RIJKSUNIVERSITEIT GRONINGEN

BACHELOR THESIS

A hunt for C-19 members: the most metal-poor structure known in the Galaxy

Using data catalogs made with Gaia and Pristine.

Author:
Heidi de Vos
s4021576

Supervisors:
Prof. Else Starkenburg
Akshara Viswanathan
Petra Awad

Abstract

This thesis searched for stars belonging to the current most metal-poor structure in the Galaxy: stellar stream C-19. With a mean metallicity of [Fe/H] = -3.38±0.06 it breaks through the globular cluster metallicity floor and has the opportunity to fill this mysterious observational gap. However, by showing signs of a globular cluster past with its low metallicity dispersion while simultaneously containing the dynamical hotness of a dwarf galaxy, C-19's progenitor remains undefined. Through the selection of stars in 4D and 5D dynamics for several parts of C-19's integrated orbit, new candidate members are gathered. These stars hint at a much wider metallicity and velocity dispersion currently visible in the C-19 members. However, follow-up research is necessary. This thesis offers 86 new stars selected based on 5 dynamical dimensions: a location (α, δ) within 5° of the orbit, a proper motion (μ_α, μ_δ) inside 1° of the obit, and a radial velocity between $v_{r,orbit} \pm 50$km/s. This includes 4 new Blue Horizontal Branch star and 1 variable Asymptotic Giant Branch star, which provide the opportunity for an improved distance calculation by employing standard candles now fitting with C-19 in 5 dimension instead of 4. In addition, the 4D selection found 78 stars containing a metallicity of [Fe/H] < -2.5 (including 36 stars with [Fe/H] < -3), several of which correspond exceptionally well in both coordinate- and proper motion space, which are in need of a radial velocity measurement.

Contents

1 Introduction

This thesis aims to investigate stellar stream C-19 and its enigmatic features. But, in order to appreciate its mystery, let's start from a wider context.

1.1 Milky Way blueprint

Our Galaxy is comprised of different parts: The bulge/bar, the thin disk, the thick disk, the stellar halo and the dark matter halo. See figure 1.
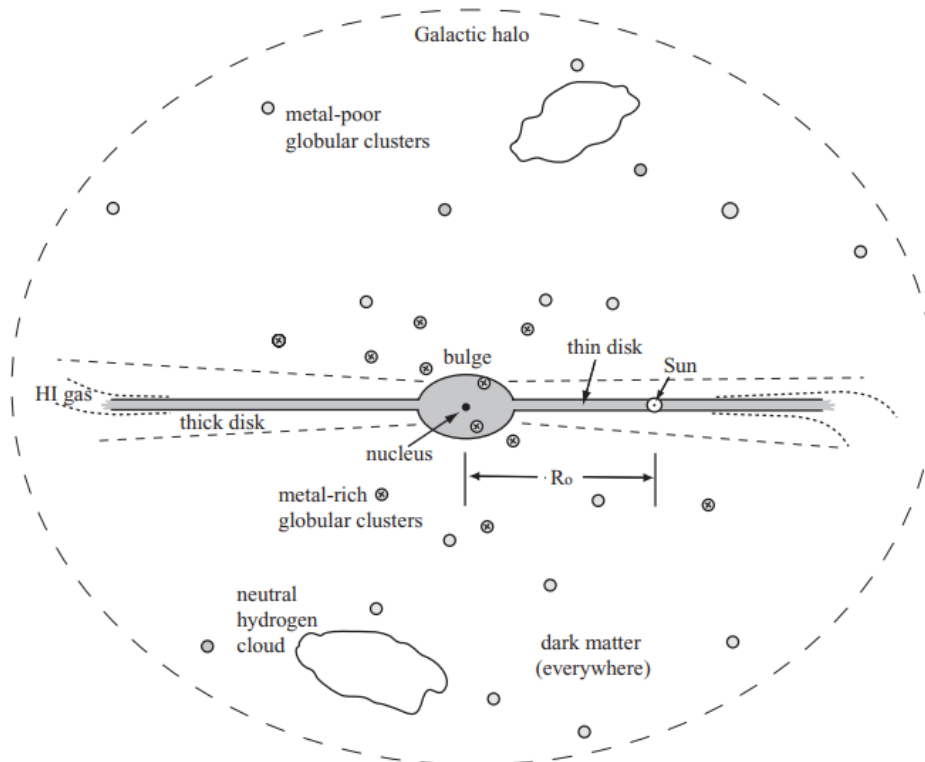

Figure 1: A schematic side view of the Milky Way (Sparke et al. 2007)

Starting in the centre with a radius of a few kpc, the buldge is the most visibly concentrated area. It holds a mass of M $\approx 20 \cdot 10^9 M_\odot$ and a luminosity of L $\approx 5 \cdot 10^9 L_\odot$. Stars in this part rotate with an average speed of about 100 km/s (Sparke et al. 2007). Spectroscopic studies indicate a mix of stellar populations, some of which have [Fe/H] contents up to +0.5 dex while others more than 13 Gyrs old, and some of which resemble populations also found in the thick disk and stellar halo (Helmi 2020).

Moving further out into the thin disk, we arrive at the flat pancake which holds 95% of all the disk stars, and all of the young massive stars. The density of which starts to drop off at the meagre height of about 300-400 pc (Sparke et al. 2007). But within this narrow slab there are stars forming at an estimated SFR of $\sim 1.6 M_\odot \text{year}^{-1}$ (Licquia et al. 2015), and have been doing so for at least 8 or 9 Gyrs (Tononi, Jordi et al. 2019). These stars neatly orbit the galactic centre in circular tracks with average speeds of 200 km/s and little random motions (Sparke et al. 2007). Following similar orbits are stars at a height of $\sim$1 kpc, located in the more diffuse and hotter component of the Milky Way: the thick disk. It holds seniority over the thin disk by $\sim$1.6 Gyr

based on white dwarfs found in the Solar vicinity (Kilic et al. 2017), but also highlighted by its lower metallicity peak at [Fe/H] $\sim$ –0.5, and its differentiating chemical footprint indicating a shorter and more intense star formation history (Helmi 2020).

Entering the stellar halo is where we tread into the neighbourhood of C-19. This region hosts very metal poor, old stars and globular clusters. All moving sporadically in random directions, usually following eccentric orbits (Sparke et al. 2007). The stellar halo is currently estimated to have a mass of M $\sim 1.3 \cdot 10^9 M_\odot$ (Mackereth et al. 2020).

By measuring orbital speeds of gas, stars and clusters located far from the center of the Milky Way, it is found that the mass required to maintain these orbit is more than what is present in the visible parts of our Galaxy. Additionally, most of the Galaxy's mass seems to lie in a sparsely stellar populated spot more than 10 kpc from the centre (Sparke et al. 2007). All this points towards the Milky Way being embedded in a dark matter halo. The current estimates of its mass is $\sim 1.3 \cdot 10^{12} M_\odot$ (Posti, Lorenzo et al. 2019 & Watkins et al. 2019) with a shape debated to be oblate in the central regions but triaxal at larger distances. It is still unclear if the dark matter is homogeneously distributed or if there exists 'lumpiness' like dark satellite's predicted by cold dark matter simulations.

A question is, how can there be streams and clusters of stars in a region like the halo which is devoid of any cold gas to create them? If the dark matter halo is an unlikely origin, then where did they come from and what did they look like before their arrival? In part, that is what the field of galactic archaeology tries to investigate. And in an even smaller part, that is what this thesis tries to figure out for stream C-19.

1.2 Galactic archaeology

1.2.1 Dynamics and merger history

Even though, on first glance, stars in the stellar halo appear to move in a more random fashion, it does not mean there is no order to the chaos. When a galaxy is pulled in by the superior gravitational potential of another galaxy (also called 'tidal stripping'), its stars will continue to follow trajectories close to the system they were stripped from. This original stellar home the stars used to inhabit, but that is now being cannibalized, is called the 'progenitor'. And so in other words, tidally stripped stars will follow the path of their progenitor, resulting in the formation of a stellar stream (Helmi and Simon D. M. White 1999). These streams can be characterized, just like a regular orbit, by the integrals of motion (IoM), such as energy E, total angular momentum for a spherical system or for a asymmetrical system just L_z (Binney and Tremaine 2008). Small progenitors will form narrow streams, whereas bigger disrupted systems will have a larger spread in IoM, and so form broader streams. Other then the size and shape of the progenitor, the stream's properties will be influenced by the amount of time passed since it became unbound (t), and the characteristic orbital timescale (t_{orb}) (Helmi 2020). In the first stages (t $\sim t_{orb}$) the stream will hold onto its density and spacial coherence, but as time goes by the debris starts to mix spatially, and maybe even become so extended it crosses with itself. Locally, the stars can still be characterized by their similar velocities, and so even in more chaotic parts of the galaxy the common origin of these crisscross or stretched out debris streams can still be traced back by their IoM.

The first application of looking for lumpiness in the IoM, resulted in the discovery of the Helmi streams (Helmi, Simon D. M. White, et al. 1999), and other later ancient merger discoveries like Thamnos and Sequoia (Myeong et al. 2019). More recently, this method combined with the abundance of data gathered by Gaia data release 2 was able to show that a large fraction of halo stars near the Sun appear to be debris from a single merger object: Gaia-Enceladus.

These stars form their own distinct sequence in a color-magnitude diagram, and a distinguishing chemical footprint setting them apart from the hot thick disk stars. And consequently, showing they originate from somewhere else. Based an the integration of the SFR estimated by a fitted chemical evolution model, the mass of Gaia-Enceladus is estimated to be $\sim 6 \cdot 10^8 M_\odot$ (Helmi, Babusiaux, et al. 2018). And had to have happened 10 Gyrs ago. The distinguishing kinematic feature of Gaia-Enceladus is visible in figure 2, and can be clearly separated from the hot thick disk stars.
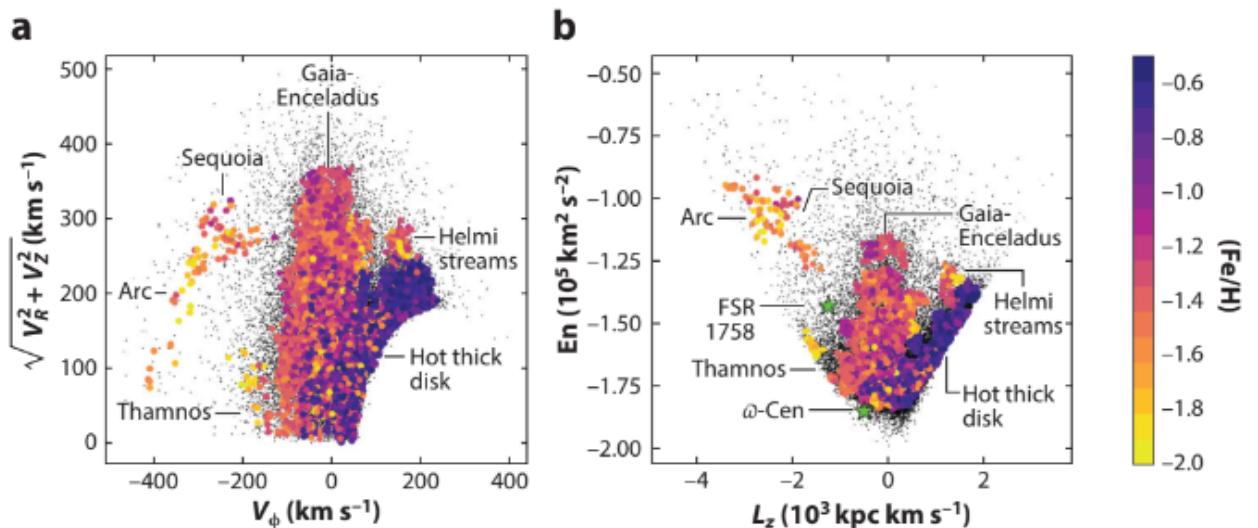

Figure 2: Distribution of stars selected kinematically within 3kpc of the sun with a reliable Gaia parallax, color coded by metallicity. Panel a shows the Tomre diagram of these stars, and panel b shows the E and z angular momentum. (Koppelman et al. 2019)

This discovery was made possible by the 6D information provided by Gaia. Unfortunately, for stars located further out, good quality parallax's and radial velocity measurements start to become sparse. However, an algorithm named STREAMFINDER by (Malhan et al. 2018) managed to successfully bridge this 2 dimensional gap, and identified a multitude of distant halo streams (Nicolas F Martin et al. 2022 and Ibata et al. 2021). One of which, was C-19.

In a nutshell, the lack of distance and radial velocity information is made up for by the crude sampling of these parameters, in combination with statistical analysis. For any given star, the algorithm calculates 3 possible distance values by fitting the observed colour ($G_{BP} - G_{RP}$) to a metal poor isochrone, and, samples radial velocities within the escape velocity range of the milky way ($[-v_{esc}, +v_{esc}]$). The observed proper motions are also sampled within their error margins. This results in 30,000 possible positions in 6D phase space for any star fed to the STREAMFINDER algorithm. Along the 30,000 integrated orbits stars with similar motions are selected and the most likely orbit is determined statistically. From spectroscopic follow-ups, this has shown to be a very successful brute-force method of finding streams (Ibata et al. 2021). As an example, see figure 3 showcasing 26 streams found using STREAMFINDER in combination with the photometric metallicity Pristine Survey.
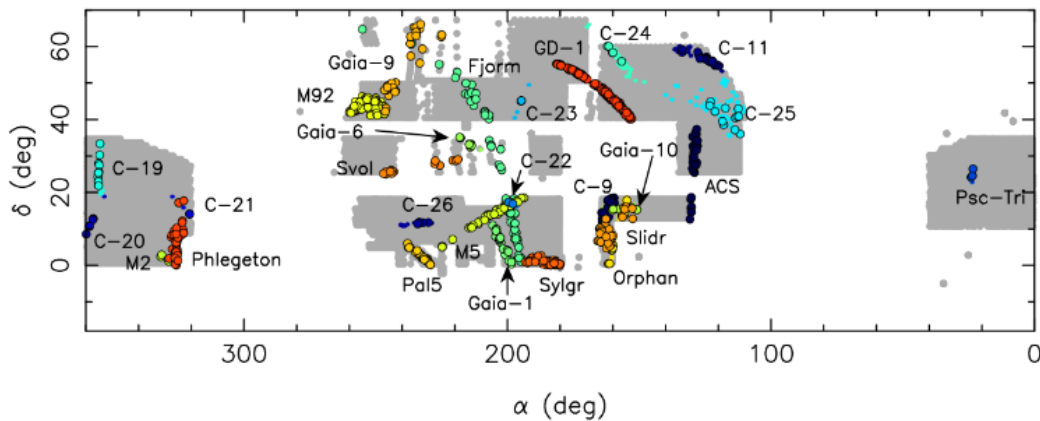
Figure 3: 26 stellar streams found by STREAMFINDER color-coded by membership in equatorial coordinates. The Pristine footprint is shown in grey (Nicolas F Martin et al. 2022).

1.2.2   Chemical tracers

Other method of discerning order within the stellar halo is through chemical abundances. The metals in stars form almost a direct link to their birthplace. Different elements are produced in different environments and on different timescales (McWilliam 1997). For example, a few million years after the birth of a massive star, α elements such as O, Mg, Si, Ca, S and Ti are released in copious amounts when it explodes as a supernova (type II SNe). Comparatively, for a binary system it can take 0.1 to a few Gyrs after its creation to die by the hand of a thermonuclear explosion of a white dwarf in the system (type Ia SNe), which then produces Iron-peak elements like Ti, V, Cr, Mn, Fe, Co, and Ni (Truran et al. 2003). However, details on the burning, the number of dwarfs involved and their masses are still debated (Maoz et al. 2014). Considering the chemical evolution of a (closed) system, this means that over time [α/Fe] will decrease as the interstellar medium (ISM) becomes polluted by type Ia SNe. Eventually, the initially nearly constant [α/Fe] relation with [Fe/H] tips over, after which [α/Fe] will only decline further, with fresh gas infall being the only way to bump it up again(Helmi 2020).

For the creation of elements heavier than the iron peak, the neutron capture process is necessary. This process comes in a slow (s) and rapid (r) form. As its name indicates, the slow process occurs when neutron captures happens more slowly then β-decays. An environment which allows this to take place is the envelope of asymptotic giant branch (AGB) stars (Busso et al. 1999). For stars with [Fe/H] = -1.5, Ba is an example of a element created in this process (Arlandini et al. 1999).

By contrast, the r-process requires a high neutron flux to create rapid neutron captures. There are suitable surroundings for this such as: type II SN, the merger of two neutron stars (Watson et al. 2019), the merger of a neutron star with black holes, or in magneto-rotational SNe (Kobayashi et al. 2023). Sr and Eu are very typical r-process elements. One element not abiding to the distinction made between the r- and s-processes is Nd, which is created in almost equal amount by both processes at solar metallicity (Arlandini et al. 1999). What is not yet completely clear, is the sites and conditions for which neutron-capture elements are produced, particularly at very low metallicities (Cowan et al. 2021).

All of this to say, the chemical DNA of a star shows the environment it came from. A caveat to this is that molecular clouds which are currently collapsing are chemically very alike, making it hard to trace back the origin of younger stars without extremely accurate measurements of

7

many different elements (Helmi 2020). For the older stars in the halo, however, their chemical footprint can tell us something about either the proto-Milky Way or the system they occupied before being cannibalized by our Galaxy. By plotting [α/Fe] vs [Fe/H] or [r/s − process/Fe] vs [Fe/H] of stars in dwarf galaxies within the Local group, we can distinguish different chemical sequences (see figure 4).
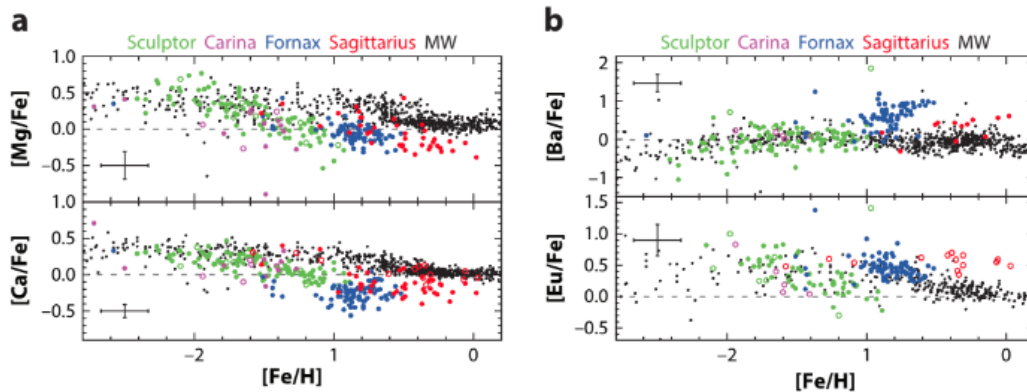

Figure 4: Chemical sequences for 4 spheroidal dwarf galaxies within the Local group shown in color, and the Milky Way shown black. Panel a shows the trends of the α elements and Panel b shows the behaviour of neutron capture elements with metallicity. (Tolstoy et al. 2009)

The galaxies are sorted according to the mass of the system (Helmi 2020). Plotting the abundances of α elements for halo stars like done in panel a could reveal an identifiable trend. For example, low-mass galaxies having formed only one generation of stars will likely show [α/Fe] at low [Fe/H], but galaxies which have been forming stars for longer and therefore more inefficiently will show low [α/Fe] even at low [Fe/H], giving them a different chemical label (Helmi 2020).
The same can be done using r-process element abundances. The possible events activating the formation of r-process elements mentioned earlier are very rare and therefore have not occurred in most ultra-faint dwarf galaxies, given their low masses. But when it does occur, it immediately enriches the galaxy and could show a chemical label with an extreme r-process abundance (Brauer et al. 2019). Massive galaxies on the other hand are more generally expected to show clustering in r-process elemental abundances (Tsujimoto et al. 2018), which, combined with the behavior of [α/Fe] or [Fe/H], could make their chemical label very distinguishable.

1.3 Progenitors

The original home of any accreted structure can be traced back by looking at an overlap between its current features, and the general features associated with existing morphology's. The simplest of which, is the globular cluster.

1.3.1 Globular Clusters

Globular clusters are compact, and roughly spherical groupings of stars found throughout the Milky Way (see figure 1) and other galaxies too. The current estimate of the number of globular clusters (GC) in the Milky Way is ~ 150, with some potentially unobserved GC's being out shined by the bulge, or simply containing a surface brightness too low to be picked up (Binney and Merrifield 1998). The absolute magnitude distribution for these dense clusters is approximately Gaussian with the most common absolute magnitude of M $\sim$ –7, and tail ends at M = –3

and M $= -10.4$ for our Galaxy (Binney and Merrifield 1998). Concentrated near the Galactic centre are globular clusters of all metallicity types, but moving further out [Fe/H] decreases. Currently, the metallicity distribution for GC's in the Milky Way is capped at $[Fe/H]_{min} \approx -2.5$. The lack of any extremely metal poor clusters with $[Fe/H] < -3$ remains a mystery (Harris et al. 1979).

The color-magnitude diagram of globular clusters can bring to light some defining physical features. For instance, where the Main-Sequence (MS) of stars in the solar vicinity extends to include some of the brightest stars in our Galaxy, the MS of globular clusters has a clear turnoff point (see figure 5 and 6).
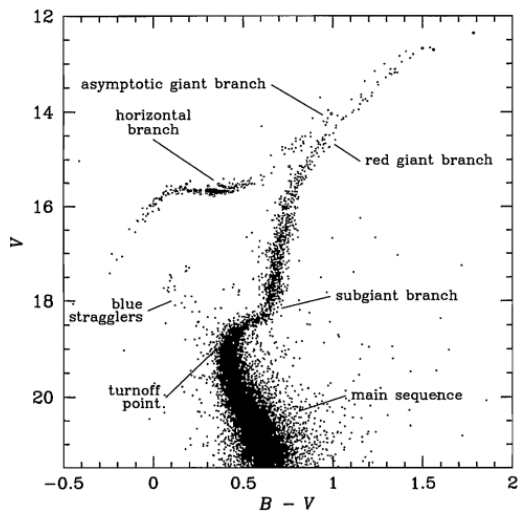

Figure 5: Color-magnitude diagram for globular cluster M3 with annotated principle sequences (Binney and Merrifield 1998).
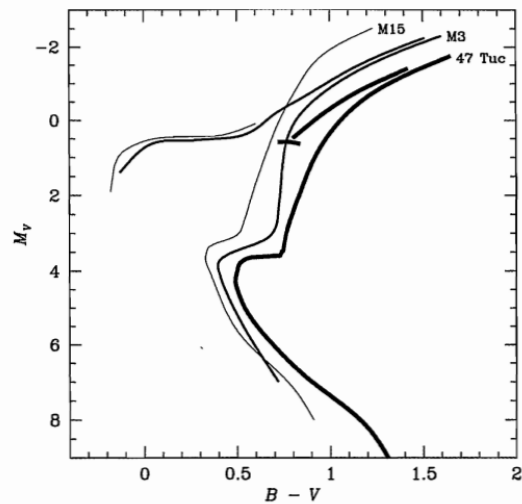

Figure 6: Schematic illustration of prinicple sequences for 3 globular clusters: the metal rich (M47), the metal intermediate (M3) and the metal poor (M15) (Binney and Merrifield 1998).

These missing upper-MS stars have short life spans, indicating that the lack of them could point towards globular clusters being old comparatively (Binney and Merrifield 1998). The sharp turnoff also means that the top MS stars are simultaneously evolving into the sub giant branch, or in other words, these MSTO stars all share nearly identical MS lifespans. which implies a very small age dispersion. Moreover, if the star formation did take place over a extended period, the SGB and the RGB would be widened by the superposition of isochrones of different metallicites, i. e. different ages (see figure 6).

Another defining feature of Globular clusters are their chemical abundances. Even though a narrow sequence in the CMD implies a low metallicity dispersion, the lighter elements in globular cluster show a significant spread. Globular clusters have a distinctive pattern of enrichment in He, N, & Na, and depletion in O & C rarely seen in field stars or open clusters, the specifics of which are unique for each GC. These unique internal variations of chemical abundances are known as multiple populations (MPs), and as of yet, they have an unresolved origin. It was long thought that this internal spread was caused by different generations of star formation, but current evidence is pointing away from this interpretation (Bastian et al. 2018). Evolutionary mixing could explain the N-C anti-correlation seen in evolved RGB stars, but not in the later observed non-evolved MS and MSTO stars. Even if sufficient mixing was achieved during MS evolution, it would result in a fuzzier turn off point contrary to what is observed in the CMD

9

(figure 5). Moreover, even though the individual abundances of C, N, O can variate, the sum C+N+O is generally observed to be constant (Dickens et al. 1991). Additionally, anti correlations of Na-O and possibly Al-Mg were almost universally found in all studied clusters (R. Gratton et al. 2004 and R. G. Gratton et al. 2012). Specifically Na, Al and Mg cannot be created in the fusion reaction of low mass stars for they lack the temperatures necessary to achieve the p-capture reactions needed for the production of said elements (Prantzos, N. et al. 2007 and Prantzos et al. 2017).

In light of all this, it seems the abundance anomalies seen in globular cluster stars we are currently observing, cannot be produced in the course of the evolution of stars but rather produced somewhere else.

1.3.2   Dwarf Galaxies

Figure 7 gives an overview of the varying different size scales for a multitude of galaxy types and globular clusters. The left plot shows this through surface brightness (μ_V) vs absolute magnitude (M_V) and the right plot with the half-light-radius ($r_{1/2}$) vs absolute magnitude (M_V). The ellipses are the typical locations of ellipctical galaxies and bulges (red), spiral galaxy disks (blue), large early (spheroidals), late-type system (dashed black) and galactic nuclei (dashed magenta). Globular clusters are visible as individual small black points. Local group dwarf galaxies are plotted in blue or green open pentagons referring to gas or no gas systems respectively. Figure 7 shows that there is no clear cut between dwarf galaxies and the larger late-type and spheroidals systems.
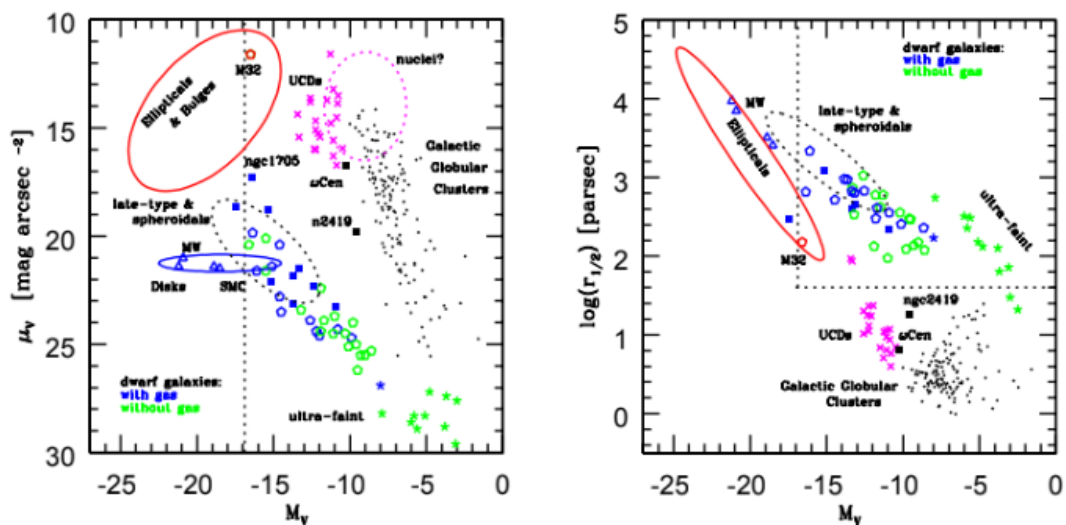

Figure 7: Structural properties for different types of galactic structures and globular clusters. Left shows the apparent surface brightness (μ_V) vs absolute magnitude (M_V), and right shows the half-light-radius ($r_{1/2}$) vs absolute magnitude (M_V) (See Tolstoy et al. 2009 and the references therein).

So, what then makes a dwarf galaxy a dwarf galaxy? A working definition was given in a 1993 conference by G. Tammann: all galaxies more spacially extended than globular clusters (see the dotted lines in figure 7), and fainter than MV ≤ 17 are dwarf galaxies. This does have a physical underpinning, as it is largely consistent with the limit of mass at which outflows start to significantly affect the baryonic mass of a galaxy. Large late-type galaxies above this limit sit

on the constant central surface brightness ridge defined by (Freeman 1970), and have retained most of their initial baryons. But for galaxies below this limit it seems the fainter they are, the more baryons that are lost. This could be explained by a potential well too shallow to hold onto its gas when Supernova winds or tidal interactions occur (Tolstoy et al. 2009). Within this definition there exist a number of different types: early-type dwarf spheroidals; late-type star-forming dwarf irregulars, ultra-faint dwarfs (green stars in fig. 7); centrally concentrated actively star-forming Blue Compact Dwarfs, and the class of even more extreme ultra-compact dwarfs are identified as dwarf galaxies from spectra but are of a similar compactness to globular clusters (see the magenta crosses Fig. 7). Other than the size difference between globular clusters and most types of dwarf galaxies visible in figure 7, the presence of a dark matter halo and the ability to hold onto its gas is another differentiating feature. As a results, this means dwarf galaxies have the ability of long(er) periods of star formation shown also by their bigger metallicity dispersion, and broader CMD sequence.

1.4   Stellar stream C-19

Now that we are equipped with some informational tools, let us have a look at stellar stream C-19. In 2022 the application of the STREAMFINDER algorithm to Gaia's Early Data Release 3 discovered a group of stars moving along a commen orbit in the Milky Way halo (Nicolas F Martin et al. 2022). The photometric metallicities of these stars gathered from the Pristine survey (Nicolas F. Martin, Starkenburg, et al. 2023) revealed an unprecedentedly low metallicity level of $< [Fe/H]_{Pr} >= -3.58 \pm 0.08$. Spectroscopic observations for 8 stars were collected using two different telescopes and spectographs (Gemini/GRACES) and (GTC/OSIRIS), confirming the coherent and extremely low metallicity at $[Fe/H] = -3.38 \pm 0.06$ (stat.) with a dispersion of $\sigma_{[Fe/h]} < 0.18$ at 95% confidence level (Nicolas F. Martin, Venn, et al. 2022). Additionaly, the observations yielded radial velocities which were in agreement with those predicted by STREAMFINDER, fully establishing the existence of the stream.

The distance was estimated by STREAMFINDER to be between 16 and 22 kpc, but was narrowed down to 18 kpc based on the average distance calculated by employing 7 blue horizontal branch stars with proper motions corresponding to members located in the same region. From the distance and the observed kinematics of 8 of the stars, an orbit was found going deep in the potential well of the Milky Way, with a pericenter of ~ 7 kpc and an apocenter of ~ 27 kpc, on an almost perpendicular plane with respect to the Milky Way disk. What is puzzeling, is that an apocenter of only ~ 27 kpc suggest it was accreted when the potential well of the Milky Way was shallower than it is now (Nicolas F. Martin, Venn, et al. 2022). But, the relatively short orbital time (<0.5 Gyrs) in combination with the baryonic or dark matter bumpiness of the halo creates a riddle as to how the structure of C-19 can still be so coherent today.
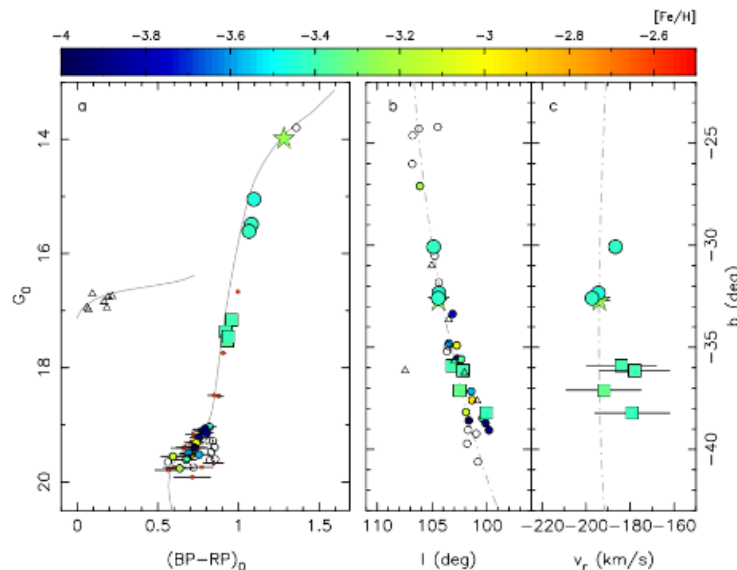

Figure 8: Properties of the C-19 member stars. **Left panel**: CMD of the stars color coded by their Pristine metallicity (small symbols) or spectroscopic metallicity (big symbols). Hollow circles indicates stars without metallicities but contain proper motions corresponding to C-19. The grey track is the 13 Gyr PARSEC isochrone of $[M/H] = -2.2$, callibrated at 18 kpc. **Mid panel:** distribution of C-19 stars in galactic coordinates. **Right panel:** radial velocity vs galactic latitude of the 8 stars with spectroscopic follow-ups. The grey track in the mid- and right-panel indicates the STREAMFINDER programmed orbit (Nicolas F. Martin, Venn, et al. 2022).

By selecting stars with proper motions similar to the STREAMFINDER selected stars, a CMD

of stars associated with the structure is made, see the left panel of figure 8. Do note, that the stars in the main-sequence turn off area contain no radial velocity information, and that some show a much more metal rich Pristine metallicity. In the color-magnitude diagram, a defined red giant branch extending into a main-sequence turnoff can be distinguished from a very blue horizontal branch. These CMD features are typical for old and metal-poor globular clusters. Additionaly, for 3 of the 8 stars high resolution spectra were gathered using the Gemini/GRACES setup providing high-precision chemical abundances. It was found that the [Na/Mg] ratios vary by a factor of 3 between these stars. A variation which is typical for ancient globular clusters, and rarely seen in halo stars or faint dwarf galaxies. Barium was measured in only one star, showing a near solar value of [Ba/Fe] also in line with a GC progenitor.

A follow up study was done using STREAMFINDER and VLT/UVES (Yuan et al. 2022), adding two new members and high quality spectra for 5 previously confirmed members (Nicolas F. Martin, Venn, et al. 2022). Both new members show metallicities similar to the mean metallicity of C-19 calculated previously ($[FeI/H]_{LTE} = -3.44 \pm 0.16$) & ($[FeI/H]_{LTE} = -3.54 \pm 0.11$). One of these stars is located $\sim 30°$ away from the bulk of C-19, but has proper-motions and a radial velocity corresponding nicely to the calculated orbit for that part of the sky (See figure 9). The other added member is found in the bulk and is also dynamically in agreement with the surrounding C-19 stars. The current stream has a width of $0.56° \pm 0.08°$, and based on the improved radial velocities of 9 members in the main body, it shows a velocity dispersion of $\sigma_v = 6.2^{+2.0}_{-1.4}$km/s (Yuan et al. 2022). This is where the C-19 progenitor stars to show less straightforward features. Both are larger than the usual width (~ 100pc) and radial velocity dispersion ($\sigma_{vr} < 1-3$km/s) measured in Globular clusters, and are more common among dwarf galaxies. In addition, N-body simulations also place C-19 structurally closer to the dwarf galaxies regime than globular cluster streams (Errani et al. 2022). With one exception, the tidal disruption of a simulated dark matter dominated globular cluster is able to reproduce the width and velocity dispersion of C-19. In this hypothesis, C-19 is ony part of a fully disrupted progenitor, where the other part is expected to hide behind the galactic plane.
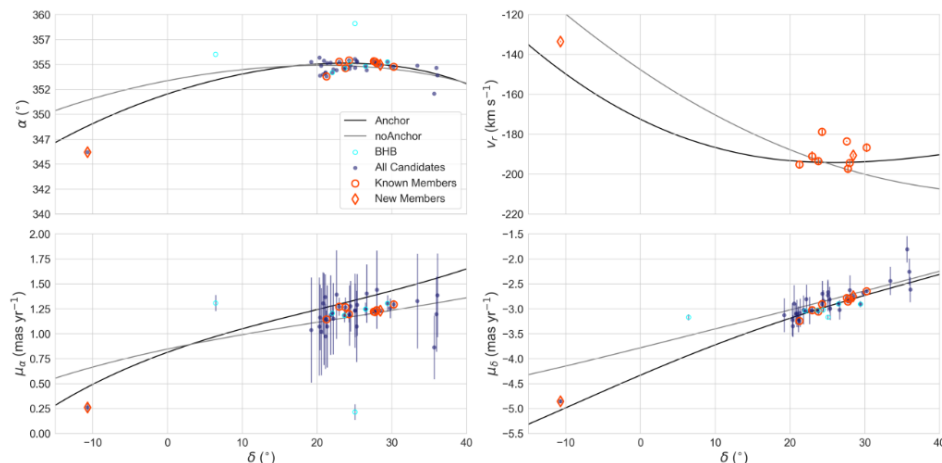

Figure 9: C-19 in 5D phase space. All members are shown with blue circles, except for the Blue Horizontal Branch members which are indicated with open cyan circles. The stars containing radial velocity measurements are annotated by orange circles. An anchored STREAMFINDER orbit at 18 kpc is shown in black, and an unanchored orbit is visible in grey (Yuan et al. 2022).

1.5   Structure of this Thesis

The aim off this thesis is to find stars moving across the sky akin to C-19, for several parts of its integrated orbit. Moreover, to look at the photometric metallicities, the CMD and the distance estimates of the selected stars to get a better grasp on potential members, and possibly a glimpse of C-19's progenitor. The various catalogs used to select stars with agreeing proper motions are described in section 2, together with the data sets employed for metallicity information. Section 3 will run through the chosen search areas, the cleaning cuts done on the data, and the two methods of dynamically selecting stars. The results are shown in section 4 and more critically discussed in section 5 together with the opportunities for follow up research.

2 Datasets

In the hunt for stars belonging to C-19, a variety of data sets are mined to gather stars, metal-licies and radial velocities. The two main catalogs for star selection are: the Reduced Proper Motion (RPM) halo catalog by Viswanathan et al. (2023) consisting of main sequence halo stars, and the Giant catalog by Foppen (2023) containing stars located in the Red Giant branch. Both use Gaia data release 3 (GDR3) as their parent sample, and both were matched with separate Pristine surveys to obtain metallicity information. These catalogs are adopted because they contain distance estimates more precise than Gaia DR3 for stars further out, and so can get rid of contamination stars which are beyond the parallax measurement reach but still too close to be C-19 members. Additionally, they provide insight on two tracks of the stellar evolution: the main sequence (turn off) and the Red Giant branch.
Nevertheless, the uncategorized GDR3 data is more suitable to look for densities along certain section of the orbit using LAAT (see section 3.3), since these regions either have no coverage for the RPM catalog in proper motion space (see section 2.2), or because the catalogs are simply too sparse to entertain LAAT.

The remainder of this section will first go over the main parent sample: Gaia Data Release 3 (section 2.1), then move on to the method of extracting photometric metallicities with the Pristine Survey (2.1.1), next it will briefly mention the radial velocity databases that are searched in (2.1.2), and finally it will elaborate on how the RPM (section 2.2) and Giant (section 2.3) catalogs are compiled.

2.1 Gaia DR3

Gaia was launched in December 2013 with the goal to produce the largest 3D sky map ever made, by providing accurate distance (parallax), proper motions, radial velocities, apparent magnitudes and colors (red and blue narrow-pass bands) for hundreds of millions of stars. The abundance of information gathered by Gaia has managed to shed light on the Milky Way's merger history and the discovery of a multitude of stellar streams. The latest Gaia data release, Gaia Data Realse 3 (GDR3) has been available since june 2022, and is the one adopted for this search (Gaia Collaboration, Vallenari, et al. 2023).
The telescope is located on an orbit around the L2 Lagrange point of the Sun and Earth-Moon system (Gaia Collaboration, Prusti, et al. 2016). The part of the sky observed by Gaia is determined by its scanning law, which is designed to maximize uniformity of measurements across the sky. However, Some parts are inevitably observed more than others, resulting in Gaia's scanning patterns manifesting in over densities visible in the all-sky-data (Gaia Collaboration, Prusti, et al. 2016).

2.1.1 The Pristine Survey

The Pristine survey was launched in 2017 (Starkenburg et al. 2017) as a way to efficiently gain metallicity information on a large scale for metal poor stars in the northern hemisphere. It takes advantage of the fact that the more metal poor stars show weaker Ca H&K absorption lines, and so can be set apart from more metal rich stars of the same temperature. This feature is visualised in figure 10 which shows the Ca H&K absorption lines for 4 stars with descending metallicities: [Fe/H] = 0.0 (red), -1.0 (orange), -2.0 (green), -3.0 (blue), and a metal absent star (black). The three panels refer to different locations (with varying temperatures and surface gravity's) on the giant branch, with the visible Ca H&K absorption lines at 3968.5 and 3933.7 Å.
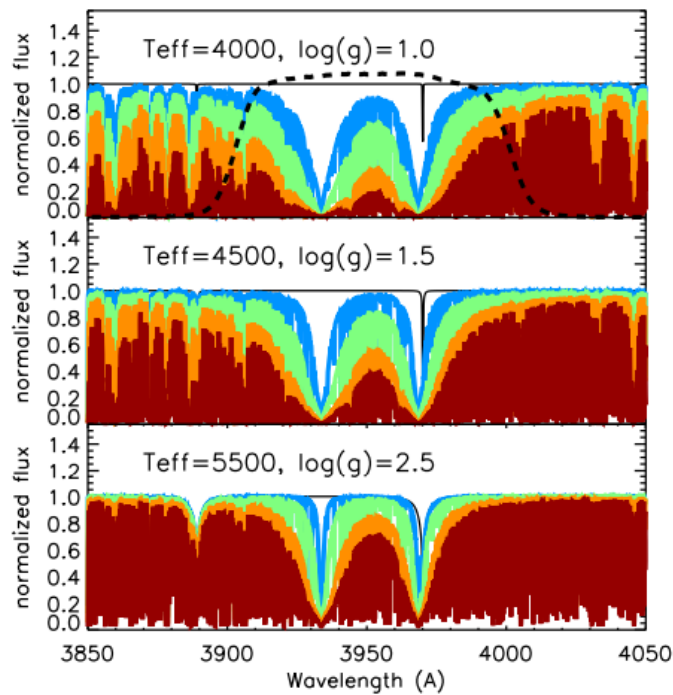
Figure 10: Synthetic spectra of stars with various metallicities: [Fe/H] = 0.0 (red), -1.0 (orange), -2.0 (green), -3.0 (blue) and a metal absent star (black) for 3 places on the giant branch corresponding to the different panels. The Ca H&K absoption lines are visible at 3968.5 and 3933.7 Å, and the narrow-band Pristine filter spanning acros it is indicated by the dashed black line (Starkenburg et al. 2017).

By comparing the relative flux in a narrow band filter (the 'CaHK filter') spanning across these strong absorption lines (shown as the black dashed line in figure 10) a metallicity value can be inferred. In part, this is allowed by the fact that the dependence of the absorption line strength on surface gravity in the metal-poor regime is much weaker then its dependence on either temperature or metallicity, and so can be ignored (Starkenburg et al. 2017). The assumption is that [Ca/Fe] linearly rises from 0.0 at [Fe/H] = 0.0, until +0.4 at [Fe/H] = -1, and remains there for lower metallicities. Ofcourse this means that stars with a discrepant [Ca/Fe] relation will have less reliable photometric metallicities. To callibrate the Pristine metallicities for more accuracy, the Ca H&K filter information is combined with the broad-band photometry of SDSS and SEGUE.
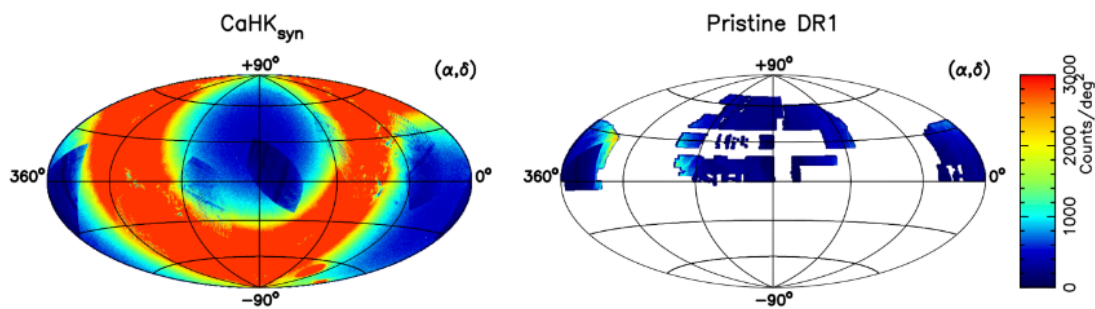

Figure 11: The pristine footprint shown for the Pristine-Gaia synthetic survey (left) and Pristine DR1 (right) (Nicolas F. Martin, Starkenburg, et al. 2023).

The star selection catalogs that will be described below (RPM and Giant), make use of separate Pristine surveys. The RPM catalog is matched with an (as of yet) unreleased updated DR1 Pristine survey, able to penetrate deep into the stellar halo and gather photometric metallicities for the faintest of stars. The footprint of this data set is shown in figure 11 on the right side. To ensure regular observations, it makes use of a bad weather program (no restrictions on observing conditions), resulting in a bad weather footprint.

The Giant catalog uses as its parent sample the Pristine-Gaia synthetic CaHK survey (Nicolas F. Martin, Starkenburg, et al. 2023), which uses the same principle described above, but instead integrates Gaia's narrow BP/RP filters under the curve of the CaHK filter. By design, this survey covers fairly bright stars since Gaia only includes BP/RP information for stars with sufficiently high signal-to-noise (S/N). It should be noted that without a proper quality cut, the uncertainties on the CaHKsyn magnitudes can be very large towards the blue end of the BP spectrum, caused by the lower S/N ratio. On the flip side, the footprint for this survey is an all sky one as shown on the left side of figure 11, with Gaia's scanning law being its only limiting factor.

Very recent high resolution spectroscopic follow ups (Viswanathan et al, upcoming) of some of the very metal poor Pristine Survey stars, allow for an indication of the relative accuracy of these photometric metallicity. Figure 14 shows the spectroscopic vs Pristine measurements for the Pristine-Gaia synthetic (left) and the Pristine DR1 (right) surveys.
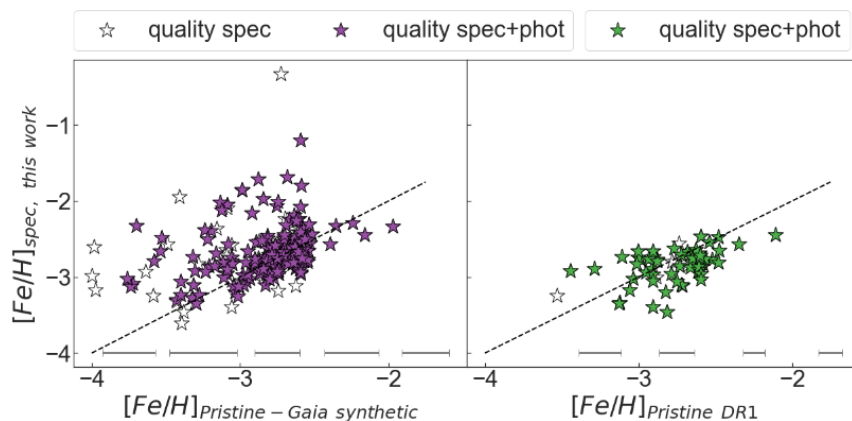

Figure 12: Comparison of photometric and spectroscopic metallicities for followed-up very metal poor Pristine stars. The left panel shows the Pristine Gaia-synthetic versus the spectroscopic metallicities derived from Calcium Triplet equivalent widths for the entire sample. The right panel shows the higher signal-to-noise Pristine CaHK magnitudes versus the spectroscopic follow-up metallicities. Median errors in bins of 0.5 dex in photometric [Fe/H] is shown in both panels (Viswanathan et al., upcomping).

The Pristine DR1 survey corresponds neatly to the spectroscopic measurements, but the Prisine-Gaia synthetic survey shows a tendency to underestimate the metallicity values. However, whatever is lost in accuracy is made up for in efficiency: narrow-band photometry can be obtained simultaneously for any star in the field of view rather than needing to narrow done on specific targets. Providing us with large scale metallicity information incredibly useful for search parties such as these.

2.1.2 Radial velocities

For radial velocities, the follow databases are looked in: LAMOST (DR8, VMP catalog), SDSS, SEGUE and Simbad.

2.2 Reduced Proper Motion catalog

A large portion of the main-sequence (MS) stars located in the more distant galactic halo are handicapped by their intrinsic faintness. Fortunately, MS stars share an almost linear relation in apparent magnitude and colour which allows us, together with their proper motion, to determine their absolute magnitude, and consequently, their distance. This process of using astrometry and photometry to calculate distances is described in (Koppelman, Helmer H. et al. 2021b) and applied by Viswanathan et al. (2023) to stars of Gaia DR3, garnering a catalog containing 47,650,376 provisional main sequence halo stars. The formula for the photometric distance is given by:

$$d_{phot} = 10^{\frac{m_{G,0} - M_{G,0} - 10}{5}} \tag{1}$$

Where $m_{G,0}$ and $M_{G,0}$ are the apparent and absolute Gaia G-magnitude after extinction correction, respectively. $M_{G,0}$ can be calculated using the following equation(s):

$$H_{G,0} = M_{G,0} - 5\log\frac{v_{tan}}{4.74057} \tag{2}$$

Where v_{tan} is the tangential velocity defined as:

$$v_{j,b} = 4.74057 \text{km/s} \left(\frac{\mu_j}{\text{mas/yr}}\right)\left(\frac{d}{\text{kpc}}\right) \tag{3}$$

(The factor of 4.74057 is due to the unit conversions to obtain the velocity in terms of km/s.) And $H_{G,0}$ is the (Gaia G-band) reduced proper motion, which can be determined using:

$$H_{G,0} = m_{G,0} - 5\log\mu - 10 \tag{4}$$

Where μ is the Gaia proper motion. The linear relation between $M_{G,0}$ and $m_{G,0}$ for MS stars allows us to combine equation 2 with 4 if we keep v_{tan} constant by selecting stars with similar tangential velocities. The reduced proper motion (RPM) versus colour diagram will then simply mimic the colour-absolute magnitude diagram for fixed v_{tan}. This feature is visualized by the isochrones (with constant v_{tan}) plotted in the RPM vs color diagram of figure 13.
The tangential velocities for disk stars are small, allowing for a clean halo star selection by keeping the high tangential velocity stars. The selected high v_{tan} stars lie within the grey dashed box shown in figure 13 with upper-bound $v_{tan} = 800$km/s and lower-bound $v_{tan} = 200$km/s, also corresponding to the blue and red isochrone, respectively. As a confirmation, the main-sequence tails of the isochrones nicely overlap the selected MS-halo-star area.
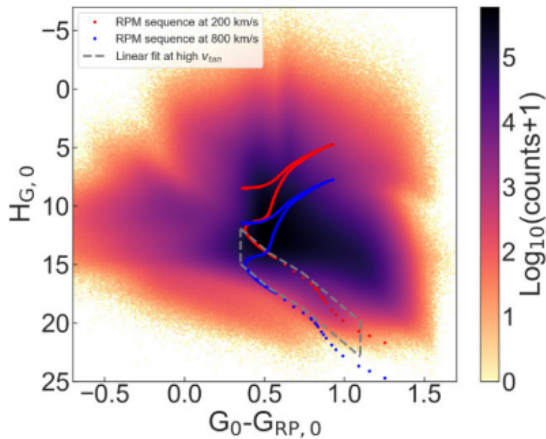
Figure 13: RPM diagram of the GDR3 parent sample. Two -1.6 [M/H] isochrones of 12 Gyr are plotted with $v_{tan} = 200$ km/s (red) and $v_{tan} = 800$ km/s (blue), the selected main sequence stars lie within the dashed gray box. The tails of the isochrones (the MS) neatly encloses the selected halo-star area (Viswanathan et al. 2023).
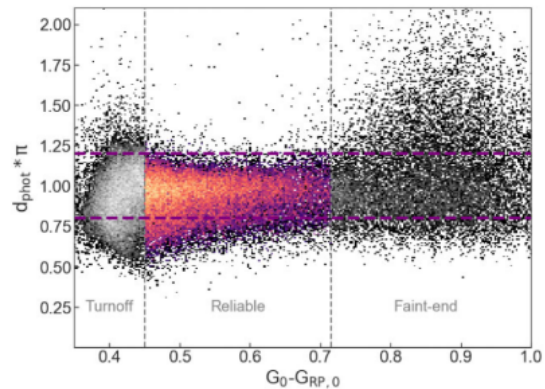
Figure 14: Comparison of the photometric distances of the RPM catalog to Gaia's good quality parallaxes, as a function of Gaia colour. The thick puruple dashed lines indicate the $\pm 10\%$ deviation of RPM with Gaia. The colour ranges are separated according to uncertainty, with the Reliable range containing a typical ($\sim 7\%$) deviation (Viswanathan et al. 2023).

The cleanness of the sample is further confirmed by the fact that intrinsically bright stars in the high v_{tan} area of the RPM diagram, would have a tangential velocity exceeding the finite escape velocity of our galaxy (Koppelman, Helmer H. et al. 2021a). Nonetheless, the sample is not a complete selection, since it will neglect halo stars which have high-line-of-sight velocities but low tangential velocities. As a reliability test, the photometric distances of the RPM-selected stars were compared to Gaia's good quality parallax measurement, the result of which are shown in figure 14 as a function of Gaia colour. The catalog is divided into a less reliable Turnoff section, a Reliable area with a typical deviation of ($\sim 7\%$) and the more unreliable Faint-end. Most of this specific search will take place in the Turnoff area.

For the purpose of C-19, the lower tangential velocity bound of $v_{tan} = 200$km/s had to be loosened to $v_{tan} = 180$km/s, to ensure coverage of C-19 in proper motion space. This does influence the cleanness of the catalog, as more giants with velocities high enough to escape the Galaxy will bleed into the sample.

2.3   Red Giant catalog

Moving over to stars further along in their stellar evolution, the Red Giant catalog makes use of a multi-step selection method following (Byström 2021 and Byström 2022), implemented on the Pristine-Gaia synthetic survey (Nicolas F. Martin, Starkenburg, et al. 2023).

This parent sample was trimmed down to a catalog containing 6,227,999 stars with a purity of 93.2% and completeness of 88.8%, using a training sample made from with the Sloan Digital Sky Survey (SDSS) (Blanton et al. 2017). The first trim was a color cut $(0.5 < (G_{BP} - G_{RP})_0 < 1.5)$ to get rid of stars which are too hot or too cold to show reliable CaH & K absorption lines (Nicolas F. Martin, Starkenburg, et al. 2023). Next was to get rid of any dwarfs by making a separation in the absolute CMD. The absolute magnitudes are calculated using the distances of stars with a parallax bigger than 0 $(\varpi > 0)$ and a fractional error smaller than 0.5 $(\frac{\sigma_\varpi}{\varpi} < 0.5)$ in order sweep away as many dwarfs as possible. Figure 15 shows the training sample stars in an absolute CMD. The stars above the black dashed line are assumed to be red giants and kept in the sample, stars below the line are discarded as dwarfs.
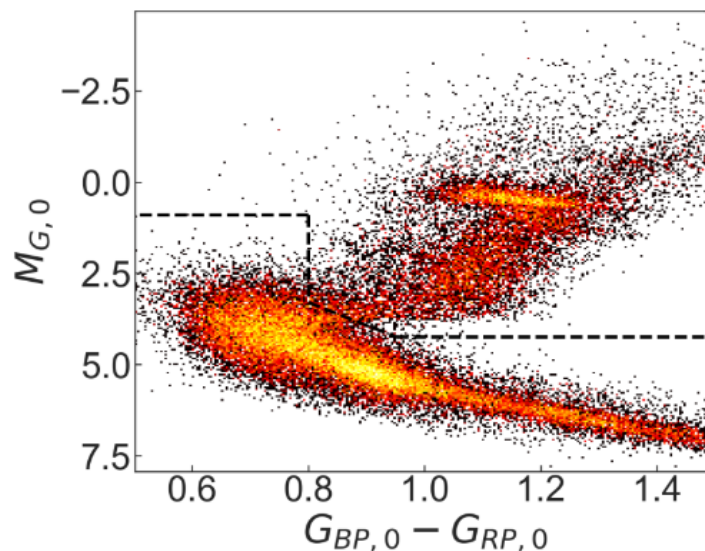

Figure 15: Absolute CMD containing training sample stars with non zero parallax and f < 0.5. The black dashed line indicates the assumed separation between the red giants (above) and the dwarfs (below). (Foppen 2023)

Stars with low quality parallax values (non zero or a big parallax error) are not shown in figure 15, but do remain in the sample since a low quality parallax indicates that the star is far away and hence in order to still be picked up by Gaia it needs to be a giant (Foppen 2023).

To shave off any Sub Giant Branch (SGB) stars, a color-metallicity cut $((G_{BP} - G_{RP})_0 > 0.14 \cdot [Fe/H] + 1.05)$ was implemented. This relational cut is based on the colors at the turning points of SGB into RGB for 9 different isochrones, covering a metalicity range between 0.0 dex and -2.0 dex with an assumed age of 10 Gyrs. Finally, the quality cuts recommended by Nicolas F. Martin, Starkenburg, et al. (2023) were implemented. This means allowing only stars with errors in the Ca H & K lines below 0.1 dex and getting rid of the unphysical peak at -4 in the metallicity grid with [Fe/H] > –3.999. But also filtering for variable stars by keeping P_{var}, the probability of it being a variable star, below 30%. Figure 16 shows the catalog data in a color magnitude diagrams for the various stages of the selection process.
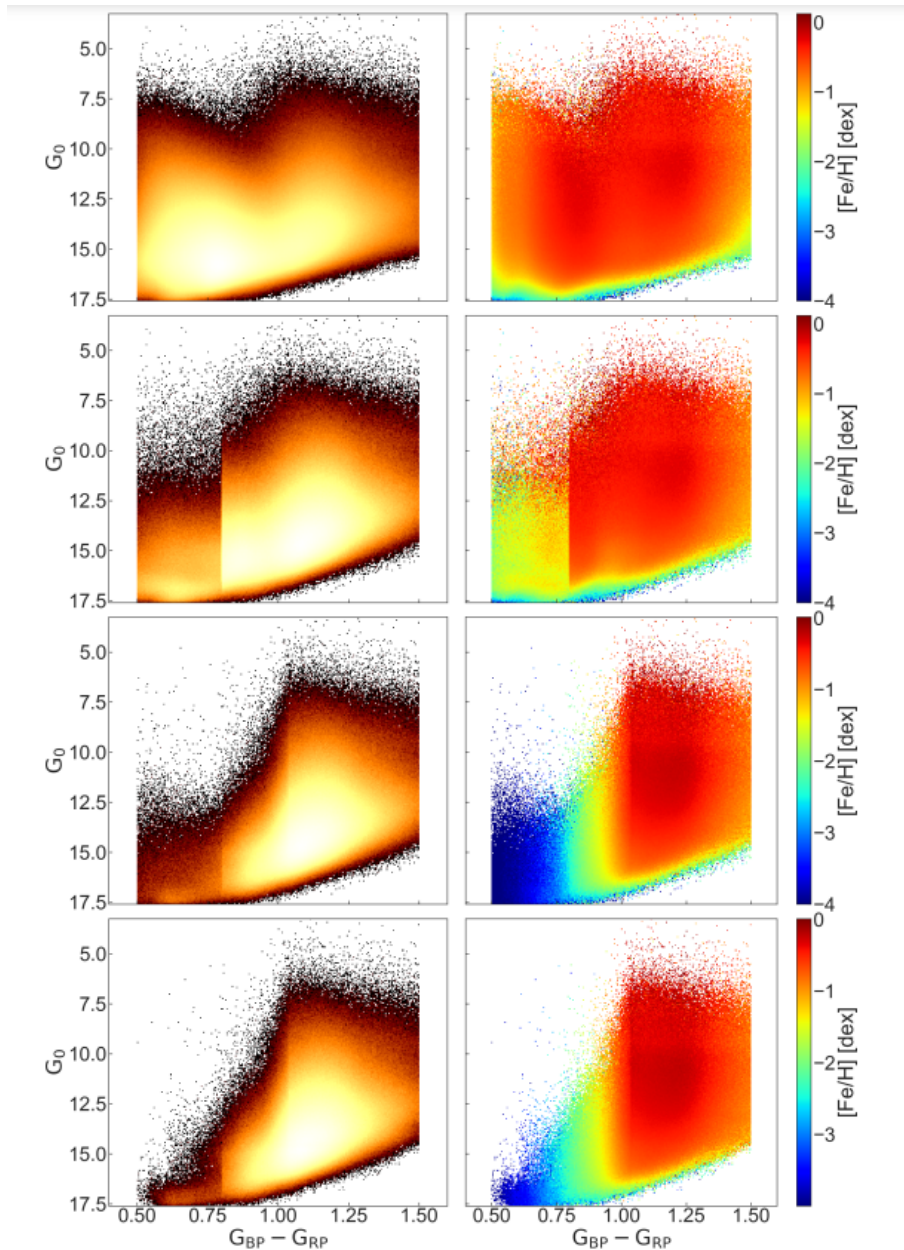
Figure 16: Left: the CMD's color-coded by the 10-logarithm of the density. Right: the CMD's color-coded by metallicity. Top row: raw data. Second row: data after the absolute CMD cut. Third row: data after the color-metallicity cut. Bottom row: data after the quality cut (Foppen 2023).

For these final stars (bottom row of figure 16), the distances were calculated using the general formula:

$$d = 10^{\frac{m-M}{5}+1} \tag{5}$$

Where d is the distance, m is the apparent magnitude and M is the absolute magnitude. The absolute magnitudes were gathered using isochrones from MESA Isochrones & Stellar Tracks (MIST) (Choi et al. 2016). For each star in the catalog, an isochrone is interpolated based on its metallicity. This isochrone provides an absolute magnitude corresponding to the observed color of the star, which can then be used to calculate the distance following formula 5.

3   Methods

In order to hunt for potential stars belonging to C-19, we use the orbit programmed by (Nicolas F. Martin, Venn, et al. 2022) following a procedure developed for STREAMFINDER detections. As described in the introduction, this procedure fits an N-body simulation to the observed configuration (taking into account its proper motions, coordinates and color) to deduce an orbit (Ibata et al. 2021). The orbit of C-19 is anchored at a heliocentric distance of 18 kpc based on the average distance calculated from 4 blue horizontal branch member stars using the relation between their absolute magnitude and (b-r) colour:

$$M_{g(BHB)} = 0.434 - 0.169(g-r) + 2.329(g-r)^2 + 20.449(g-r)^3 + 94.517(g-r)^4$$

(Taken from Deason et al. (2011)). It should be noted that this relation is only calibrated for systems with [Fe/H] > -2.3, and since the current mean metallicity of C-19 is estimated at [Fe/H] = 3.38 ± 0.06 (Yuan et al. 2022), it might not be perfectly suitable.
However, compared to a non anchored orbit, the 18 kpc orbit is marginally favored to minimize the difference between FeI and FeII iron abundances observed from 3 member stars. Additionally, the PARSEC isochrone ([M/H] = -2.2) of 13-Gyrs callibrated at a 18 kpc distance fits nicely through the members in a color magnitude diagram (see figure 8). Note again, however, the much higher discrepant metallicity used for the calibration.

Along this orbit, 3 separate parts were selected to search for more signs of C-19:

- **Part 1:** α (345, 360) and δ (-15, 40) **or** l (-22, -61) and b (52, 112)

- **Part 2:** α (345, 360) and δ (40, 60) **or** l (-18, -2) and b (101, 117)

- **Part 3:** α (260, 340) and δ (55, 80) **or** l (46, 19) and b (86, 117)

These parts are shown on the celestial sphere of figure 17 in violet, blue and green respectively. Part 1 is the violet patch surrounding C-19, part 2 is located on top of it in blue, and the tail end of the orbit positioned at the opposite side of the disk in green is part 3.
The orbit was integrated for ± 1 Gyrs where part 3 is located at the end of its integration, i.e. the place where C-19 is expected to be 1 Gyr in the future. Finding potential members in this area would mean C-19 is significantly more spread out then currently known, and would simultaneously be a confirmation of the validity of the integrated orbit. Additionally, part 3 is significantly more close to us then parts 1 & 2 (see the right side of figure 17) making it an interesting place to look around for more intrinsically faint MS or MSTO stars.
A practical reason to investigate both part 2 & 3 of the orbit, is because they are located close to the disk, an area that is computationally very taxing for STREAMFINDER and which it therefore leaves unexamined. Theoretically, the disk area is interesting because of the hypothesis that C-19 is only one part of the remains of a cold dark matter dominated globular cluster, the other part hiding behind the galactic plane (Errani et al. 2022). It should be noted however that these parts will suffer from more contamination as a consequence of looking through the disk at a lower angle. With increasing extinction the Pristine-Gaia synthetic metallicities becomes more unreliable/biased (Nicolas F. Martin, Starkenburg, et al. 2023).
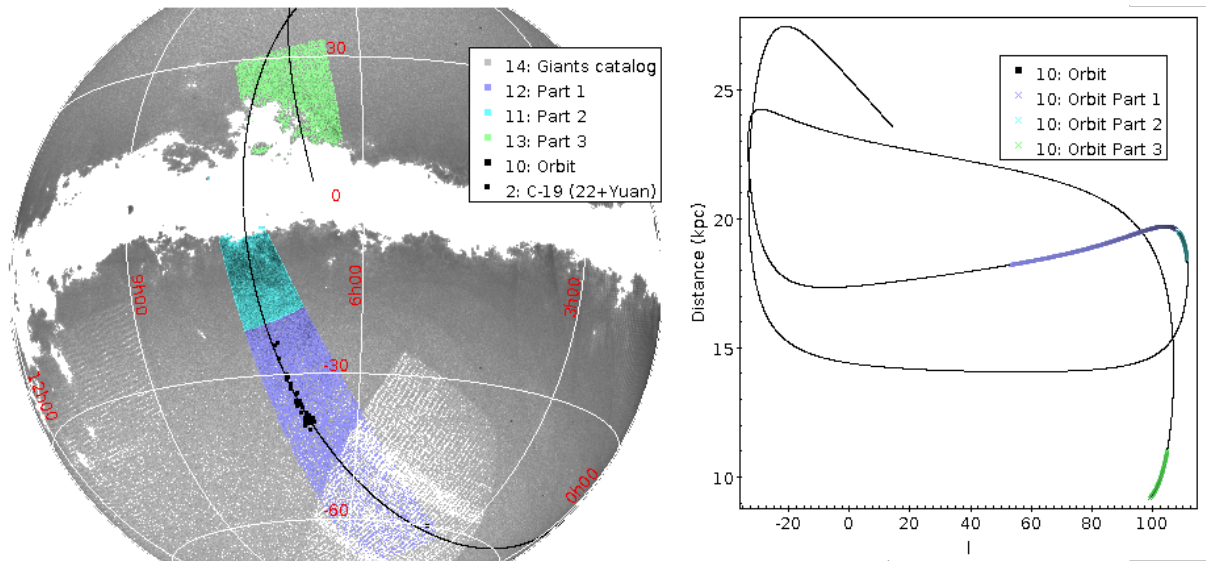
Figure 17: The left plot showcases the 3 selected parts along the orbit on the celestial sphere in galactic coordinates using data (grey) from the Giant catalog. The C-19 members (Yuan et al. 2022) and the STREAMFINDER orbit are visible in black. Part 1 surrounds C-19 in violet, above which part 2 is shown in blue and on the other side of the disk part 3 is visible in green. The right plot shows the full orbit in distance versus galactic longitude (l) where the 3 parts are highlighted by their corresponding colors. Note that part 3 is a lot nearer to us than parts 1 and 2.

Shown in figure 17 are also the current C-19 members (Yuan et al. 2022) in black, located in part 1. Note the single member at $\sim -60°$ (b). This area was chosen to get a better look at the surrounding stars of C-19, and to potentially fill the gap between the bulk and the sole member $\sim 15°$ further down. Do note however the blatantly apparent streaming patters in this part of the sky. The sparse patch below the bulk of C-19 could explain why STREAMFINDER did not find more members in this area already, and could potentially limit our search here as well.

The remainder of this section will go over the quality cuts done on all the catalogs (section 3.1) and the two separate types of dynamical selection that were implemented: looping of the data in 4 and 5D (section 3.2) and using the LAAT algorithm (section 3.3).

3.1   Pre-selection of stars

Before initiating the search, all 3 data sets undergo a filtering procedure to get rid of nearby noise and unreliable measurements. The quality cut is slightly different for each catalog, and since part 3 of the orbit is at a closer distance than parts 1 & 2, a separate CMD and distance cut is needed for this part. The trimming is tested by crossmatching the initial and final selection with the 38 C-19 members given by (Yuan et al. 2022). For all catalogs, every C-19 member matched before the quality, CMD and distance filtering, could be matched after the cleaning process too.

3.1.1   Quality cut

The **RPM catalog** is matched with Pristine (Nicolas F. Martin, Starkenburg, et al. 2023) and the corresponding quality cuts are made. This includes keeping the photometric metallicities between $[Fe/H]_{phot} < 0.0$ and $[Fe/H]_{phot} > -3.999$ to remove stars located at the edge of the metallicity grid (see Fig 12 of Nicolas F. Martin, Starkenburg, et al. (2023)). The photometric uncertainty was kept at $\delta[Fe/H]_{phot} < 0.5$ dex, and the variable stars were cut using $P_{var} < 0.3$. Additionally, to select stars which are (very) likely point sources and not extended objects, $CASU\_flag = -1$ or -2 is held. A general Gaia RUWE < 1.4 cut, and a more inclusive $|C^{\star}| < 3 \cdot \sigma_{C^{\star}}$ is also done.

The **Giant catalog** is prefiltered for quality, however, for this search a $P_{var} > 0.55$ was chosen instead of $P_{var} > 0.3$ in order to avoid the exclusion of the brightest C-19 member with $P_{var} = 0.53$. For bright stars (G < 13), the S/N of the Gaia magnitude is so high that even the slightest variation results in a high P_{var} value while not significantly affecting $[Fe/H]_{phot}$ (Nicolas F. Martin, Starkenburg, et al. 2023). Therefore it should be safe to be more lenient on the P_{var} cutoff in this catalog. Note however, that it might influence the purity and completeness values mentioned in section 2.3.

The **GDR3** data has a simple RUWE < 1.4 cut, and is not matched with Pristine or any other catalog to keep it as vast as possible. This does necessitate an additional dust extinction step in order to get the dereddened G, G_{BP} and G_{RP} band values. Using the dust maps by (Schlegel et al. 1998), the reddening value (ebv) at the galactic coordinates for each star is collected, which is then translated into the extinction value (A_v) for each bandpass:

$$A_G = R_V \cdot ebv \cdot 0.83627$$

$$A_{BP} = R_V \cdot ebv \cdot 1.08337$$

$$A_{RP} = R_V \cdot ebv \cdot 0.63439$$

Where R_v is taken as 3.1. To get the dereddened magnitudes, A_v is subtracted from the corresponding G magnitudes.

3.1.2   CMD cut

The bulk of C-19, which lies roughly in the middle of **parts 1 & 2**, is located at ~ 18kpc. like (Nicolas F. Martin, Venn, et al. (2022)) the most metal poor PARSEC isochrone is fitted through the member stars in a color magnitude diagram (see figure 18). The grey area around the isochrone indicate the stars that remain in the selection. The margins are kept fairly wide to accommodate for the variation in distance throughout these two parts of the orbit, the fact that an isochrone of $[Fe/H] = -2.2$ might be biased, and that the distance estimate of 18 kpc
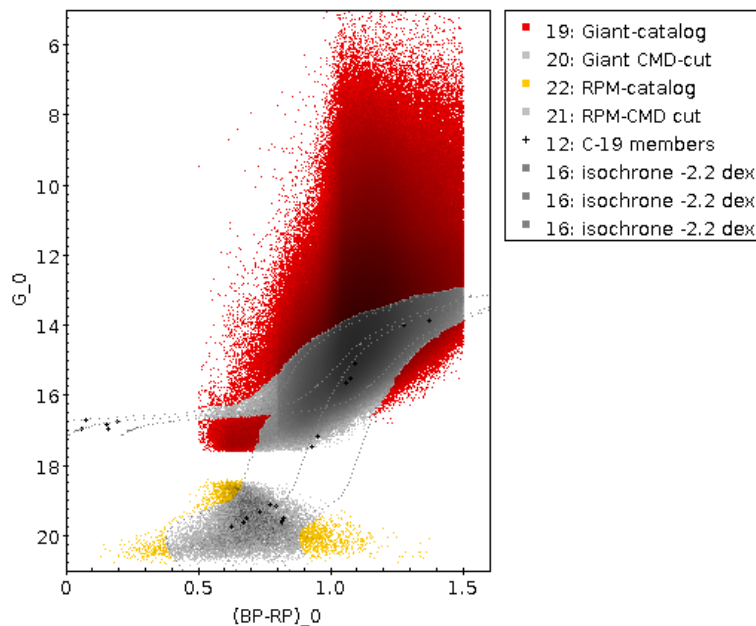
could be narrowed down more.

Figure 18: CMD of the Giant stars (red) and RPM stars (yellow) stars taken from parts 1 and 2. The member stars (Yuan et al. 2022) are shown in black with the 13 Gyr PARSEC isochrone of -2.2 dex fitted through them. The grey area indicates the area of stars selected through the CMD cut.

According to the integrated orbit, **part 3** ranges over a distance of $\sim 9\text{--}11\text{kpc}$ (see also figure 17). Using the general formula: $m - M = 5\log(\frac{d}{10\text{pc}})$, the isochrone offset at a centre distance of 10 kpc is given by 15 mag, with a lower and upper bound of ~ 14.8 mag and~ 15.2 mag corresponding to distances of 9 and 11 kpc, respectively. For this reason and the reasons mentioned above, the margins are again lenient, and a similar CMD cut to figure 18 is performed.

3.1.3   Distance cut

In order to get rid of any disk stars obstructing the view, a distance cut is done on all 3 data sets.

The **RPM catalog** was able to match with one potential C-19 member star but assigned it a much closer distance of ~ 12.4 kpc. The photometric distances for stars on the turn-off can be quite underestimated by the RPM sample, because it assumes the stars are on the main-sequence turn off when they could be on the sub-giant turn off instead which would mean they are further out than the photometric distance. However, the star does contain a much higher photometric metallicity ($[\text{Fe/H}] = -0.14$) than the mean metallicity of C-19, and so the distance estimate might not be wrong, the star might just be a non-member(or if it is a member, an indication of a dwarf galaxy progenitor). Nevertheless, since our selection is partly located in the turn off $G_0 - RP_0$ range where the photometric distances of the RPM catalog are slightly less reliable (see figure 14), a conservative distance cut of < 10 kpc is done for parts 1 & 2, and < 5 kpc for part 3.

The **Giant catalog** was able to match with the 5 brightest member stars. The color-metallicity cut in the Giant catalog to shave off any sub giant branch stars might explain why the two members on the lower end of the Giant branch (see figure 18) are not included in the catalog.

The estimated distances provided by the Giant catalog in comparison to the distances given by (Yuan et al. 2022) are shown in Table 1.

Giant catalog (kpc)	Yuan et al. 2022 (kpc)
27.9	20.5
23.0	20.5
-	19.8
32.3	21.2
-	20.7

Table 1: Distance comparison of 5 potential member stars between the Giants catalog and (Yuan et al. 2022)

Two stars are missing a distance input in the Giant catalog, presumably because they couldn't be interpolated into an isochrone for the distance calculation step (see section 2.3). Table 1 shows quite a big discrepancy between the distance estimates. For this reason the same distance cut of < 10 kpc for parts 1 & 2, and < 5 kpc for part 3 applies to this catalog. All stars without a distance input like the two members in table 1 are also included in the selection.

Unsurprisingly, the **GDR3** data matched with all 38 members. It is cut to include stars with a parallax $\varpi < 1$ for all three parts of the orbit. For parts 1 & 2, stars which have an admissible relational error $\frac{\varpi}{\sigma_\varpi} > 5$ and a distance $\frac{1}{\varpi} < 7$ kpc are discarded. This means that stars which have a distance closer than 7 kpc but an unreliable parallax are kept, and any star further out with an unreliable parallax too. The same is repeated for part 3 of the orbit but at 5 kpc instead of 7 kpc.

3.2   Dynamical selection in 4D and 5D

A star located on the orbit in proper motion space (μ_α vs μ_δ) will not necessarily be located near the orbit in coordinate space (α vs δ). It can be anywhere within the selected RA and DEC bounds of the data. In order to find stars corresponding to C-19's orbit in all 4 dimensions ($\alpha, \delta, \mu_\alpha, \mu_\delta$) a nested looping process is implemented using python (see appendix for code). This loop does the following **4D selection** for the data it is fed:

- A point along the integrated orbit is selected in coordinate space. All the stars of the dataset located within a radius of $5°$ around this orbit point (α, δ) are selected.

- The selected stars are transferred over to proper motion space. Of these stars, all located within a radius of $1°$ of the same orbit point (μ_α, μ_δ) but now in proper motion space, are selected.

- This process is repeated until the desired part of the orbit is fully iterated over. The doubly selected stars are thrown out.

This way, stars with coordinates close to the orbit, and proper motions similar to the expected motion of C-19 for each point along the orbit are gathered from the data. Since the Giant- and RPM-catalog are made using GDR3 data, the 4D selection is done using data from GDR3, and then cross matched with the two catalogs afterwards. To ease the computational effort, a bulk

cut is done in proper motion- and coordinate-space before feeding the stars to the nested loop. Only after the 4D selection is made, are any radial velocities collected from the databases mentioned in section 2.1.2. For stars located as far out as 18 kpc there is but a sparse amount of radial velocities available, to include them prior to the proper motion- and coordinate selection would extremely limit the sample of stars.

The stars with existing radial velocities are put back into the loop, but now a **5D selection** is possible:

- Stars within ±50km/s of the radial velocity (v_r) for each orbit point are selected.

Stars with deviating radial velocities are taken out of the sample. The results after applying this method are plotted in section 4.1 (including radial velocities) and 4.2 (excluding radial velocities).

3.3   Application of LAAT

The Locally Aligned Ant Technique (LAAT) offers another way of selecting stars located close together in multi-dimensional space. An example of its success, was its application on data of the Jhelum stellar stream, where it was able to expose two components of the stream moving with different proper motions (Awad et al. 2023).

The algorithm mimics the operation of Ant Colony Optimization (Dorigo et al. 2004) that disperses a number of virtual 'ants' in a dataset which randomly walk around and deposit 'pheromone' on the data points they visit. These pheromones contain an 'evaporation rate' that reduces the pheromone quantity over time. The ants are attracted to the pheromone and so form a positive feedback loop: points that are visited more frequently will accumulate more pheromone and will evaporate less quickly, thus attracting more ants. The algorithm can be set up in such a way that it focuses on different kinds of spacial structures within the data that forms it landscape. In our case (and similarly to (Awad et al. 2023)) during the walk, for any data points the ant passes by, the main direction along which data points are distributed in the neighbourhood is calculated. This is done by the principal component analysis (PCA) which is performed within neighborhoods of radius r centered at the location of the data point. The ants are programmed in such way they will then, 1) follow the dominant direction along which stars are distributed within the neighbourhood pinpointed by PCA, and 2) jump towards neighborhoods that have accumulated more pheromone as the walk is allowed to continue. This way, as the ants are re-dispersed over multiple iterations, the pheromone will start to accumulate on data points belonging to embedded structures, and evaporate on more randomly distributed regions.

For our search, the data landscape was either 2D coordinate space, or 4D coordinate and proper motions space. After running the data in LAAT, it can be plotted in its original form but now color-coded by the pheromone counts associated with each data point in order to highlight the embedded structures hidden in the data. The results are shown in section 4.4.

4 Results

We will now go over the stars that are selected based on 5D (section 4.1) and 4D (section 4.2) dynamics, and look at their photometric metallicities, distance estimates and place in the CMD. Additionally, section 4.3 will give a broader view of the (uncut) proper-motion space of part 1, and its metallicity distribution. Finally, selection 4.4 will go over the results of applying LAAT to the 3 search areas.

4.1 The 5D selection

After the cleaning and bulk proper motion cut for stars in parts 1-3, a total of 301,080 GDR3 stars went into the 4D selection process, with 67,194 stars making it out the other side. Unfortunately, only 954 stars had an existing radial velocity measurement, and after the final 5D selection a total of 84 stars remained. These 84 stars are made up of: 68 GDR3 stars, 15 Giants and 1 RPM MS-halo star.

In part 3 all radial velocities deviated more than 100 km/s from the orbit and were accordingly excluded. Note however, that any small deviation in the integrated orbit compared to the physical orbit can extrapolate the further we move away from the C-19 bulk. And because part 3 is located at the tail end of the orbit, a bigger error margin could be considered in future research.

For part 1 & 2 the radial velocity (v_r) vs declination (δ) are plotted in figure 19. The top panel shows to which data catalog the stars belong: black dots indicate the members from (Yuan et al. 2022), red triangles are what remains of the giant catalog, the yellow square was the only survivor of the RPM-catalog, and the small green points show the GDR3 data. Because the distance cut on the GDR3 data is less strict compared to the RPM- and Giants catalog, it is less certain these stars are actually as far out comparatively, and so they are indicated by smaller points. The second panel shows the correspondence between the orbit and stars for the otherwise invisible coordinate dimension α. The distance estimates are color-coded in the third panel, but as the distances of GDR3 data are unreliable and sometimes nonphysical (negative parallax), they are left grey. One of the giants did not contain a distance estimate and is also left grey. The bottom panel shows the metallicities of the catalog stars, including 7 GDR3 stars that could be matched with the Pristine-Gaia synthetic metallicity survey. The 4D phase space for the 5D selected stars is shown in figure 20, color coded by metallicity, and the color-magnitude-diagram is given in figure 21 together with the PARSEC isochrone ([M/H]= -2.2) fitted at 18 kpc. Note that ratio of Giants to Main-sequence (Turnoff) stars of this 5D selection does not represent the physical ratio because this far out a MS(TO) star is probably not intrinsically bright enough to have a reliable radial velocity measurement, and so is underrepresented.

Panel 4 shows a wide range in the metallicities among these stars, and all seem to fit equally will in 5D dynamics. That is, a lower metallicity does not necessarily mean a closer correspondence to the orbit. Starting with the metal poorest, two giants contain metallicity values of ([Fe/H] = –3.08 and [Fe/H] = –2.88) and can be recognized as the two blue triangles in the bottom panel of figure 19. But taking the cake is one of the GDR3 stars matched with Pristine containing a metallicity of ([Fe/H] = –3.59). This star will be referenced as 'star A' and is annotated in the bottom panel of figure 19 with a red 'A', and similarly in figure 21. The GDR3-5D looped stars are matched with the 'raw' Pristine catalog, meaning that no quality cuts are done. The extremely metal poor star has a P_{var} of 0.8, but an acceptable d_CaHK of 0.04.
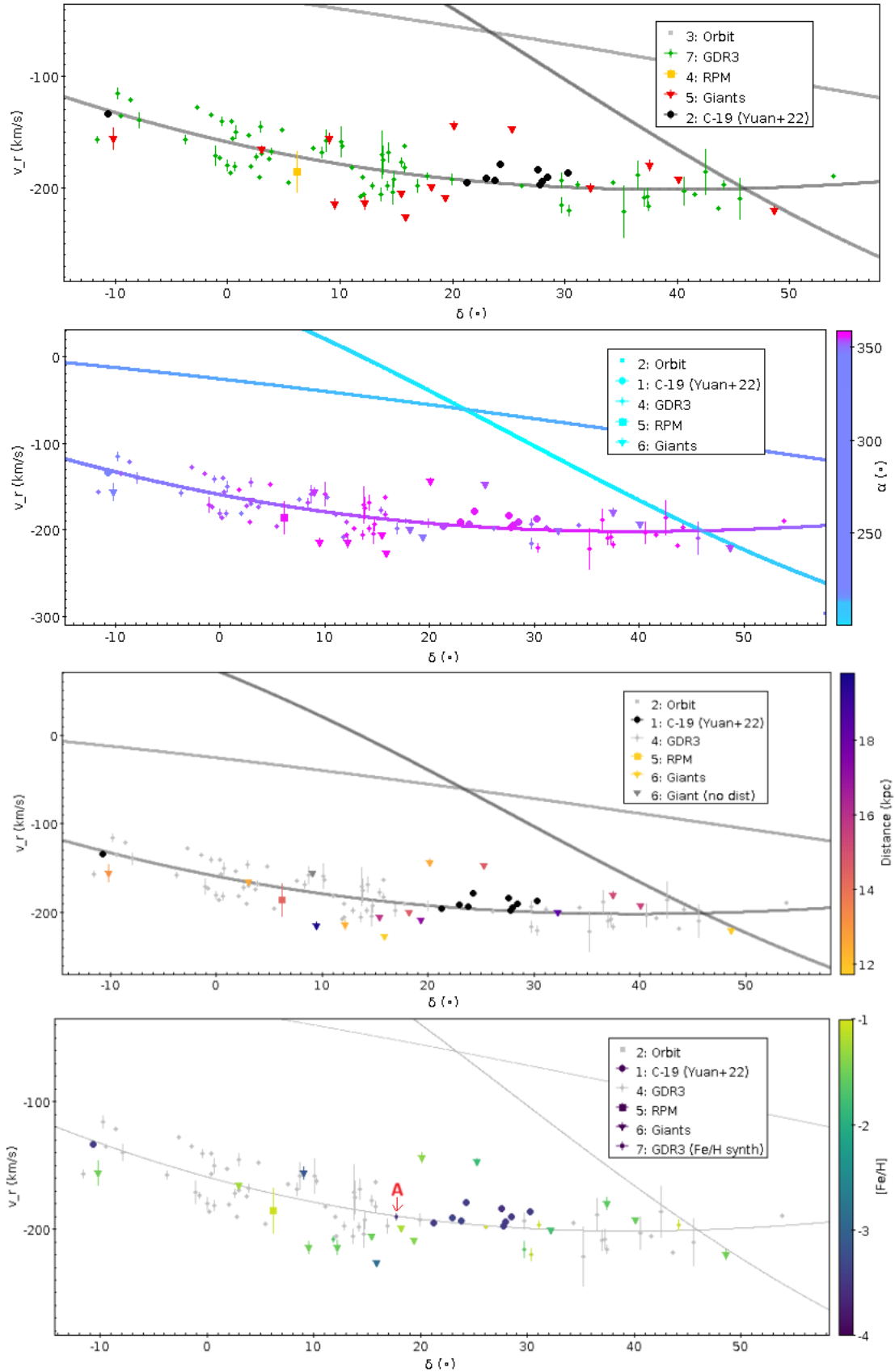
Figure 19: The radial velocity (v_r) versus the declination (δ) for parts 1&2 of the orbit. Showing stars from GDR3 (green dots), the giants catalog (Red triangles), the RPM catalog (yellow square) and the members containing RV measurements (black circles) of (Yuan et al. 2022). The second panel is color coded by α to indicate the correspondence between orbit and stars for this dimension. The third panel is color coded by distance, but note that the GDR3 data and one of the giants are not coded. The last panel shows the metallicies of all the stars, including 6 of the GDR3 stars. Star A is annotated by a red 'A' in the bottom panel.

Looking at its place in the CMD diagram (figure 21), there is an off chance it is a red giant with an exacerbated P_{var} value, but more likely would be an asymptotic giant branch (AGB) star at the end of its life experiencing periodic instabilities, explaining the high P_{var} value. Figure 20 shows it is located very close to the bulk in both coordinate- and proper motion-space, which combined with its extremely low metallicity and perfect fitting radial velocity makes star A a very likely C-19 member.

Comparatively, the two metal poor giants are located slightly further away from orbit in all 5 dimensions (see figure 19 and 20). The metal poorest of the two has no distance input, which is not necessarily a bad sign since two C-19 members were also lacking a distance value in the Giant catalog (see table 1). But, the slightly less metal poor giant is located at an estimated distance of 11.7 kpc and rather awkwardly placed in the CMD. Following to the PARSEC isochrone it could be an AGB star, but likewise a red giant shifted up by a distance modulus.

Moving on to the metal richer RPM star. It is located slightly more on the outskirts in proper motion space (with big error margins), but is a decent fit in coordinate space and has an good fitting radial velocity. For a presumably underestimated turn off star, a distance of 14.42 ± 2.0 is also not bad. Do note that the star is a lot bluer than the other turn off members (see figure 21).
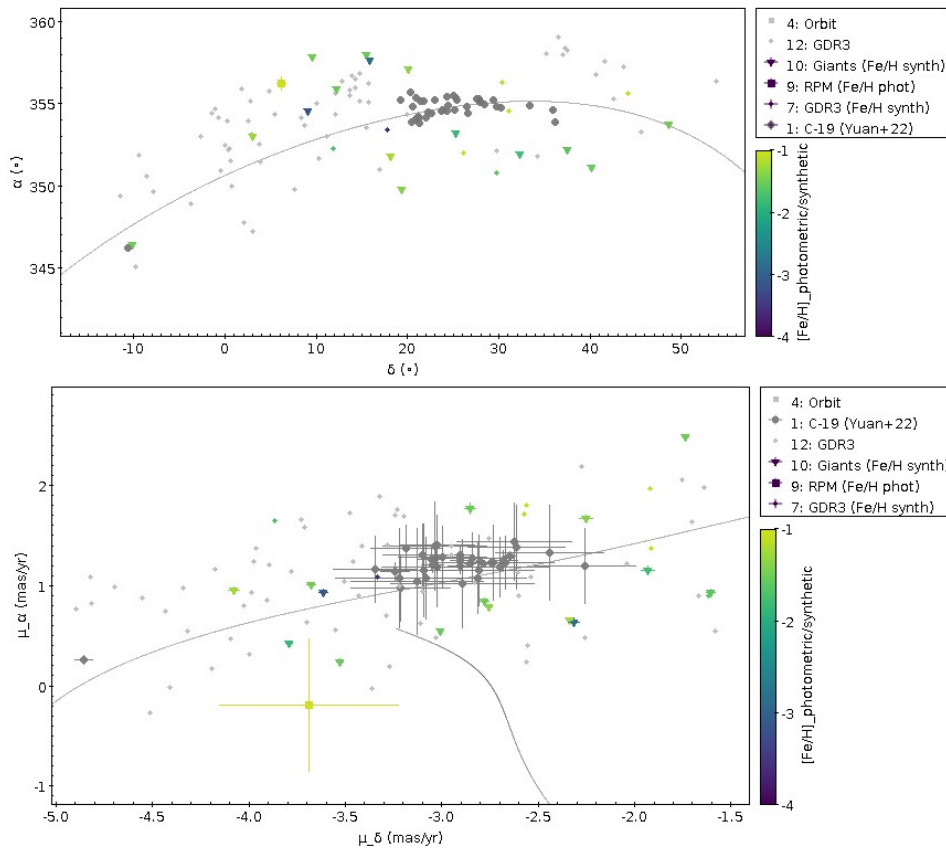

Figure 20: 4D phase space of the 5D selected stars of parts 1&2, color coded by metallicites. The giants are represented by triangles, the RPM star by a square, the GDR3 data by small dots and the C-19 members (Yuan et al. 2022) are shown in grey circles. The STREAMFINDER orbit ancored at 18 kpc is visible as a grey line. Star A is not annotated, but is visible very close to the bulk in both panels.

In fact, figure 21 shows a slightly more blue than red deviation in the GDR3 stars from to the isochrone. It could be the case that these bluer stars are located at a closer distance than 18

kpc, where the upward discrepancy with the isochrone callibrated for 18kpc shows this distance modulus. Another explanation is that, when following the general trend in isochrones (Bressan et al. 2012), a more metal poor isochrone would be bluer and brighter and so create a better fit for these stars. Interestingly, the 68 GDR3 stars also contained 4 Blue Horizontal Branch stars, annotated on the left side of figure 21 with blue arrows. One of them significantly extending the branch towards the red.
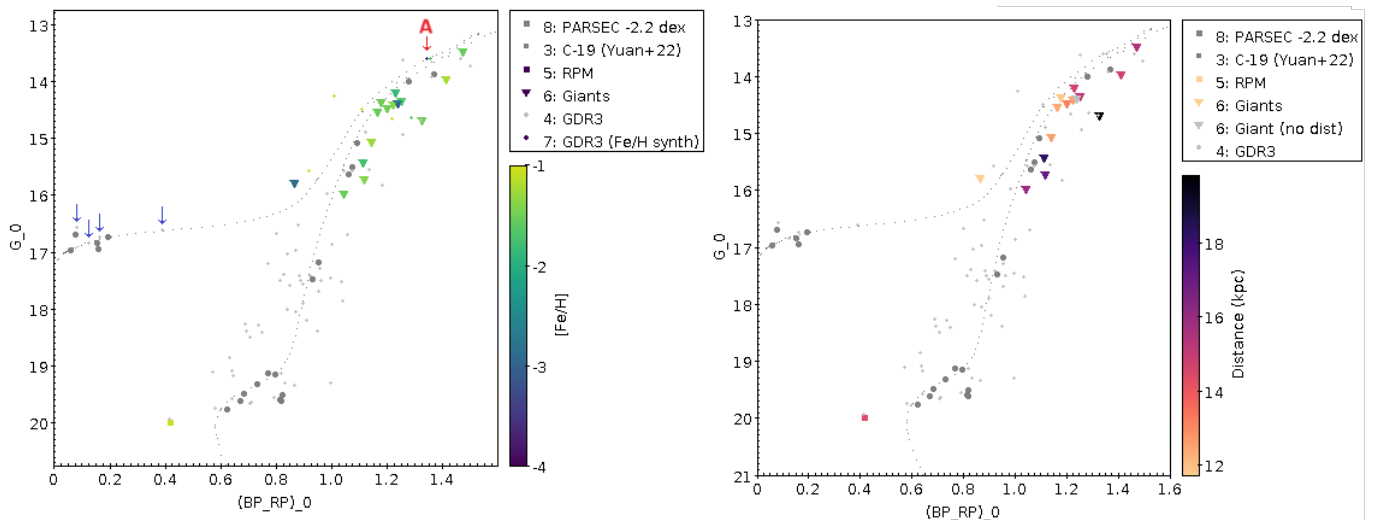

Figure 21: CMD of the 5D selected stars of parts 1&2. The giants are represented by triangles, the RPM star by a square, the GDR3 data by small light grey dots and the C-19 members (Yuan et al. 2022) are shown in dark grey circles. A 13-Gyr PARSEC isochrone of [M/H] = -2.2 calibrated for a 18 kpc distance is visible as a grey track. Stars on the left hand side are color coded by the metallicities, and on the right hand side by distances. In the left panel star A is annotated by a red 'A', and the 4 BHB stars are indicated by blue arrows.

Taken at face value, the 5D selected stars show a much wider metallicity and velocity dispersion then the current C-19 members (Yuan et al. 2022). Additionally, the stars trace out a broader sequence in the CMD.

4.2   The 4D selection

The process of 4D looping with pre-selected stars amounted to 66,240 GDR3 stars with unknown radial velocities. Even though the 5th dynamical dimension leaves us in the dark here, parameters like metallicity and distance available in the RPM and Giant catalogs can balance out an otherwise incomplete physical view. For this reason, but also to avoid indecipherable plots, the GDR3 stars are left out of the section. This does mean that the figures do not depict all the stars in the area which move desirably to C-19 in 4 dimensions.

Figure 22 shows the 4D selected Giant and RPM stars in coordinate space for parts 1&2 as triangles and crosses, respectively. The C-19 members (Yuan et al. 2022) are shown as black or grey circles, and the STREAMFINDER orbit is visible in grey. The top panel is color coded by metallicity. The RPM stars are tied to the pristine DR1 footprint which is why they do not extend along the whole orbit like the giants do.
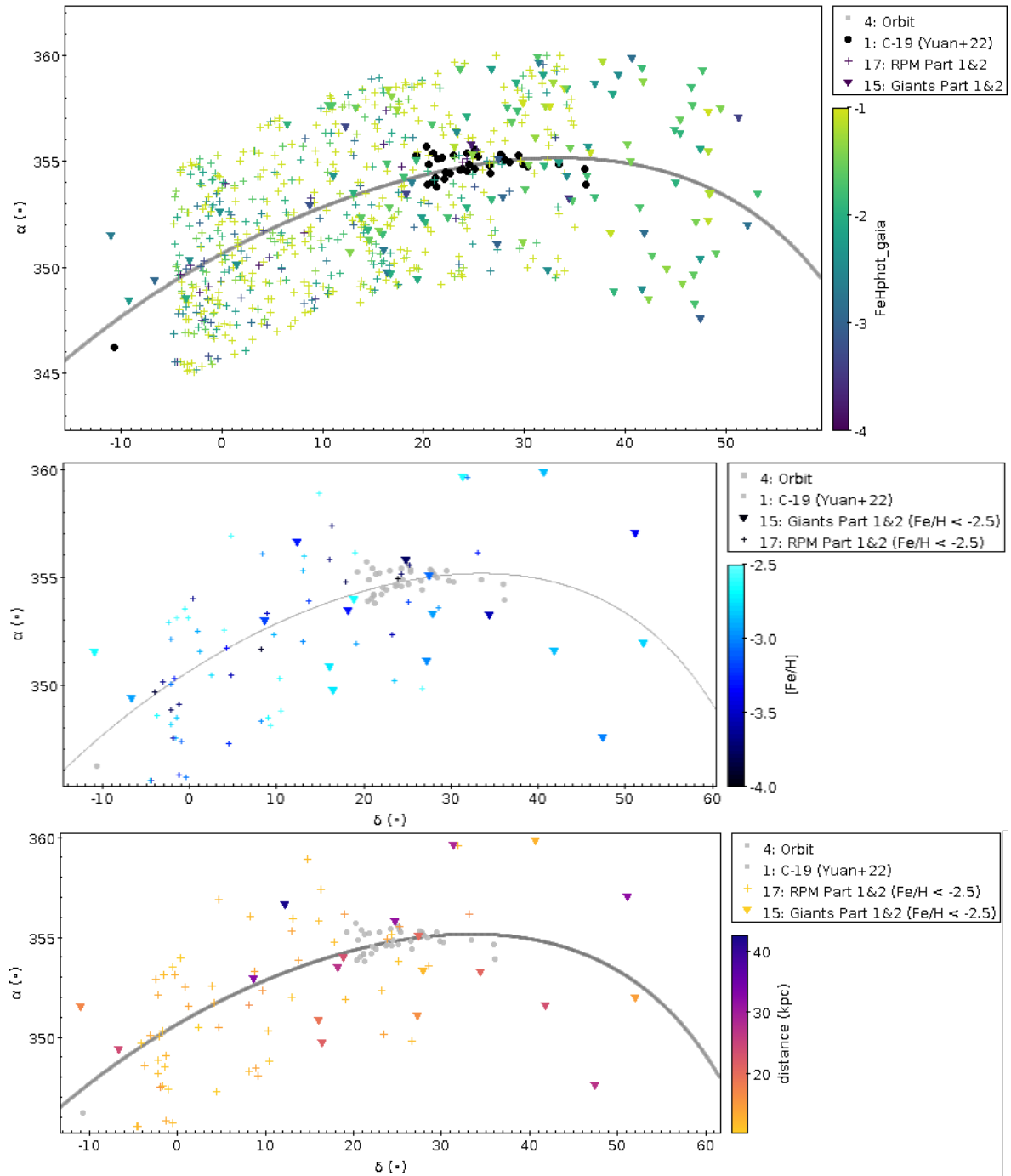
Figure 22: Coordinate space (RA α(°) vs DEC δ(°)) of the 4D selected stars of parts 1&2. The giants are shown in triangles, the RPM stars are visible as crosses and the C-19 stars (Yuan et al. 2022) as grey or black circles. The PARSEC orbit ancored at 18 kpc is shown in grey. **Top panel:** all 4D selected stars color coded by metallicity. **Middle panel:** 4D selected stars with [Fe/H] < -2.5, color-coded by metallicity. **Bottom panel:** 4D selected stars with [Fe/H] < -2.5, color coded by distance. The GDR3-4D selected stars are not shown.

Because we are blind in one extra dynamical dimension for this selection of stars, the attention is redirected to C-19's low metallicity feature instead, and consequently a metallicity cut of [Fe/H] < -2.5 is done for the 4D selected stars. This results in a remainder of 19 Giants (8 with [Fe/H]

< -3) and 59 RPM stars (27 with [Fe/H] < -3). The second and third panel of figure 22 show these metal poor stars of the sample, color-coded by metallicity and distance, respectively. The corresponding proper motion space is shown in Figure 23, color coded by metallicity, and the bottom panel is coded by declination (δ) to visualize the likeness between orbit and stars for this dimension. The CMD is shown in Figure 24 color coded by metallicity (left) and distance (right) together with the C-19 members and PARSEC isochrone in grey.
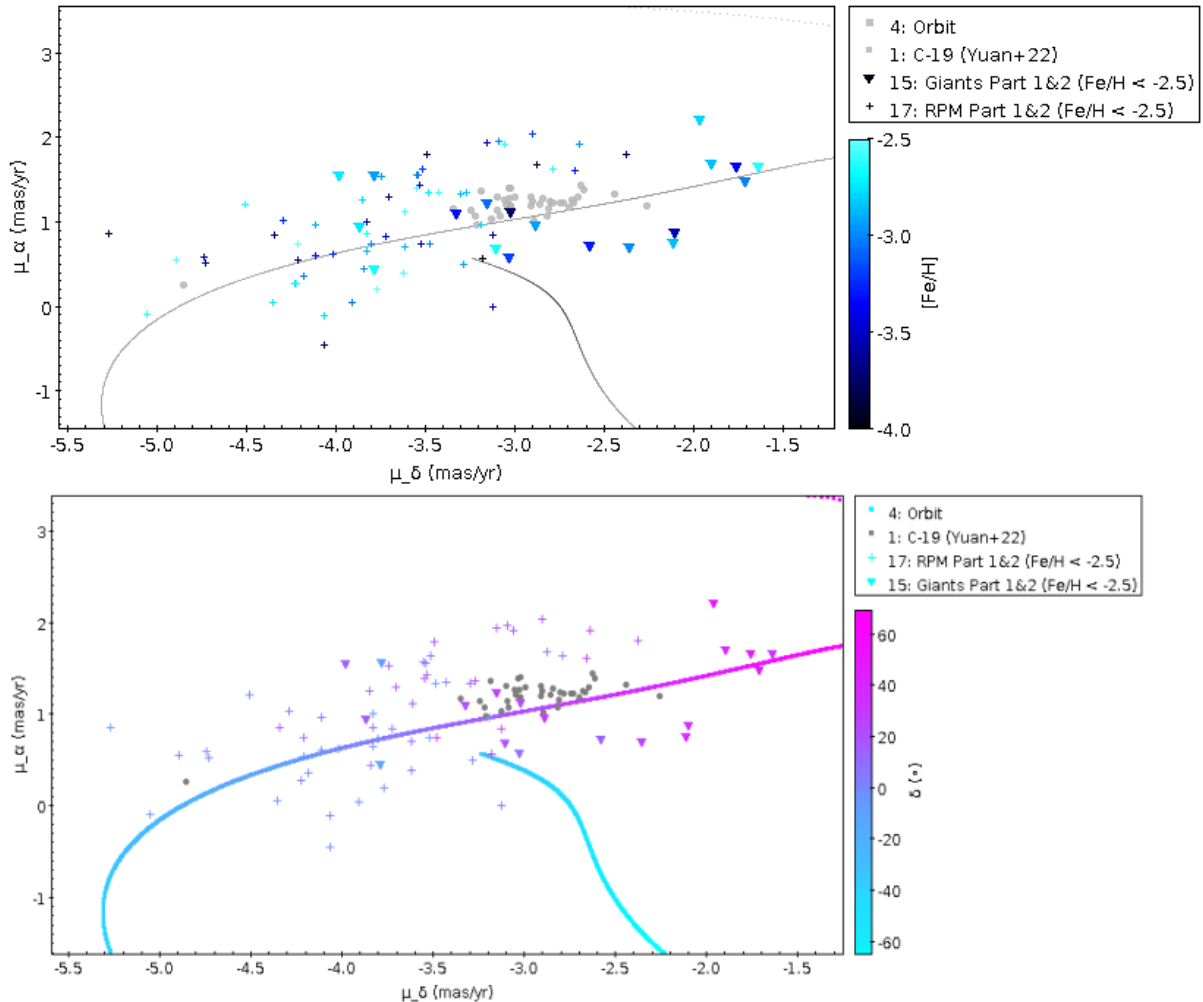
Figure 23: Proper motion space (μ_α vs μ_δ) of the 4D selected stars with [Fe/H] < -2.5 of parts 1&2. The giants are shown in triangles, the RPM stars are visible as crosses and the C-19 stars (Yuan et al. 2022) as grey or black circles. The STREAMFINDER orbit ancored at 18 kpc is shown in grey or color-coded. The top panel is color-coded by metallicity and the bottom panel by $\delta(°)$ to show the correspondence between orbit and stars for this coordinate dimension in proper motion space.

Visible in the bottom panel of figure 22 and the right panel of figure 24, is that the Giants show a significantly bigger distance distribution then the 5D selection, some estimated to be more than 40 kpc away. Additionally, because we are not bound by radial velocity measurements, many more RPM stars survived the selection.

Strikingly, the mid panel of Figure 22 shows an extremely metal poor giant ([Fe/H] = -3.76) located right in the bulk of C-19, visible as the nearly black triangle. This star can also be spotted in the top panel of figure 23, again, right in the bulk but now in proper motions space.

The giant has an estimated distance of 31.4 kpc, which, considering a C-19 member got estimated at 32.3 (see table 1), is not bad. Additionally, the CMD shows it fits in perfectly with the C-19 members shown in grey (see nearly black triangle in the left panel of figure 24). All things considered, it is an interesting potential member that could benefit from a spectroscopic follow-up. Similarly, three slightly more metal rich giants are also visible inside the bulk, both in coordinate and proper motion space, which also show acceptable distance estimates (bottom row of figure 22). The RPM sample shows many more [Fe/H] < -3.5 stars, three of which are located right in the bulk close to the extremely metal poor giant in coordinate space (see the three nearly black crosses next to the giant in panel 2 figure 22). The proper motion space also shows a few extremely metal poor RPM stars very close to, or on top of, the C-19 orbit.

However, generally speaking, all 4D selected stars do not show a more favorable correspondence to the orbit for the metal poorer stars. That is, just like the 5D selection, the stars show a wide, and seemingly random metallicity distribution (see specifically the top panel of figure 22). And again, the stars trace out a wider sequence in the CMD (figure 24).
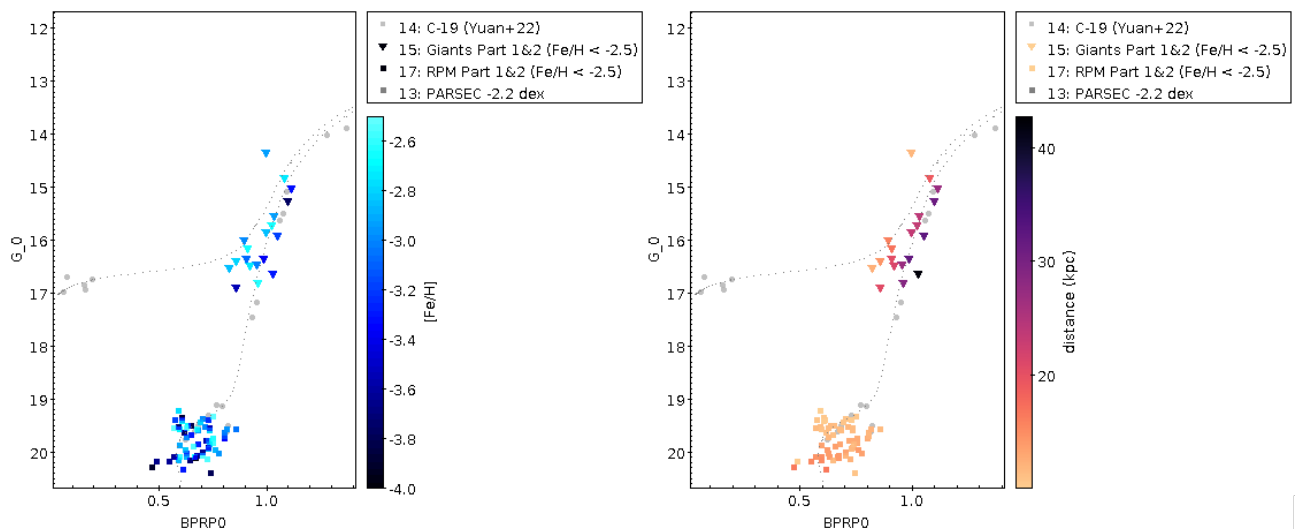

Figure 24: CMD of the 4D selected stars with [Fe/H] < -2.5 of parts 1&2. The giants are represented by triangles, the RPM star by squares and the C-19 members (Yuan et al. 2022) are shown in grey circles. A 13-Gyr PARSEC isochrone of [M/H] = -2.2 calibrated for a 18 kpc distance is visible as a grey track. Stars on the left hand side are color coded by the metallicities, and on the right hand side by distances.

4.3 Densities in proper motion

This section takes on more of a bird's eye view of the stars positioned in the area that contains C-19 (part 1). Figure 25 shows the proper motion space for all the stars taken from part 1 (RA(345° - 360°) DEC(-15° - 40°)) after the quality, distance and CMD are cut, but before any looping or proper motion trimming is done. The top panel contains stars from the Giant catalog, color coded by metallicity, and the bottom panel shows the GDR3 data stars, uncoded. The black circles and line indicate the C-19 members (Yuan et al. 2022) together with the STREAMFINDER orbit of part 1. The RPM data is not shown due to the large gap in coverage near $\mu_\alpha = \mu_\delta = 0$ stemming from the catalogs sole selection of stars with high v_{tan}.

The giants in the top panel showcase two different density regions centred roughly at ($\mu_\alpha = -2, \mu_\delta = -2$) and ($\mu_\alpha = 1.5, \mu_\delta = -2$) with contrasting metallicity distributions. To make sure this is not simply a side effect of how the giant catalog was assembled, GDR3 stars of the same coordinate region are plotted in proper motion space in the bottom panel. The two density regions are also visible in the GDR3 data, with the addition of a density near $\mu_\alpha = \mu_\delta = 0$. Most likely, this addition can be explained as a contamination of disk stars orbiting in a similar motion as the sun around the centre of the galaxy (see section 1.1), and thus seeming stationary (proper motions close to 0) from Gaia's point of view. The giant catalog being cut more strictly on distance would explain why the disk stars did not bleed into the giant sample.
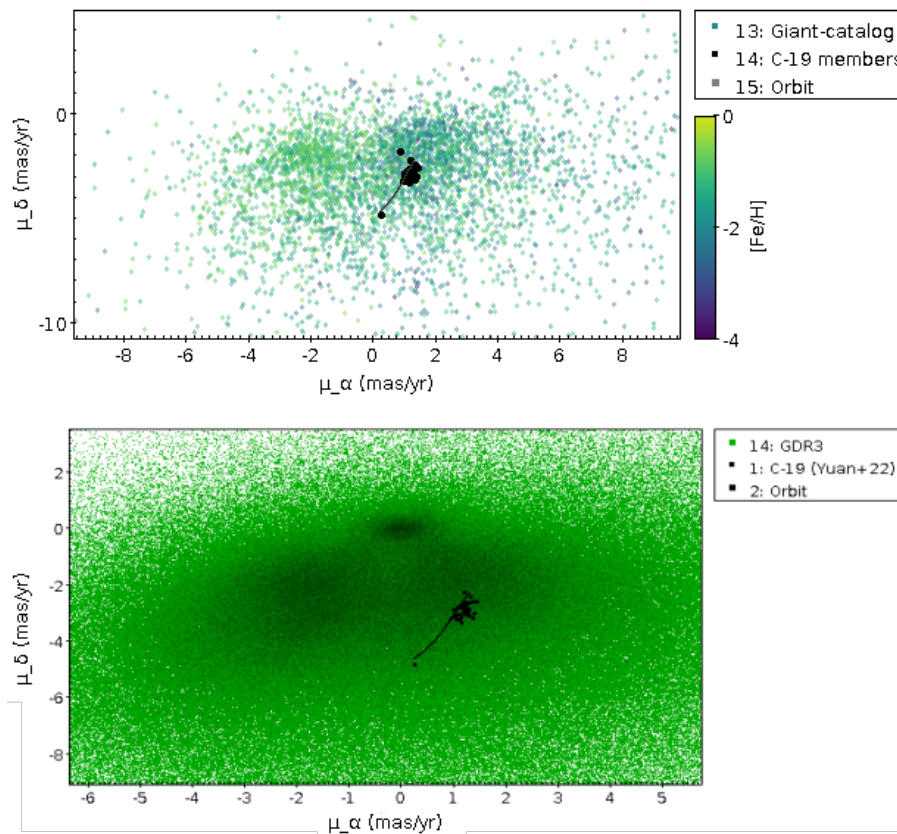

Figure 25: Proper motions space (μ_α vs μ_δ) of the giant-catalog (top panel) and the GDR3 data (bottom panel) for unlooped stars of part 1 (RA(345° - 360°) DEC(-15° - 40°)) coordinate space. The giants are color coded by their metallicity and shown with a slight translucency to elucidate the two density regions. The GDR3 stars are shown in green, uncoded and without translucency. The C-19 members (Yuan et al. 2022 are indicated by black circles together with the STREAMFINDER orbit of part 1, also in black.

In a somewhat unconventional plot, figure 26 displays the same giant sample but now in [Fe/H] versus (μ_α or μ_δ) to visualize the metallicty distribution of any set of stars moving in a similar α- or δ-proper motion. The giants are color coded by their distance estimates and the C-19 members (Yuan et al. 2022) are shown in black using their Pristine photometric metallicities. The two density regions shown in figure 25 are visible in the left panel of figure 26 as two broad vertical streaks. The density located behind C-19 in figure 25 starts to dense up at a lower metallicity, and with a far longer extension into the metal poor realm compared to the left sided density streak. Additionally, the distance coding indicates a trend of more distant stars roughly below [Fe/H] < -1.7. As described in section 2.3, the distances of the giants are estimated by obtaining absolute magnitudes through the color of a star and the interpolated isochrones from their metallicity, which together with the apparent magnitudes can calculate the distance modulus. It could be that the apparent trend of increasing distance with decreasing metallicity in figure 26 is merely manufactured by the Giants catalog. Even so, the apparent magnitude still plays an independent role. Looking closely at the more metal poor stars in figure 26 shows that not all have a far out distance estimate, disfavouring the possibility of this processing side-effect.
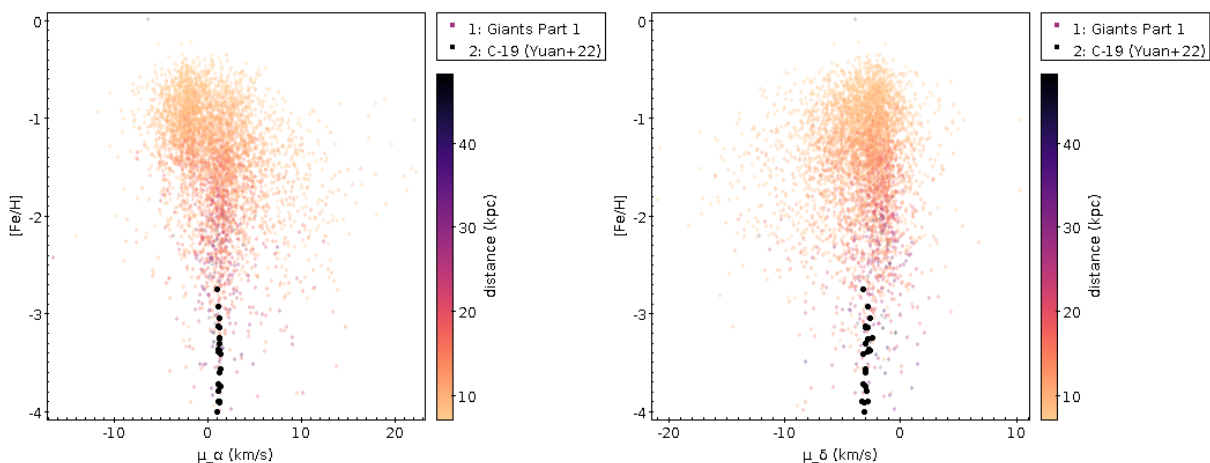

Figure 26: Pristine metallicity versus propermotion: [Fe/H] vs μ_α (left), and [Fe/H] vs μ_δ (right). The giants of part 1 are shown color coded by distance, and the C-19 members (Yuan et al. 2022) with Pristine photometric metallicities are indicated with black circles, not color coded.

Keep in mind that figure 25 and 26 do not tell us anything about the radial velocity of these stars. Nonetheless, they showcases a large distribution of stars in part 1 of the sky that have (fitting) far out distance estimates, moving with analogous μ_α (and slightly more disperse μ_δ) to C-19, but a surprisingly large, not globular cluster like, metallicity distribution.

4.4   LAAT Results

Running LAAT in 4D is an way to highlight densities existing both in coordinate- and proper motion space. Unfortunately, doing this for the Giant and RPM catalogs resulted in an under-saturated gradient. Even using just 2D coordinates, they struggled to pick up a density near C-19 and in some cases were heavily disadvantaged by Gaia's scanning patterns. The unfiltered GDR3 data on the other hand was able to highlight an over density near the bulk of C-19. This is possibly explained by the much stricter proper motion cut which is allowed by the dense GDR3 data, but not the sparser RPM and Giant catalogs. These data sets needed a wider, and hence possibly noisier, proper motion cut in order to allow meaningful runs with LAAT.

For the GDR3 data in part 1, a narrow elliptical proper motion cut is done around the bulk of C-19. This selection is run in LAAT in 2D coordinate space. The left plot of figure 27 shows the data color coded by the resulting pheromone counts. Behind the C-19 members (the black dots), a region with high pheromone count is visible. In order to account for misleading densities lighting up in LAAT manufactured by Gaia's scanning patterns, the right side of figure 27 shows the same data points but weighed instead by the observation count (astrometric_n_good_obs_al).
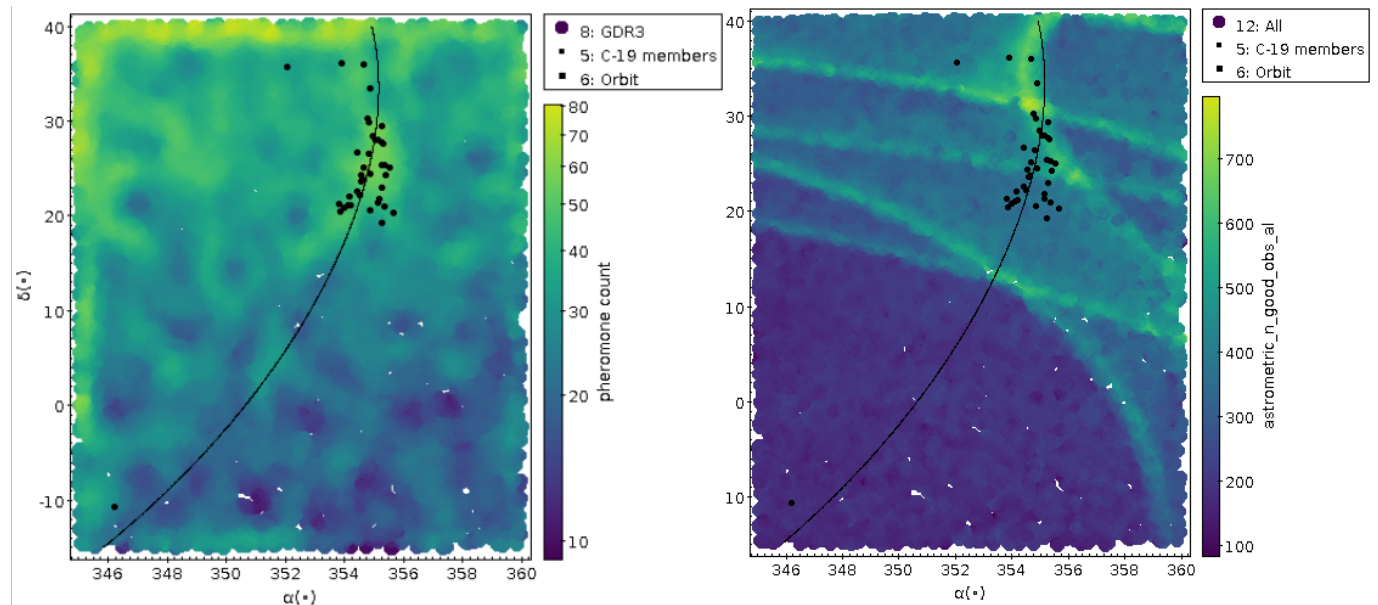

Figure 27: The GDR3 data (with strict proper motion cut) for part 1 in coordinate space ($\delta(°)$ vs $\alpha(°)$) color coded on the left by the pheromone count, and on the right by Gaia's observation count. The C-19 members (Yuan et al. 2022) and the STREAMFINDER orbit are shown in black. Behind the members LAAT lights up in a banana shaped cloud.

There are some clear scanning patterns sweeping over C-19, however, LAAT explores a region below the C-19 bulk with a more homogeneous scanning pattern which should leave LAAT unaffected in that area. In the over density of the GDR3 data, all 37 of the (bulk) member stars were able to be crossmatched. The cloud was selected and fed back into LAAT, this time in 4D ($\alpha, \delta, \mu_\alpha, \mu_\delta$) in an attempt to highlight any embedded structures. Figure 28 shows the stars color coded by their 4D LAAT run pheromone count on the left, and with an enforced pheromone threshold on the right.
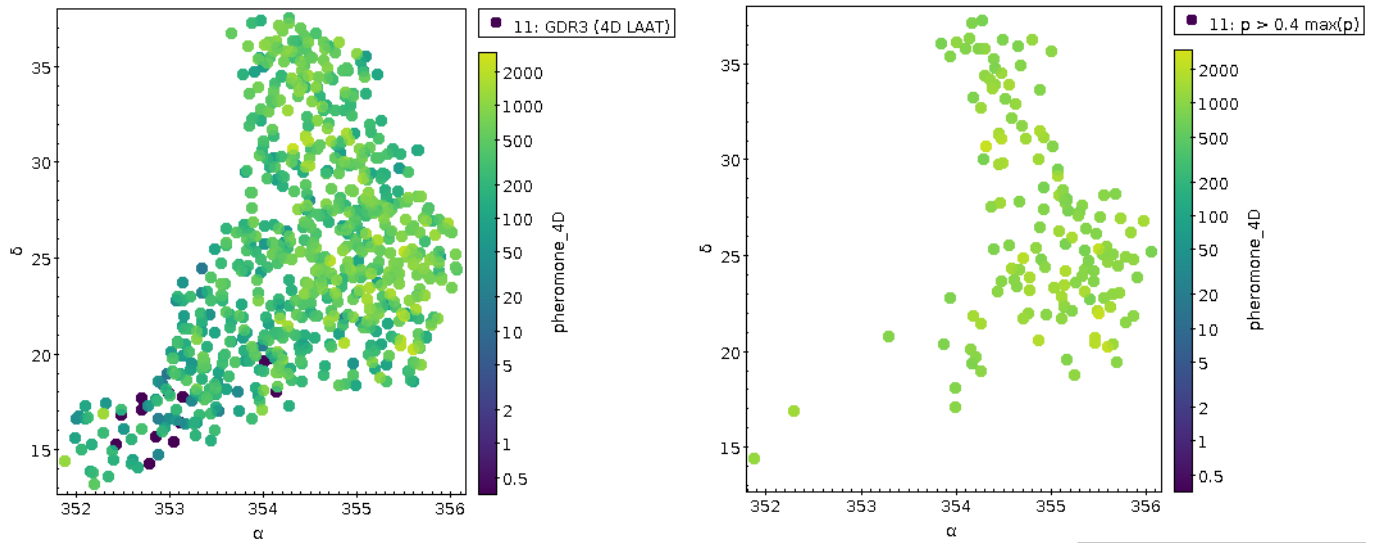
Figure 28: The GDR3 data surrounding the C-19 members (with high 2D pheromone counts) plotted in coordinate space. The left plot is color coded by the pheromone count after running the data in 4 dimensions in LAAT, the right plot shows the same data but now with an enforced pheromone threshold (pheromone $> 0.4 \cdot$ max(pheromone)).

The data points containing high pheromone counts shown on the right side of figure 28 do highlight a vertical downward stream in $(35 > \delta > 20)$ and $(355.5 > \alpha > 354)$, but this seems to correspond to the scanning pattern in that same region visible on the right side of figure 27. The proper-motion space (not shown here) of the same data color coded by the 4D pheromone count showed no coherence and was randomly distributed. After the 4D LAAT run, only 34 of the member stars could be crossmatched. However, the 3 members that were lost, are from the Main-Sequence Turn off area and contained no radial velocity or metallicity information. As described in section 1.4, these stars were selected based on their coordinates and propermotions. And so, it is worth considering that the loss of these members is not a failure on the algorithm's end but rather an indication that the missing stars could be non-members.

For parts 2 and 3, a narrow cut around the orbit in proper motion space was done for the GDR3 data, which was fed in 2D coordinate space to LAAT. Nothing significant was found around these parts, but the runs were extremely saturated by the increased density of the disk and so were deemed inconclusive. If there were any (thin) streams in these parts, LAAT was not able to set it apart from the overwhelming amount of stars surrounding it.

5   Discussion

The main question that remains for both the $4D(\alpha, \delta, \mu_\alpha, \mu_\delta)$ and $5D(\alpha, \delta, \mu_\alpha, \mu_\delta, v_r)$ selected stars, is if the wide metallicity and velocity dispersion shown in these samples are indicators that C-19 is dynamically hotter, and chemically more disperse than previously thought, or conversely, if it indicates these sample stars simply do not belong to C-19. For the 4D selected stars, radial velocities are necessary, and for the 5D selection spectroscopic metallicities and chemical abundances would help answer this question. The lack of radial velocities available in our distant search area (only 954 out of 67,194 stars) also inhibits any wider 5D contextual view. This means that is is hard to compare the stars found in the 5D selection with the movements of the surrounding stars, because we simply lack the access to this information.

The 6th dimension, distance, is represented in this thesis by fuzzy estimates rather than concrete measurements. Even though a lot of nearby contamination could be cut away for the Giant and RPM catalogs, the distance estimates of the final selection of stars cannot be used in a very meaningful way. Moreover, the 68 GDR3 stars of the 5D selection could only afford a loose distance cut and so the chance of noisy stars within this sample is also higher. Although, from the bottom panel of figure 25 it is clear that by selecting stars within $1°$ of the orbit in proper motion space, the density of disk stars moving similarly to the sun are at least avoided.

Another missing piece of the puzzle is that at a distance of 18 kpc the more numerous, but intrinsically fainter main sequence stars seem to be hidden behind GDR3's ~ 22 G magnitude barrier (see figures 21 and 24). As a consequence, our access is limited to stars situated in the maturer end of their stellar evolution.

It is also worth emphasising, that all the metallicities used in this thesis are photometric, and especially at the metal-poor end the Giants made from the Pristine-Gaia synthetic survey start to become more unreliable (see figure 12). All of this to say, that follow-up research would be good. The section below will go over the opportunities this thesis offers to do just that.

5.1   Future research

When it comes to an improved distance calculation, there are 5 new suitable stars at our disposal. The current distance estimate of 18 kpc originates from Blue Horizontal Branch stars selected for their coordinate position and proper motion (Nicolas F. Martin, Venn, et al. 2022). From the 5D selection process 4 new BHB members are found (see figure 21), which can be used as standard candles fitting with C-19 in 5 dimensions instead of just 4. Additionally, the variable AGB star (star A), offers the same employability if it is carbon rich (Lee 2023). Aside from fitting with C-19 in 5 dimension, star A has the additional extremely low metallicity of [Fe/H] = -3.59 working in its favour. Promisingly, one of the current members(C-19_346.2010.70) has slightly enhanced C/H of 0.3 dex (Yuan et al. 2022).

The 84 5D-selection stars form interesting candidates to look for chemical abundances, and to gather high resolution spectroscopic metallicities in order to obtain a more reliable view on the metallicity dispersion of these stars shown in figure 19. In addition, the chemical abundances would create the opportunity to look at any chemical tracers visible in the sample stars, which could allow us to distinguish (different) stellar populations.

In a like manner, the 4D selected stars would benefit from radial velocity measurements. In this case the 8 Giants and 27 RPM stars with [Fe/H] < -3 form interesting candidates, especially the extremely metal poor ones located right in the bulk of C-19 in 4 dimensions (as described in section 4.2).

To garner more potential members, it could be good to conduct a 'low quality' search. Meaning that the Pristine stars with [Fe/H] < -4 which are excluded for this thesis, are included. And to

put a less strict cutoff on the P_{var} parameter, which is motivated by the fact that star A was only found because it was matched with the uncut Pristine survey.

6   Conclusion

This thesis looked at several parts along the programmed orbit of C-19, and selected stars based on a 4D or 5D dynamical correspondence. Taken at face value, the wide (photometric) metallicity dispersion shown in figures 19 & 26, and the dynamical hotness visible in figures 19 & 20 point towards a dwarf galaxy or nuclear star cluster progenitor. Additionaly, the widened CMD (figure 21) disfavors the identity of a globular cluster progenitor. However, at this stage, these hints of a non-GC progenitor are less of a conclusive statement and more so an invitation to do follow up research. Figure 22 shows many extremely metal poor candidates, but radial velocities are needed for these 4D selected stars. Similarly, the 5D selected stars would benefit from more robust spectroscopic metallicties and chemical abundances to get a clearer view on C-19's chemical footprint, and to help decide which of the candidate stars can actually be grouped together chemically. What can be said with more certainty, is that star A is very likely a member of C-19 based on 5D dynamics and a (synthetic) metallicity of $[Fe/H] = 3.59$. It has a high probability of being a variable star ($P_{var} = 0.8$) and together with its location in the CMD, it could very well be an AGB star nearing the end of its life. This means it can be used, together with the 4 newly found BHB stars, for a more reliable distance estimate of C-19. Where the previous distance was based on BHB stars with a correspondence to C-19 in 4D, this would add the certainty of the 5th radial velocity dimension for these 4 BHBs and 1 AGBs, and the low metallicity of star A as the cherry on top.

7 Acknowledgements

Nearing the end of this thesis I began to struggle a little bit. But nonetheless, my supervisor Else Starkenburg continued to help me with patience and kindness, for which I am very grateful. She made this project a lot of fun and I am happy I got the opportunity to take a look inside the realm of Galactic Archaeology from her. This project made me gain a genuine appreciation for all the efforts being done in the Kapteyn Institute like the Pristine Survey, STREAMFINDER, but also after being confronted by the lack of radial velocities available, WEAVE. The small scientist inside of me now understands the significance of these projects in a more personal way after being put face to face with the observational limits of astronomy today through this project. But also makes me excited for what is to come.

I would like to thank my PhD supervisors : Akshara Viswanathan and Petra Awad. Both have provided incredible guidance in different areas. Akshara helped me find radial velocities, crossmatched the RPM catalog with Pristine, but also loosened the tangential velocity cut on her RPM catalog. She consistently managed to make time for me, and our meetings always helped me get a step further.
Petra helped me setting up LAAT, but also provided code for the dust extinction of the GDR3 data. She explained everything in a very calm and clear manner without fail, making the adoption of an algorithm like LAAT suprisingly smooth.

Last but not least, I want to thank topcat for being a steady partner in crime. It might seem weird to show appreciation for a (presumably) inanimate piece of code that won't be able to appreciate my appreciation, but nevertheless I can't fight the human urge to shout my love into the void. Thank you topcat.

8 Appendix

The following is the python code used for the 4D and 5D selection process:

```
pip install haversine
import vaex
import numpy as np
from haversine import haversine, Unit

# import stars
df_s = vaex.read_csv("star_filename.csv")


# import orbit
df_orbit = vaex.open("orbit_filename.fits")

# select part 1 & 2 of orbit:
#df_o = df_orbit[(df_orbit.ra < 360) & (df_orbit.ra > 345) &
(df_orbit.dec < 60) & (df_orbit.dec > -15)]

# select part 2 of orbit:
#df_o = df_orbit[(df_orbit.ra < 360) & (df_orbit.ra > 345) &
(df_orbit.dec < 60) & (df_orbit.dec > 40)]

# select part 3 of orbit:
#df_o = df_orbit[(df_orbit.dis < 11)]

# convert relevant columns to numpy arrays:

# orbit:
# -180 for haversine formula (180,180) instead of (0,360)
ra_o = df_o.ra.to_numpy() - 180
dec_o = df_o.dec.to_numpy()
pmra_o = df_o.mu_r.to_numpy()
pmdec_o = df_o.mu_d.to_numpy()
rv_o = df_o.vh.to_numpy()

# stars:
# -180 for haversine formula (180,180) instead of (0,360)
ra_s = df_s.ra.to_numpy() - 180
dec_s = df_s.dec.to_numpy()
pmra_s = df_s.pmra.to_numpy()
pmdec_s = df_s.pmdec.to_numpy()
rv_s = df_s.RV.to_numpy()

# relevant columns of stars (not used in loop):
BPRP0_s = df_s.bp_rp.to_numpy()
G_0_s = df_s.phot_g_mean_mag.to_numpy()
parallax_s = df_s.parallax.to_numpy()
```

```python
RPlx_s = df_s.parallax_over_error.to_numpy()
source_id_s = df_s.source_id.to_numpy()
rv_e_s = df_s.e_RV.to_numpy()

#index list of selected stars:
index = []

# loop over orbit:
for i in range(0,len(ra_o)):

    # loop over stars:
    for j in range(0,len(ra_s)):

        # distance between star and orbitpoint in coordinate space:
        dist_cord = haversine( (dec_o[i], ra_o[i]),  (dec_s[j],
        ra_s[j]), unit=Unit.DEGREES )
        # distance between star and orbitpoint in propermotion space:
        dist_pm = haversine( (pmdec_o[i], pmra_o[i]),
        (pmdec_s[j], pmra_s[j]), unit=Unit.DEGREES )

        # set radius limit:
        r_cord = 5 # degrees
        r_pm = 1 # degrees
        rv_min = rv_o[i] - 50 # km/s
        rv_max = rv_o[i] + 50 # km/s

        # select stars within radius
        if dist_cord < r_cord and dist_pm < r_pm and rv_s[j] > rv_min
        and rv_s[j] < rv_max:
            # add their index to list
            index.append(j)

# get rid of double items in selected stars list:

index_unique = np.unique(index)

# create list containing coordinates and propermotions of selected stars:
ra = []
dec = []
pmra = []
pmdec = []
RV = []

# other relevant info:
BPRP0 = []
G_0 = []
parallax = []
RPlx = []
source_id = []
```

```
RV_e = []

# fill lists above:
for i in index_unique:
    ra.append(ra_s[i] + 180) # add 180 back to regain original value
    dec.append(dec_s[i])
    pmra.append(pmra_s[i])
    pmdec.append(pmdec_s[i])
    RV.append(rv_s[i])
    BPRP0.append(BPRP0_s[i])
    G_0.append(G_0_s[i])
    parallax.append(parallax_s[i])
    RPlx.append(RPlx_s[i])
    source_id.append(source_id_s[i])
    RV_e.append(rv_e_s[i])

# combine arrays in dataframe:
df = vaex.from_arrays(source_id=source_id, ra=ra, dec=dec, pmra=pmra,
                      pmdec=pmdec, BPRP0=BPRP0, G_0=G_0, parallax=parallax,
                      RPlx=RPlx, RV=RV, RV_e=RV_e)
# save selected stars:
df.export_csv('selected_stars.csv')
```

References

Arlandini, Claudio et al. (Nov. 1999). "Neutron Capture in Low-Mass Asymptotic Giant Branch Stars: Cross Sections and Abundance Signatures". In: *The Astrophysical Journal* 525.2, p. 886. DOI: 10.1086/307938. URL: https://dx.doi.org/10.1086/307938.

Awad, P et al. (Sept. 2023). "Swarming in stellar streams: unveiling the structure of the Jhelum stream with ant colony-inspired computation". In: *A&A*.

Bastian, Nate and Carmela Lardo (Sept. 2018). "Multiple Stellar Populations in Globular Clusters". In: *Annual Review of Astronomy and Astrophysics* 56.1, pp. 83–136. DOI: 10.1146/annurev-astro-081817-051839. URL: https://doi.org/10.1146%2Fannurev-astro-081817-051839.

Binney, James and Michael Merrifield (1998). *Galactic Astronomy*. Vol. 62. Princeton University Press. ISBN: 9780691025650. URL: http://www.jstor.org/stable/j.ctv1nxcw51 (visited on 11/08/2023).

Binney, James and Scott Tremaine (2008). *Galactic Dynamics: Second Edition*.

Blanton, Michael R. et al. (June 2017). "Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe". In: *The Astronomical Journal* 154.1, p. 28. DOI: 10.3847/1538-3881/aa7567. URL: https://dx.doi.org/10.3847/1538-3881/aa7567.

Brauer, Kaley et al. (Feb. 2019). "The Origin of r-process Enhanced Metal-poor Halo Stars In Now-destroyed Ultra-faint Dwarf Galaxies". In: *The Astrophysical Journal* 871.2, p. 247. DOI: 10.3847/1538-4357/aafafb. URL: https://dx.doi.org/10.3847/1538-4357/aafafb.

Bressan, Alessandro et al. (Nov. 2012). "PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code". In: *MNRAS* 427.1, pp. 127–145. DOI: 10.1111/j.1365-2966.2012.21948.x. arXiv: 1208.4498 [astro-ph.SR].

Busso, M., R. Gallino, and G. J. Wasserburg (1999). "Nucleosynthesis in Asymptotic Giant Branch Stars: Relevance for Galactic Enrichment and Solar System Formation". In: *Annual Review of Astronomy and Astrophysics* 37.1, pp. 239–309. DOI: 10.1146/annurev.astro.37.1.239. eprint: https://doi.org/10.1146/annurev.astro.37.1.239. URL: https://doi.org/10.1146/annurev.astro.37.1.239.

Byström, Amanda (2021). *The Metallicity Structure of the Milky Way halo I : Creating a stellar catalogue of the distant halo's red giants*.

— (2022). "The Metallicity Structure of the Milky Way Halo II : Characterising the distant halo substructure". MA thesis.

Choi, Jieun et al. (May 2016). "MESA ISOCHRONES AND STELLAR TRACKS (MIST). I. SOLAR-SCALED MODELS". In: *The Astrophysical Journal* 823.2, p. 102. DOI: 10.3847/0004-637X/823/2/102. URL: https://dx.doi.org/10.3847/0004-637X/823/2/102.

Cowan, John J. et al. (Feb. 2021). "Origin of the heaviest elements: The rapid neutron-capture process". In: *Reviews of Modern Physics* 93.1. DOI: 10.1103/revmodphys.93.015002. URL: https://doi.org/10.1103%2Frevmodphys.93.015002.

Deason, A. J., V. Belokurov, and N. W. Evans (Sept. 2011). "The Milky Way stellar halo out to 40 kpc: squashed, broken but smooth". In: *Monthly Notices of the Royal Astronomical Society* 416.4, pp. 2903–2915. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2011.19237.x. eprint: https://academic.oup.com/mnras/article-pdf/416/4/2903/2980439/mnras0416-2903.pdf. URL: https://doi.org/10.1111/j.1365-2966.2011.19237.x.

Dickens, R. J. et al. (May 1991). "Evidence from stellar abundances for a large age difference between two globular clusters". In: 351.6323, pp. 212–214. DOI: 10.1038/351212a0.

Dorigo, Marco and Thomas Stützle (June 2004). *Ant Colony Optimization*. The MIT Press. ISBN: 9780262256032. DOI: 10.7551/mitpress/1290.001.0001. URL: https://doi.org/10.7551/mitpress/1290.001.0001.

Errani, Raphaël et al. (June 2022). "The *Pristine* survey – XVIII. C-19: tidal debris of a dark matter-dominated globular cluster?" In: *Monthly Notices of the Royal Astronomical Society* 514.3, pp. 3532–3540. DOI: 10.1093/mnras/stac1516. URL: https://doi.org/10.1093%2Fmnras%2Fstac1516.

Foppen, Anne (Aug. 2023). "The metallicity structure of the outer Galactic halo". Bachelor's Thesis.

Freeman, K. C. (June 1970). "On the Disks of Spiral and S0 Galaxies". In: *ApJ* 160, p. 811. DOI: 10.1086/150474.

Gaia Collaboration, T. Prusti, et al. (Nov. 2016). "The Gaia mission". In: 595, A1, A1. DOI: 10.1051/0004-6361/201629272. arXiv: 1609.04153 [astro-ph.IM].

Gaia Collaboration, A. Vallenari, et al. (June 2023). "Gaia Data Release 3. Summary of the content and survey properties". In: 674, A1, A1. DOI: 10.1051/0004-6361/202243940. arXiv: 2208.00211 [astro-ph.GA].

Gratton, Raffaele, Christopher Sneden, and Eugenio Carretta (Sept. 2004). "Abundance Variations Within Globular Clusters". In: 42.1, pp. 385–440. DOI: 10.1146/annurev.astro.42.053102.133945.

Gratton, Raffaele G., Eugenio Carretta, and Angela Bragaglia (Feb. 2012). "Multiple populations in globular clusters". In: *The Astronomy and Astrophysics Review* 20.1. DOI: 10.1007/s00159-012-0050-3. URL: https://doi.org/10.1007%2Fs00159-012-0050-3.

Harris, William E. and René Racine (1979). "Globular Clusters in Galaxies". In: *Annual Review of Astronomy and Astrophysics* 17.1, pp. 241–274. DOI: 10.1146/annurev.aa.17.090179.001325. eprint: https://doi.org/10.1146/annurev.aa.17.090179.001325. URL: https://doi.org/10.1146/annurev.aa.17.090179.001325.

Helmi, Amina (Aug. 2020). "Streams, Substructures, and the Early History of the Milky Way". In: 58, pp. 205–256. DOI: 10.1146/annurev-astro-032620-021917. arXiv: 2002.04340 [astro-ph.GA].

Helmi, Amina, Carine Babusiaux, et al. (Oct. 2018). "The merger that led to the formation of the Milky Way's inner stellar halo and thick disk". In: 563.7729, pp. 85–88. DOI: 10.1038/s41586-018-0625-x. arXiv: 1806.06038 [astro-ph.GA].

Helmi, Amina and Simon D. M. White (Aug. 1999). "Building up the stellar halo of the Galaxy". In: *Monthly Notices of the Royal Astronomical Society* 307.3, pp. 495–517. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02616.x. eprint: https://academic.oup.com/mnras/article-pdf/307/3/495/2842682/307-3-495.pdf. URL: https://doi.org/10.1046/j.1365-8711.1999.02616.x.

Helmi, Amina, Simon D. M. White, et al. (Nov. 1999). "Debris streams in the solar neighbourhood as relics from the formation of the Milky Way". In: 402.6757, pp. 53–55. DOI: 10.1038/46980. arXiv: astro-ph/9911041 [astro-ph].

Ibata, Rodrigo et al. (June 2021). "Charting the Galactic Acceleration Field. I. A Search for Stellar Streams with Gaia DR2 and EDR3 with Follow-up from ESPaDOnS and UVES". In: *The Astrophysical Journal* 914.2, p. 123. DOI: 10.3847/1538-4357/abfcc2. URL: https://dx.doi.org/10.3847/1538-4357/abfcc2.

Kilic, Mukremin et al. (Mar. 2017). "The Ages of the Thin Disk, Thick Disk, and the Halo from Nearby White Dwarfs". In: *ApJ* 837.2, 162, p. 162. DOI: 10.3847/1538-4357/aa62a5. arXiv: 1702.06984 [astro-ph.SR].

Kobayashi, Chiaki et al. (Jan. 2023). "Can Neutron Star Mergers Alone Explain the r-process Enrichment of the Milky Way?" In: *The Astrophysical Journal Letters* 943.2, p. L12. DOI: 10.3847/2041-8213/acad82. URL: https://dx.doi.org/10.3847/2041-8213/acad82.

Koppelman, Helmer H. et al. (Nov. 2019). "Multiple retrograde substructures in the Galactic halo: A shattered view of Galactic history". In: *Astronomy &amp Astrophysics* 631, p. L9. DOI: 10.1051/0004-6361/201936738. URL: https://doi.org/10.1051%2F0004-6361%2F201936738.

Koppelman, Helmer H. and Helmi, Amina (2021a). "Determination of the escape velocity of the Milky Way using a halo sample selected based on proper motion". In: *A&A* 649, A136. DOI: 10.1051/0004-6361/202038777. URL: https://doi.org/10.1051/0004-6361/202038777.

— (2021b). "The reduced proper motion selected halo: Methods and description of the catalogue". In: *A&A* 645, A69. DOI: 10.1051/0004-6361/202038178. URL: https://doi.org/10.1051/0004-6361/202038178.

Lee, Abigail J. (Oct. 2023). "Carbon Stars as Standard Candles: An Empirical Test for the Reddening, Metallicity, and Age Sensitivity of the J-region Asymptotic Giant Branch (JAGB) Method". In: *The Astrophysical Journal* 956.1, p. 15. DOI: 10.3847/1538-4357/acee69. URL: https://dx.doi.org/10.3847/1538-4357/acee69.

Licquia, Timothy C. and Jeffrey A. Newman (June 2015). "Improved Estimates of the Milky Way's Stellar Mass and Star Formation Rate from Hierarchical Bayesian Meta-Analysis". In: *ApJ* 806.1, 96, p. 96. DOI: 10.1088/0004-637X/806/1/96. arXiv: 1407.1078 [astro-ph.GA].

Mackereth, J. Ted and Jo Bovy (Mar. 2020). "Weighing the stellar constituents of the galactic halo with APOGEE red giant stars". In: *MNRAS* 492.3, pp. 3631–3646. DOI: 10.1093/mnras/staa047. arXiv: 1910.03590 [astro-ph.GA].

Malhan, Khyati and Rodrigo A Ibata (Apr. 2018). "STREAMFINDER – I. A new algorithm for detecting stellar streams". In: *Monthly Notices of the Royal Astronomical Society* 477.3, pp. 4063–4076. DOI: 10.1093/mnras/sty912. URL: https://doi.org/10.1093%2Fmnras%2Fsty912.

Maoz, Dan, Filippo Mannucci, and Gijs Nelemans (Aug. 2014). "Observational Clues to the Progenitors of Type Ia Supernovae". In: *Annual Review of Astronomy and Astrophysics* 52.1, pp. 107–170. DOI: 10.1146/annurev-astro-082812-141031. URL: https://doi.org/10.1146%2Fannurev-astro-082812-141031.

Martin, Nicolas F et al. (Aug. 2022). "The iPristine/i survey – XVI. The metallicity of 26 stellar streams around the Milky Way detected with the ttSTREAMFINDER/tt in iGaia/i EDR3". In: *Monthly Notices of the Royal Astronomical Society* 516.4, pp. 5331–5354. DOI: 10.1093/mnras/stac2426. URL: https://doi.org/10.1093%2Fmnras%2Fstac2426.

Martin, Nicolas F., Else Starkenburg, et al. (2023). *The Pristine survey – XXIII. Data Release 1 and an all-sky metallicity catalogue based on Gaia DR3 BP/RP spectro-photometry*. arXiv: 2308.01344 [astro-ph.GA].

Martin, Nicolas F., Kim A. Venn, et al. (Jan. 2022). "A stellar stream remnant of a globular cluster below the metallicity floor". In: 601.7891, pp. 45–48. DOI: 10.1038/s41586-021-04162-2. arXiv: 2201.01309 [astro-ph.GA].

McWilliam, Andrew (1997). "Abundance ratios and galactic chemical evolution". In: *Annual Review of Astronomy and Astrophysics* 35.1, pp. 503–556.

Myeong, G C et al. (July 2019). "Evidence for two early accretion events that built the Milky Way stellar halo". In: *Monthly Notices of the Royal Astronomical Society* 488.1, pp. 1235–1247. ISSN: 0035-8711. DOI: 10.1093/mnras/stz1770. eprint: https://academic.oup.com/mnras/article-pdf/488/1/1235/28936943/stz1770.pdf. URL: https://doi.org/10.1093/mnras/stz1770.

Posti, Lorenzo, Fraternali, Filippo, and Marasco, Antonino (2019). "Peak star formation efficiency and no missing baryons in massive spirals". In: *A&A* 626, A56. DOI: 10.1051/0004-6361/201935553. URL: https://doi.org/10.1051/0004-6361/201935553.

Prantzos, N., C. Charbonnel, and C. Iliadis (Dec. 2017). "Revisiting nucleosynthesis in globular clusters". In: *Astronomy &amp Astrophysics* 608, A28. DOI: 10.1051/0004-6361/201731528. URL: https://doi.org/10.1051%2F0004-6361%2F201731528.

Prantzos, N., Charbonnel, C., and Iliadis, C. (2007). "Light nuclei in galactic globular clusters: constraints on the self-enrichment scenario from nucleosynthesis". In: *A&A* 470.1, pp. 179–190. DOI: 10.1051/0004-6361:20077205. URL: https://doi.org/10.1051/0004-6361:20077205.

Schlegel, David J., Douglas P. Finkbeiner, and Marc Davis (June 1998). "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds". In: *ApJ* 500.2, pp. 525–553. DOI: 10.1086/305772. arXiv: astro-ph/9710327 [astro-ph].

Sparke, Linda S. and III Gallagher John S. (2007). *Galaxies in the Universe: An Introduction.*

Starkenburg, Else et al. (May 2017). "The Pristine survey – I. Mining the Galaxy for the most metal-poor stars". In: *Monthly Notices of the Royal Astronomical Society* 471.3, pp. 2587–2604. ISSN: 0035-8711. DOI: 10.1093/mnras/stx1068. eprint: https://academic.oup.com/mnras/article-pdf/471/3/2587/19496762/stx1068.pdf. URL: https://doi.org/10.1093/mnras/stx1068.

Tolstoy, Eline, Vanessa Hill, and Monica Tosi (Sept. 2009). "Star-Formation Histories, Abundances, and Kinematics of Dwarf Galaxies in the Local Group". In: *Annual Review of Astronomy and Astrophysics* 47.1, pp. 371–425. DOI: 10.1146/annurev-astro-082708-101650. URL: https://doi.org/10.1146%2Fannurev-astro-082708-101650.

Tononi, Jordi et al. (2019). "Effects of 22Ne sedimentation and metallicity on the local 40 pc white dwarf luminosity function". In: *A&A* 628, A52. DOI: 10.1051/0004-6361/201834267. URL: https://doi.org/10.1051/0004-6361/201834267.

Truran J. W., Jr. and A. Heger (Dec. 2003). "Origin of the Elements". In: *Treatise on Geochemistry* 1, p. 711. DOI: 10.1016/B0-08-043751-6/01059-8.

Tsujimoto, Takuji and Nobuya Nishimura (Aug. 2018). "Early Chemical Evolution of Zn Driven by Magnetorotational Supernovae and the Pathway to the Solar Zn Composition". In: *The Astrophysical Journal Letters* 863.2, p. L27. DOI: 10.3847/2041-8213/aad86b. URL: https://dx.doi.org/10.3847/2041-8213/aad86b.

Viswanathan, Akshara et al. (Mar. 2023). "Hidden deep in the halo: selection of a reduced proper motion halo catalogue and mining retrograde streams in the velocity space". In: *Monthly Notices of the Royal Astronomical Society* 521.2, pp. 2087–2102. ISSN: 0035-8711. DOI: 10.1093/mnras/stad380. eprint: https://academic.oup.com/mnras/article-pdf/521/2/2087/49534827/stad380.pdf. URL: https://doi.org/10.1093/mnras/stad380.

Watkins, Laura L. et al. (Mar. 2019). "Evidence for an Intermediate-mass Milky Way from Gaia DR2 Halo Globular Cluster Motions". In: *ApJ* 873.2, 118, p. 118. DOI: 10.3847/1538-4357/ab089f. arXiv: 1804.11348 [astro-ph.GA].

Watson, Darach et al. (2019). "Identification of strontium in the merger of two neutron stars". In: *Nature* 574.7779, pp. 497–500.

Yuan, Zhen et al. (May 2022). "The Pristine survey – XVII. The C-19 stream is dynamically hot and more extended than previously thought". In: *Monthly Notices of the Royal Astronomical Society* 514.2, pp. 1664–1671. DOI: 10.1093/mnras/stac1399. URL: https://doi.org/10.1093%2Fmnras%2Fstac1399.