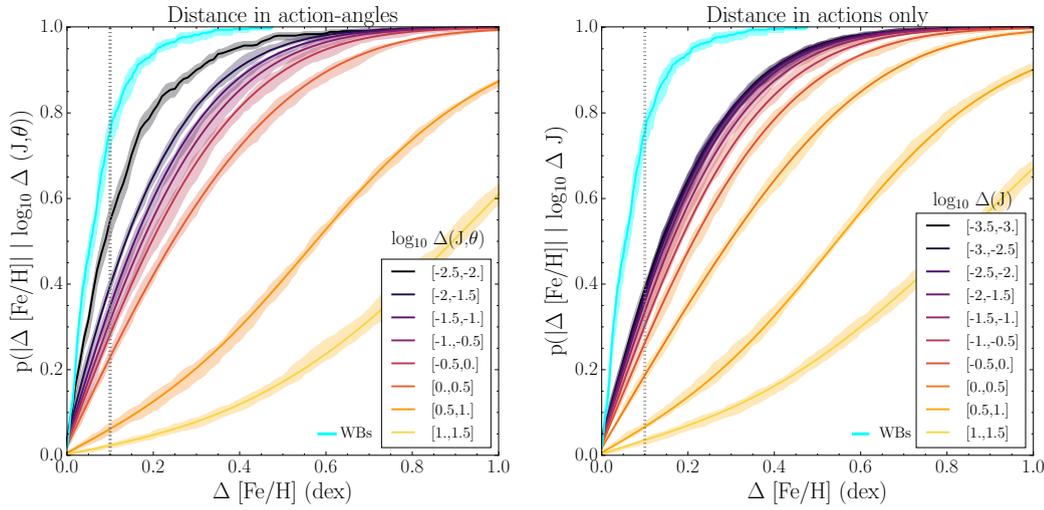
FIGURE A.3: Distribution of pairwise distances in action-angle and actions only. The left side of this two panel figure is analogous to Fig. 4.3. The right side of this figure now shows the CDF of pairs in given distance bins $\Delta(J)$ as a function of Δ[Fe/H] for the LMDR5 $\times$ GDR2 MS stars. Each colored line on the left plot, again corresponds to the bins from Fig. 4.2 at different $\log_{10}\Delta(J,\theta)$ and to bins at different $\log_{10}\Delta(J)$ on the right side. The width of these lines show the 5th and the 95th percentile of a bootstrap re-sampling. The cyan line shows the complete distribution of WBs, for comparison. The dashed line is located at Δ[Fe/H] $= 0.1$, that we consider as an upper limit for the uncertainties in [Fe/H]. We observe that the distance in actions only reaches smaller values than $\log_{10}\Delta(J,\theta)$. Even though for the first 4 bins it seems that each line lie in the same position, the rest of the bins show the same trend as in $\log_{10}\Delta(J,\theta)$, and we still see some signature present when considering actions only.

not including the angles information. Analogous to Fig. 4.3, the right side of this figure presents the results of p($|\Delta$ [Fe/H]$|$ $\log_{10}\Delta J$). Each line here is colored at different bins of the $\log_{10}\Delta J$ histogram. For the smallest bin, we find that $\sim 40\%$ of these pairs is at Δ [Fe/H] $= 0.1$ dex. We notice that we find smaller values of $\log_{10}\Delta J$ as compared to the ones found for $\log_{10}\Delta(J,\theta)$, with the smallest bin at $\log_{10}\Delta J = -3.5$. The first 4 bins are overlapped and don't show much difference between them, but the rest of them, from $\log_{10}\Delta J = -1.5$ on wards show the same features as $\log_{10}\Delta(J,\theta)$.

Finally, Fig. A.4 shows the mapping of $\log_{10}\Delta J$ into velocity-distance space to the right side. With the actions only metric we find at the smallest bin, pairs of stars between 0.1-1 kpc in $\Delta\vec{r}$ not as different to what we find with $\log_{10}\Delta(J,\theta)$ at 0.01-0.5 kpc. For $\Delta\vec{v}$ we find a larger difference, however the spread for $\log_{10}\Delta(J,\theta)$ is much larger in velocity space. These plots are not intended as a direct comparison between $\log_{10}\Delta(J,\theta)$ and $\log_{10}\Delta$ J. When moving from 6D to 3D coordinates inevitably we lose information. Nevertheless, we want to show that actions are still a valid coordinate system, where we can still find valuable information for pairs that are close in both $\log_{10}\Delta J$-Δ[Fe/H].
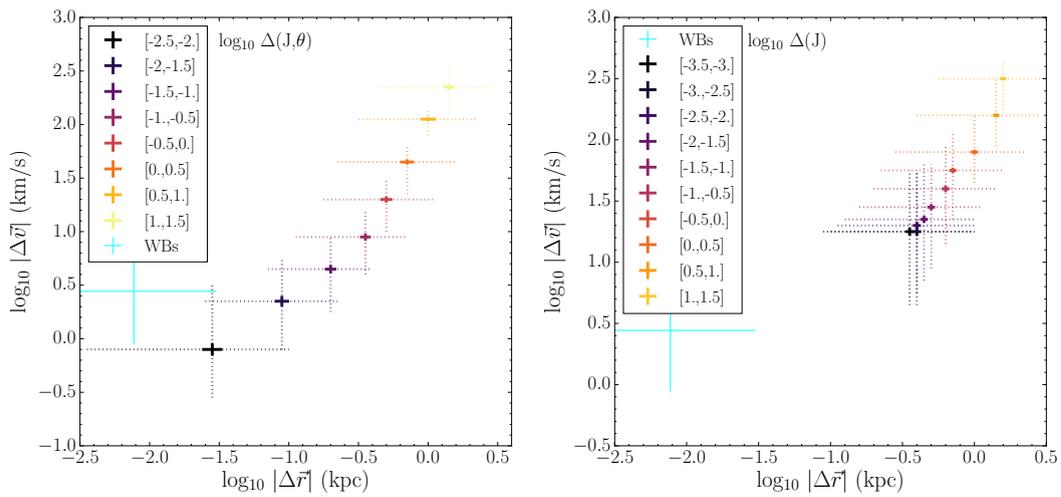
FIGURE A.4: Differences in 3D velocities $\Delta\vec{v}$ and positions $\Delta\vec{r}$, for the same bins in $\log_{10}\Delta(J,\theta)$ as shown in Fig. A.3 to the left, and $\log_{10}\Delta(J)$ to the right. The left side of this plot is analogous to Fig. 4.5, but now we are considering bins of 0.5 in $\log_{10}\Delta(J,\theta)$. Again, we show the WBs in cyan, that are located at small $\Delta\vec{v}$ and $\Delta\vec{r}$. For the bins, the solid lines show the uncertainty of the mean value (calculated via bootstrapping), and the dashed line shows the 5th and 95th percentile. For stars close in $\log_{10}\Delta(J)$ we see that it maps into large values in $\Delta\vec{r}$ and $\Delta\vec{v}$ as compared to what we observe in when we combine both actions and angles. However, this is expected as we are considering less information. We still notice that some information is present, as the first bin in $\log_{10}\Delta(J)$ seems to correspond to the third one in $\log_{10}\Delta(J,\theta)$, finding pairs with similar $\Delta(J)$-Δ[Fe/H] at $\Delta(\vec{r})$=0.39 kpc, extending up to 1kpc.

Publications

The content presented in this thesis has been largely based on the following published research papers:

Chapter 2:

>**Coronado, Johanna**, Rix, Hans-Walter & Trick, Wilma H., 2018, "*Unbiased TGAS×LAMOST distances and the role of binarity*", MNRAS, 481, 2970

Chapter 3:

>Trick, Wilma H., **Coronado, Johanna** & Rix, Hans-Walter, 2019, "*The Galactic disc in action space as seen by Gaia DR2*", MNRAS, 484, 3291

Chapter 4:

>**Coronado, Johanna**, Rix, Hans-Walter, Trick, Wilma H., El-Badry Kareem, Rybizki, Jan, Xiang, Maosheng, 2020, "*From birth associations to field stars: mapping the small-scale orbit distribution in the Galactic disc*", MNRAS, 495, 4098

I am also first and co-author on the following publications, which are not included in this thesis:

>Xiang, Maosheng, Rix, Hans-Walter, Yuan-Sen, Ting, Ludwig, Hans-Günter, **Coronado, Johanna**, et al., 2020, *Chemically peculiar A and F stars with enhanced s-process and iron-peak elements: stellar radiative acceleration at work*, accepted for publication in ApJ

>Sellwood, J. A., Trick, Wilma H., Carlberg, R. G., **Coronado, Johanna**, Rix, Hans-Walter, 2019, "*Discriminating among theories of spiral structure using Gaia DR2*", MNRAS, 484, 3154

>Placco, Vinicius M., Santucci, Rafael M., Beers, Timothy C., Chanamé, Julio, Sepúlveda, María Paz, **Coronado, Johanna**, et al., 2019, *The R-Process Alliance: Spectroscopic*

Follow-up of Low-metallicity Star Candidates from the Best & Brightest Survey, ApJ, 870, 122

Coronado, Johanna, Sepúlveda, María Paz, Gould, Andrew, Chanamé, Julio, 2018, *A distant sample of halo wide binaries from SDSS*, MNRAS, 480, 4302

Placco, Vinicius M., Beers, Timothy C., Santucci, Rafael M., Chanamé, Julio, Sepúlveda, María Paz, **Coronado, Johanna**, et al., 2018, *Spectroscopic Validation of Low-metallicity Stars from RAVE*, AJ, 155, 256

Coronado, J., Helminiak, K. G., Vanzi, L., Espinoza, N., Brahm, R., Jordán, A., Catelan, M., Ratajczak, M., Konacki, M. *Orbital and physical parameters of eclipsing binaries from the ASAS catalogue - VII. V1200 Centauri: a bright triple in the Hyades moving group*, MNRAS, 448, 1937

Bibliography

Adamo, Angela & Bastian, Nate (2018). In: *The Birth of Star Clusters*. Ed. by Steven Stahler. **424**. Astrophysics and Space Science Library, p. 91.

Adibekyan, V. Zh. et al. (2011). *A&A* **535**, L11, p. L11.

Allison, Richard J. (2012). *MNRAS* **421**, pp. 3338–3342.

Antoja, T. et al. (2008). *A&A* **490**, pp. 135–150.

Antoja, T. et al. (2009). *ApJ* **700**, pp. L78–L82.

Arnold, Vladimir Igorevich (1978). *Mathematical methods of classical mechanics*. Springer.

Bailer-Jones, C. A. L. et al. (2013). *A&A* **559**, A74, A74.

Bailer-Jones, Coryn A. L. (2015). *PASP* **127**, p. 994.

Bekki, Kenji & Chiba, Masashi (2001). *ApJ* **558**, pp. 666–686.

Bensby, T., Feltzing, S., & Lundström, I. (2003). *A&A* **410**, pp. 527–551.

Beuther, H. et al. (2017). *A&A* **597**, A85, A85.

Binney, J. (2012). *MNRAS* **426**, pp. 1324–1327.

Binney, J. & Spergel, D. (1984). *MNRAS* **206**, pp. 159–177.

Binney, J. & Tremaine, S. (2008). *Galactic Dynamics: Second Edition*. Princeton University Press.

Bird, Jonathan C., Kazantzidis, Stelios, & Weinberg, David H. (2012). *MNRAS* **420**, pp. 913–925.

Bird, Jonathan C. et al. (2013). *ApJ* **773**, 43, p. 43.

Bissantz, Nicolai, Englmaier, Peter, & Gerhard, Ortwin (2003). *MNRAS* **340**, pp. 949–968.

Bissantz, Nicolai & Gerhard, Ortwin (2002). *MNRAS* **330**, pp. 591–608.

Bizyaev, D. V. et al. (2014). *ApJ* **787**, 24, p. 24.

Blanco-Cuaresma, S. & Soubiran, C. (2016). In: *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pp. 333–336.

Blanco-Cuaresma, S. et al. (2015). *A&A* **577**, A47, A47.

Bland-Hawthorn, Joss & Gerhard, Ortwin (2016). *ARA&A* **54**, pp. 529–596.

Bland-Hawthorn, Joss, Krumholz, Mark R., & Freeman, Ken (2010). *ApJ* **713**, pp. 166–179.

Boesgaard, Ann Merchant, Roper, Brian W., & Lum, Michael G. (2013). *ApJ* **775**, 58, p. 58.

Bonnell, Ian A. & Dobbs, Clare L. (2007). In: *Triggered Star Formation in a Turbulent ISM*. Ed. by Bruce G. Elmegreen & Jan Palous. **237**. IAU Symposium, pp. 344–350.

Bovy, J. (2015). *ApJS* **216**, 29, p. 29.

Bovy, J. & Rix, H.-W. (2013). *ApJ* **779**, 115, p. 115.

Bovy, J. et al. (2016a). *ApJ* **823**, 30, p. 30.

Bovy, Jo (2014). *ApJ* **795**, 95, p. 95.

Bovy, Jo (2016). *ApJ* **817**, 49, p. 49.

Bovy, Jo, Rix, Hans-Walter, & Hogg, David W. (2012). *ApJ* **751**, 131, p. 131.

Bovy, Jo et al. (2012). *ApJ* **759**, 131, p. 131.

Bovy, Jo et al. (2016b). *ApJ* **833**, 31, p. 31.

Brand, J. & Blitz, L. (1993). *A&A* **275**, pp. 67–90.

Brandt, Timothy D. & Huang, Chelsea X. (2015). *ApJ* **807**, 24, p. 24.

Bressan, Alessandro et al. (2012). *MNRAS* **427**, pp. 127–145.

Bressert, E. et al. (2010). *MNRAS* **409**, pp. L54–L58.

Buder, S. et al. (2018). *MNRAS* **478**, pp. 4513–4552.

Bullock, James S. & Johnston, Kathryn V. (2005). *ApJ* **635**, pp. 931–949.

Burnett, B. & Binney, J. (2010). *MNRAS* **407**, pp. 339–354.

Burstein, D. (1979). *ApJ* **234**, pp. 829–836.

Busha, Michael T. et al. (2011). *ApJ* **743**, 40, p. 40.

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. (1989). *ApJ* **345**, pp. 245–256.

Carlin, J. L. et al. (2015). *AJ* **150**, 4, p. 4.

Carpenter, John M., Heyer, Mark H., & Snell, Ronald L. (2000). *ApJS* **130**, pp. 381–402.

Casagrande, L. et al. (2011). *A&A* **530**, A138, A138.

Casey, A. R. et al. (2017). *ApJ* **840**, 59, p. 59.

Ceverino, Daniel et al. (2012). *MNRAS* **420**, pp. 3490–3520.

Chou, Mei-Yin et al. (2010). *ApJ* **708**, pp. 1290–1309.

Chumak, Ya. O., Rastorguev, A. S., & Aarseth, S. J. (2005). *Astronomy Letters* **31**, pp. 308–314.

Combes, F. et al. (1990). *A&A* **233**, p. 82.

Conselice, Christopher J. (2014). *ARA&A* **52**, pp. 291–337.

Conselice, Christopher J. et al. (2003). *AJ* **126**, pp. 1183–1207.

Conselice, Christopher J. et al. (2016). *ApJ* **830**, 83, p. 83.

Contenta, Filippo et al. (2017). *MNRAS* **466**, pp. 1741–1756.

Contigiani, O., Rossi, E. M., & Marchetti, T. (2019). *MNRAS* **487**, pp. 4025–4036.

Contopoulos, G. (1963). *AJ* **68**, p. 1.

Coronado, J., Rix, H.-W., & Trick, W. H. (2018). *MNRAS* **481**, pp. 2970–2980.

Cropper, M. et al. (2018). *A&A* **616**, A5, A5.

Cui, X.-Q. et al. (2012). *Research in Astronomy and Astrophysics* **12**, pp. 1197–1242.

Curtis, Jason L. et al. (2019). *AJ* **158**, 77, p. 77.

Daflon, Simone & Cunha, Katia (2004). *ApJ* **617**, pp. 1115–1126.

Dalcanton, Julianne J. & Bernstein, Rebecca A. (2002). *AJ* **124**, pp. 1328–1359.

de Boer, T. J. L., Belokurov, V., & Koposov, S. (2015). *MNRAS* **451**, pp. 3489–3503.

de Bruijne, J. H. J. (2012). *Ap&SS* **341**, pp. 31–41.

de Jong, R. S. (1996). *A&AS* **118**, pp. 557–573.

De Silva, G. M. et al. (2007a). *AJ* **133**, pp. 1161–1175.

De Silva, G. M. et al. (2007b). *AJ* **133**, pp. 694–704.

de Vaucouleurs, Gerard (1948). *Annales d'Astrophysique* **11**, p. 247.

de Vaucouleurs, Gerard (1959). *Handbuch der Physik* **53**, p. 275.

Debernardi, Y. et al. (2000). *VizieR Online Data Catalog*, J/A+A/354/881, J/A+A/354/881.

Dehnen, Walter (1998). *AJ* **115**, pp. 2384–2396.

Dehnen, Walter (2000). *AJ* **119**, pp. 800–812.

Deng, L.-C. et al. (2012). *Research in Astronomy and Astrophysics* **12**, pp. 735–754.

Dobbs, C. L. et al. (2018). *MNRAS* **478**, pp. 3793–3808.

Douglas, S. T. et al. (2014). *ApJ* **795**, 161, p. 161.

Duarte, Manuel & Mamon, Gary A. (2014). *MNRAS* **440**, pp. 1763–1778.

Duchêne, Gaspard & Kraus, Adam (2013). *ARA&A* **51**, pp. 269–310.

Dwek, E. et al. (1995). *ApJ* **445**, p. 716.

Edvardsson, B. et al. (1993). *A&A* **500**, pp. 391–442.

Eggen, O. J. (1959). *The Observatory* **79**, pp. 143–148.

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. (1962). *ApJ* **136**, p. 748.

Eggen, Olin J. (1960). *MNRAS* **120**, p. 540.

Eilers, Anna-Christina et al. (2019). *ApJ* **871**, 120, p. 120.

El-Badry, K. et al. (2018). *MNRAS* **476**, pp. 528–553.

El-Badry, Kareem & Rix, Hans-Walter (2018). *MNRAS* **480**, pp. 4884–4902.

El-Badry, Kareem et al. (2019). *MNRAS* **489**, pp. 5822–5857.

Elmegreen, B. G. & Elmegreen, D. M. (1983). *MNRAS* **203**, pp. 31–45.

Ester, Martin et al. (1996). In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD-96. Portland, Oregon: AAAI Press, pp. 226–231.

Evans, D. W. et al. (2018). *A&A* **616**, A4, A4.

Fall, S. M. & Efstathiou, G. (1980). *MNRAS* **193**, pp. 189–206.

Famaey, B. et al. (2005). *A&A* **430**, pp. 165–186.

Feltzing, S., Bensby, T., & Lundström, I. (2003). *A&A* **397**, pp. L1–L4.

Feng, Y. & Modi, C. (2017). *Astronomy and Computing* **20**, pp. 44–51.

Ferguson, A. M. N. & Clarke, C. J. (2001). *MNRAS* **325**, pp. 781–791.

Ferguson, Annette et al. (2007). *Astrophysics and Space Science Proceedings* **3**, p. 239.

Ferguson, Annette M. N. et al. (1998). *ApJ* **506**, pp. L19–L22.

Foreman-Mackey, D. et al. (2013). *PASP* **125**, p. 306.

Fouvry, Jean-Baptiste, Binney, James, & Pichon, Christophe (2015). *ApJ* **806**, 117, p. 117.

Fragkoudi, F. et al. (2017). *A&A* **606**, A47, A47.

Frankel, Neige et al. (2018). *ApJ* **865**, 96, p. 96.

Freeman, K. C. (1970). *ApJ* **160**, p. 811.

Freeman, Ken & Bland-Hawthorn, Joss (2002). *ARA&A* **40**, pp. 487–537.

Fuhrmann, Klaus (1998). *A&A* **338**, pp. 161–183.

Fuhrmann, Klaus (2008). *MNRAS* **384**, pp. 173–224.

Gaia Collaboration et al. (2016a). *A&A* **595**, A2, A2.

Gaia Collaboration et al. (2016b). *A&A* **595**, A1, A1.

Gaia Collaboration et al. (2017). *A&A* **601**, A19, A19.

Gaia Collaboration et al. (2018a). *A&A* **616**, A11, A11.

Gaia Collaboration et al. (2018b). *A&A* **616**, A1, A1.

Garcia-Dias, Rafael et al. (2019). *A&A* **629**, A34, A34.

Gibbons, S. L. J., Belokurov, V., & Evans, N. W. (2014). *MNRAS* **445**, pp. 3788–3802.

Gilmore, G. & Reid, N. (1983). *MNRAS* **202**, pp. 1025–1047.

Gilmore, G. et al. (2012). *The Messenger* **147**, pp. 25–31.

Gilmore, Gerard, Wyse, Rosemary F. G., & Kuijken, Konrad (1989). *ARA&A* **27**, pp. 555–627.

Gossage, Seth et al. (2018). *ApJ* **863**, 67, p. 67.

Grand, R. J. J. et al. (2018). *MNRAS* **474**, pp. 3629–3639.

Grand, Robert J. J. et al. (2016). *MNRAS* **459**, pp. 199–219.

Green, G. M. et al. (2015). *ApJ* **810**, 25, p. 25.

Hawkins, K. et al. (2017). *MNRAS* **471**, pp. 722–729.

Hawkins, Keith, Lucey, Madeline, & Curtis, Jason (2020). *arXiv e-prints*, arXiv:2003.04940, arXiv:2003.04940.

Hawkins, Keith et al. (2020). *MNRAS* **492**, pp. 1164–1179.

Hayden, M. R. et al. (2017). *A&A* **608**, L1, p. L1.

Haywood, M. (2008). *MNRAS* **388**, pp. 1175–1184.

Haywood, M. et al. (2013). *A&A* **560**, A109, A109.

Hoeg, E. et al. (1997). *A&A* **323**, pp. L57–L60.

Hogg, David W. et al. (2016). *ApJ* **833**, 262, p. 262.

Holmberg, E. (1958). *Meddelanden fran Lunds Astronomiska Observatorium Serie II* **136**, p. 1.

Houk, Nancy & Swift, Carrie (1999). *Michigan catalogue of two-dimensional spectral types for the HD Stars ; vol. 5*. **5**. Dept. of Astronomy, University of Michigan.

Howard, Christian D. et al. (2009). *ApJ* **702**, pp. L153–L157.

Hubble, E. P. (1926). *ApJ* **64**, pp. 321–369.

Humphreys, R. M. & Sandage, A. (1980). *ApJS* **44**, pp. 319–381.

Hunt, Jason A. S. et al. (2019). *MNRAS* **490**, pp. 1026–1043.

Hurley, J. & Tout, C. A. (1998). *MNRAS* **300**, pp. 977–980.

Ibukiyama, A. & Arimoto, N. (2002). *A&A* **394**, pp. 927–941.

Ivezić, Ž. et al. (2008). *ApJ* **684**, 287-325, pp. 287–325.

Jeans, J. H. (1916). *MNRAS* **76**, pp. 552–572.

Jofré, Paula, Heiter, Ulrike, & Soubiran, Caroline (2019). *ARA&A* **57**, pp. 571–616.

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

Jurić, Mario et al. (2008). *ApJ* **673**, pp. 864–914.

Kamdar, Harshil et al. (2019). *ApJ* **884**, L42, p. L42.

Katz, D. et al. (2019). *A&A* **622**, A205, A205.

Kautsch, S. J. et al. (2006). *A&A* **451**, pp. 1171–1171.

Kendall, S., Kennicutt, R. C., & Clarke, C. (2011). *MNRAS* **414**, pp. 538–564.

Kennicutt R. C., Jr. (1981). *AJ* **86**, pp. 1847–1858.

Kennicutt Robert C., Jr. (1989). *ApJ* **344**, p. 685.

Kharchenko, N. V. et al. (2013). *A&A* **558**, A53, A53.

Kordopatis, G. et al. (2011). *A&A* **535**, A107, A107.

Kormendy, J. & Norman, C. A. (1979). *ApJ* **233**, pp. 539–552.

Kormendy, John & Kennicutt Robert C., Jr. (2004). *ARA&A* **42**, pp. 603–683.

Kormendy, John et al. (2009). *ApJS* **182**, pp. 216–309.

Kovács, Géza et al. (2014). *MNRAS* **442**, pp. 2081–2093.

Kruijssen, J. M. Diederik (2012). *MNRAS* **426**, pp. 3008–3040.

Kruk, Sandor J. et al. (2018). *MNRAS* **473**, pp. 4731–4753.

Krumholz, Mark R. (2014). *Phys. Rep.* **539**, pp. 49–134.

Krumholz, Mark R., McKee, Christopher F., & Bland-Hawthorn, Joss (2019). *ARA&A* **57**, pp. 227–303.

Krumholz, Mark R. & Ting, Yuan-Sen (2018). *MNRAS* **475**, pp. 2236–2252.

Lada, Charles J. & Lada, Elizabeth A. (2003). *ARA&A* **41**, pp. 57–115.

Law, David R., Johnston, Kathryn V., & Majewski, Steven R. (2005). *ApJ* **619**, pp. 807–823.

Lee, Young Sun et al. (2011). *ApJ* **738**, 187, p. 187.

Leung, Henry W. & Bovy, Jo (2019). *MNRAS* **483**, pp. 3255–3277.

Lilly, Simon J. et al. (2013). *ApJ* **772**, 119, p. 119.

Lin, C. C. & Shu, Frank H. (1964). *ApJ* **140**, p. 646.

Lindblad, Bertil (1963). *Stockholms Observatoriums Annaler* **5**, p. 5.

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

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

Liu, F. et al. (2016a). *MNRAS* **463**, pp. 696–704.

Liu, F. et al. (2016b). *MNRAS* **457**, pp. 3934–3948.

Loebman, Sarah R. et al. (2014). *ApJ* **794**, 151, p. 151.

López-Corredoira, M. et al. (2002). *A&A* **394**, pp. 883–899.

MacArthur, Lauren A., Courteau, Stéphane, & Holtzman, Jon A. (2003). *ApJ* **582**, pp. 689–722.

Majewski, S. R. et al. (2017a). *AJ* **154**, 94, p. 94.

Majewski, Steven R. et al. (2017b). *AJ* **154**, 94, p. 94.

Marasco, Antonino et al. (2015). *MNRAS* **451**, pp. 4223–4237.

Martinez-Valpuesta, Inma, Shlosman, Isaac, & Heller, Clayton (2006). *ApJ* **637**, pp. 214–226.

McConnachie, Alan W. et al. (2006). *ApJ* **647**, pp. L25–L28.

McMillan, Paul J. (2011a). *MNRAS* **414**, pp. 2446–2457.

McMillan, Paul J. (2011b). *MNRAS* **418**, pp. 1565–1574.

McMillan, Paul J. et al. (2018). *MNRAS* **477**, pp. 5279–5300.

McWilliam, Andrew & Zoccali, Manuela (2010). *ApJ* **724**, pp. 1491–1502.

Meingast, Stefan & Alves, João (2019). *A&A* **621**, L3, p. L3.

Meingast, Stefan, Alves, João, & Fürnkranz, Verena (2019). *A&A* **622**, L13, p. L13.

Melnikov, S. & Eislöffel, J. (2012). *A&A* **544**, A111, A111.

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

Michalik, D. et al. (2014). *A&A* **571**, A85, A85.

Miller, R. H. & Smith, B. F. (1979). *ApJ* **227**, pp. 785–797.

Minchev, I. & Famaey, B. (2010). *ApJ* **722**, pp. 112–121.

Minchev, I. et al. (2011). *A&A* **527**, A147, A147.

Minchev, I. et al. (2017). *ApJ* **834**, 27, p. 27.

Mo, Houjun, van den Bosch, Frank C., & White, Simon (2010). *Galaxy Formation and Evolution*. Cambridge University Press.

Munn, Jeffrey A. et al. (2004). *AJ* **127**, pp. 3034–3042.

Naab, Thorsten & Ostriker, Jeremiah P. (2017). *ARA&A* **55**, pp. 59–109.

Navarro, Julio F. et al. (2011). *MNRAS* **412**, pp. 1203–1209.

Ness, M. et al. (2015). *ApJ* **808**, 16, p. 16.

Ness, M. et al. (2018). *ApJ* **853**, 198, p. 198.

Ness, M. K. et al. (2019). *ApJ* **883**, 177, p. 177.

Ness, Melissa & Lang, Dustin (2016). *AJ* **152**, 14, p. 14.

Nordström, B. et al. (2004). *A&A* **418**, pp. 989–1019.

Norris, John (1987). *ApJ* **314**, p. L39.

Oh, Semyeong et al. (2017). *AJ* **153**, 257, p. 257.

Ollongren, A. (1965). *ARA&A* **3**, p. 113.

Peebles, P. J. E. (1969). *ApJ* **155**, p. 393.

Perryman, M. A. C. et al. (1997). *A&A* **500**, pp. 501–504.

Perryman, M. A. C. et al. (1998). *A&A* **331**, pp. 81–120.

Pfalzner, S. (2013). *A&A* **549**, A82, A82.

Pichardo, Bárbara et al. (2012). *AJ* **143**, 73, p. 73.

Piffl, T. et al. (2014). *MNRAS* **445**, pp. 3133–3151.

Portail, M., Wegg, C., & Gerhard, O. (2015). *MNRAS* **450**, pp. L66–L70.

Posti, Lorenzo & Helmi, Amina (2019). *A&A* **621**, A56, A56.

Price-Jones, Natalie & Bovy, Jo (2019). *MNRAS* **487**, pp. 871–886.

Queiroz, A. B. A. et al. (2018). *MNRAS* **476**, pp. 2556–2583.

Queiroz, A. B. A. et al. (2019). *arXiv e-prints*, arXiv:1912.09778, arXiv:1912.09778.

Quillen, A. C. & Minchev, Ivan (2005). *AJ* **130**, pp. 576–585.

Quillen, Alice C. et al. (2015). *MNRAS* **450**, pp. 2354–2366.

Ragan, S. E. et al. (2018). *MNRAS* **479**, pp. 2361–2373.

Randich, S. et al. (2006). *A&A* **450**, pp. 557–567.

Recio-Blanco, A. et al. (2016). *A&A* **585**, A93, A93.

Reddy, Arumalla B. S., Giridhar, Sunetra, & Lambert, David L. (2015). *MNRAS* **450**, pp. 4301–4322.

Reid, M. J. et al. (2014). *ApJ* **783**, 130, p. 130.

Renaud, Florent (2018). *New Astron. Rev.* **81**, pp. 1–38.

Rix, Hans-Walter & Bovy, Jo (2013). *A&ARv* **21**, 61, p. 61.

Rix, Hans-Walter et al. (2016). *ApJ* **826**, L25, p. L25.

Robin, A. C. et al. (2003). *A&A* **409**, pp. 523–540.

Rodrigues, T. S. et al. (2014). *MNRAS* **445**, pp. 2758–2776.

Rodríguez-Puebla, Aldo et al. (2017). *MNRAS* **470**, pp. 651–687.

Röser, Siegfried & Schilbach, Elena (2020). *arXiv e-prints*, arXiv:2002.03610, arXiv:2002.03610.

Röser, Siegfried, Schilbach, Elena, & Goldman, Bertrand (2019). *A&A* **621**, L2, p. L2.

Roškar, Rok et al. (2008). *ApJ* **684**, p. L79.

Rubin, V. C. (1983). *Science* **220**, pp. 1339–1344.

Ruiz-Dern, L. et al. (2018). *A&A* **609**, A116, A116.

Rybizki, J. et al. (2018). *PASP* **130**, p. 074101.

Sacchi, E. et al. (2019). *ApJ* **878**, 1, p. 1.

Sachdeva, Sonali et al. (2015). *MNRAS* **451**, pp. 2–16.

Saintonge, Amélie et al. (2013). *ApJ* **778**, 2, p. 2.

Salaris, M., Weiss, A., & Percival, S. M. (2004). *A&A* **414**, pp. 163–174.

Salo, Heikki & Laurikainen, Eija (2000). *MNRAS* **319**, pp. 377–392.

Sanders, J. L. & Binney, J. (2015). *MNRAS* **449**, pp. 3479–3502.

Schiavon, Ricardo P. et al. (2017). *MNRAS* **465**, pp. 501–524.

Schmidt, Maarten (1959). *ApJ* **129**, p. 243.

Schönrich, R. & Aumer, M. (2017). *MNRAS* **472**, pp. 3979–3998.

Schönrich, Ralph & Binney, James (2009). *MNRAS* **399**, pp. 1145–1156.

Schönrich, Ralph, Binney, James, & Dehnen, Walter (2010). *MNRAS* **403**, pp. 1829–1833.

Schönrich, Ralph, McMillan, Paul, & Eyer, Laurent (2019). *MNRAS* **487**, pp. 3568–3580.

Sellwood, J. A. (2011). *MNRAS* **410**, pp. 1637–1646.

Sellwood, J. A. (2012). *ApJ* **751**, 44, p. 44.

Sellwood, J. A. (2014). *Reviews of Modern Physics* **86**, pp. 1–46.

Sellwood, J. A. & Binney, J. J. (2002). *MNRAS* **336**, pp. 785–796.

Sellwood, J. A. & Carlberg, R. G. (1984). *ApJ* **282**, pp. 61–74.

Sellwood, J. A. et al. (2019). *MNRAS* **484**, pp. 3154–3167.

Sérsic, J. L. (1963). *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* **6**, pp. 41–43.

Sesar, Branimir et al. (2017). *ApJ* **838**.

Sharma, S. et al. (2011). *ApJ* **730**, 3, p. 3.

Sheth, Kartik et al. (2008). *ApJ* **675**, pp. 1141–1155.

Shou-kun, Xu et al. (2019). *Chinese Astron. Astrophys.* **43**, pp. 225–236.

Simpson, Jeffrey D. et al. (2019). *MNRAS* **482**, pp. 5302–5315.

Steinmetz, M. et al. (2006). *AJ* **132**, pp. 1645–1668.

Steinmetz, Matthias & Muller, Ewald (1995). *MNRAS* **276**, pp. 549–562.

Steinmetz, Matthias et al. (2020a). *arXiv e-prints*, arXiv:2002.04377, arXiv:2002.04377.

Steinmetz, Matthias et al. (2020b). *arXiv e-prints*, arXiv:2002.04512, arXiv:2002.04512.

Tacconi, L. J. et al. (2013). *ApJ* **768**, 74, p. 74.

Tian, H.-J. et al. (2017). *ApJS* **232**, 4, p. 4.

Tian, Hai-Jun et al. (2020). *ApJS* **246**, 4, p. 4.

Ting, Yuan-Sen, Conroy, Charlie, & Goodman, Alyssa (2015). *ApJ* **807**, 104, p. 104.

Ting, Yuan-Sen & Rix, Hans-Walter (2019). *ApJ* **878**, 21, p. 21.

Ting, Yuan-Sen et al. (2017). *ApJ* **843**, 32, p. 32.

Ting, Yuan-Sen et al. (2019). *ApJ* **879**, 69, p. 69.

Toomre, A. (1981). In: *Structure and Evolution of Normal Galaxies*. Ed. by S. M. Fall & D. Lynden-Bell, pp. 111–136.

Toomre, Alar (1969). *ApJ* **158**, p. 899.

Toomre, Alar (1977). In: *Evolution of Galaxies and Stellar Populations*. Ed. by Beatrice M. Tinsley & D. Campbell Larson Richard B. Gehret, p. 401.

Trick, W. H., Bovy, J., & Rix, H.-W. (2016). *ApJ* **830**, 97, p. 97.

Trick, Wilma H., Coronado, Johanna, & Rix, Hans-Walter (2019). *MNRAS* **484**, pp. 3291–3306.

Trick, Wilma H. et al. (2019). *arXiv e-prints*, arXiv:1906.04786, arXiv:1906.04786.

Tsikoudi, V. (1980). *ApJS* **43**, pp. 365–377.

Twarog, Bruce A., Ashman, Keith M., & Anthony-Twarog, Barbara J. (1997). *AJ* **114**, p. 2556.

Valenti, E. et al. (2016). *A&A* **587**, L6, p. L6.

Vallée, Jacques P. (2008). *AJ* **135**, pp. 1301–1310.

Vallée, Jacques P. (2014). *AJ* **148**, 5, p. 5.

Vallée, Jacques P. (2017). *The Astronomical Review* **13**, pp. 113–146.

van den Bergh, Sidney (1960). *ApJ* **131**, p. 558.

van der Kruit, P. C. & Freeman, K. C. (2011). *ARA&A* **49**, pp. 301–371.

van der Kruit, P. C. & Searle, L. (1981). *A&A* **95**, pp. 105–115.

van der Wel, A. et al. (2014). *ApJ* **788**, 28, p. 28.

van Dokkum, Pieter G. et al. (2013). *ApJ* **771**, L35, p. L35.

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

Villalobos, Álvaro & Helmi, Amina (2008). *MNRAS* **391**, pp. 1806–1827.

Wang, Wenting et al. (2015). *MNRAS* **453**, pp. 377–400.

Watkins, Laura L. et al. (2019). *ApJ* **873**, 118, p. 118.

Webb, Jeremy J. et al. (2013). *ApJ* **764**, 124, p. 124.

Wegg, Christopher & Gerhard, Ortwin (2013). *MNRAS* **435**, pp. 1874–1887.

Wegg, Christopher, Gerhard, Ortwin, & Bieth, Marie (2019). *MNRAS* **485**, pp. 3296–3316.

White, Martin, Cohn, J. D., & Smit, Renske (2010). *MNRAS* **408**, pp. 1818–1834.

White, S. D. M. & Rees, M. J. (1978). *MNRAS* **183**, pp. 341–358.

White, Simon D. M. & Frenk, Carlos S. (1991). *ApJ* **379**, p. 52.

Wielen, R., Fuchs, B., & Dettbarn, C. (1996). *A&A* **314**, p. 438.

Wilkinson, M. I. & Evans, N. W. (1999). *MNRAS* **310**, pp. 645–662.

Wojno, Jennifer et al. (2018). *MNRAS* **477**, pp. 5612–5624.

Wu, Y. et al. (2011). *Research in Astronomy and Astrophysics* **11**, pp. 924–946.

Wu, Y. et al. (2014). In: *Statistical Challenges in 21st Century Cosmology*. Ed. by A. Heavens, J.-L. Starck, & A. Krone-Martins. **306**. IAU Symposium, pp. 340–342.

Xiang, M.-S. et al. (2017). *MNRAS* **467**, pp. 1890–1914.

Xiang, Maosheng et al. (2019). *ApJS* **245**, 34, p. 34.

Xu, Ye, Hou, Li-Gang, & Wu, Yuan-Wei (2018). *Research in Astronomy and Astrophysics* **18**, 146, p. 146.

Xue, Xiang-Xiang, Rix, Hans-Walter, & Zhao, Gang (2009). *Research in Astronomy and Astrophysics* **9**, pp. 1230–1240.

Yadav, R. K. S. et al. (2008). *A&A* **484**, pp. 609–620.

Yanny, B. et al. (2009). *AJ* **137**, 4377-4399, pp. 4377–4399.

Yoachim, Peter & Dalcanton, Julianne J. (2006). *AJ* **131**, pp. 226–249.

Yoshii, Y. (1982). *PASJ* **34**, pp. 365–379.

Young, J. S. & Scoville, N. Z. (1991). *ARA&A* **29**, pp. 581–625.

Yuan, H. B., Liu, X. W., & Xiang, M. S. (2013). *MNRAS* **430**, pp. 2188–2199.

Zaritsky, Dennis (1999). In: *The Third Stromlo Symposium: The Galactic Halo*. **165**. Astronomical Society of the Pacific Conference Series. Astronomical Society of the Pacific, p. 34.

Zaritsky, Dennis, Kennicutt Robert C., Jr., & Huchra, John P. (1994). *ApJ* **420**, p. 87.

Zhao, G. et al. (2012). *Research in Astronomy and Astrophysics* **12**, pp. 723–734.

Zibetti, Stefano & Ferguson, Annette M. N. (2004). *MNRAS* **352**, pp. L6–L10.

Zucker, D. B. et al. (2012). In: *Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way*. Ed. by W. Aoki et al. **458**. Astronomical Society of the Pacific Conference Series, p. 421.