Doctoral Dissertations and Master's Theses

Summer 2023

Development and Testing of a New Method for Velocity-Selecting White Dwarfs from Gaia by Galactic Population

Joseph Hammill
Embry-Riddle Aeronautical University, hammilj1@my.erau.edu

DEVELOPMENT AND TESTING OF A NEW METHOD FOR VELOCITY-SELECTING WHITE DWARFS FROM GAIA BY GALACTIC POPULATION

BY

JOSEPH HAMMILL

A Thesis

Submitted to the Department of Physical Sciences

In partial fulfillment of the requirements

for the degree of

Master of Science in Engineering Physics

08/2023

Embry-Riddle Aeronautical University
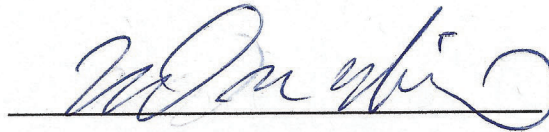
Daytona Beach, Florida

DEVELOPMENT AND TESTING OF A NEW METHOD FOR VELOCITY-SELECTING WHITE DWARFS FROM GAIA BY GALACTIC POPULATION

by

Joseph Hammill

This thesis was prepared under the direction of the candidate's Thesis Committee Chair, Dr. Ted von Hippel, Professor, Daytona Beach Campus, and has been approved by the Thesis Committee. It was submitted to the Department of Physical Sciences in partial fulfillment of the requirements of the degree of Master of Science in Engineering Physics

THESIS COMMITTEE:

_____

Dr. Ted von Hippel,
Committee Chair

Jason Aufdenberg
Digitally signed by Jason Aufdenberg
Date: 2023.08.15 16:47:08 -04'00'
_____
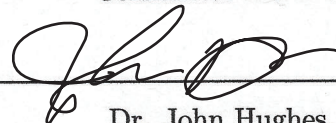
Dr. Jason Aufdenberg,
Committee Member

_____
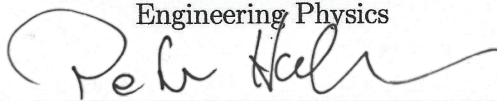
Dr. Saida Caballero-Nieves,
Committee Member

Edwin Mierkiewicz
Digitally signed by Edwin Mierkiewicz
DN: cn=Edwin Mierkiewicz, o, ou,
email=mierkiee@erau.edu, c=US
Date: 2023.08.17 13:28:21 -05'00'
_____

Dr. Edwin Mierkiewicz,
Graduate Program Chair,
Engineering Physics

_____

Dr. John Hughes,
Department Chair,
Physical Sciences

_____

Dr. Peter Hoffmann
Dean, College of Arts and Sciences

_____

Dr. Christopher Grant,
Associate Provost of Academic Support

iii

Abstract

The detailed processes by which spiral galaxies form remains an open question in modern cosmology. Observations of the current configuration of spiral galaxies including the Milky Way reveal thin and thick disk and halo populations which must all be accounted for in formation theories and likely have distinct ages. Using the Milky Way as an example to probe this question, we are studying the formation history of these structures.

This work details our approach to age-dating the galaxy, velocity-selecting targets from a sample of white dwarfs from the Gaia DR3 catalog that have also 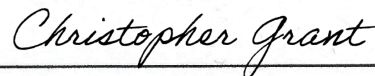been age-analysed using BASE-9. BASE-9 uses photometry, as well as prior measurements and values from ancillary data, such as trigonometric parallaxes, fitting up to nine stellar parameters for a given target. White dwarfs provide a valuable target type, as they are long−lived and stable objects which remain from the earliest epochs of star formation in each Galactic population. Gaia astrometry provides us with proper motions of these stars, though not radial velocities. We have developed a scheme to complete the velocity vector by estimating radial velocities in terms of the proper motion, and assigning stars by probability of membership to the thin disk, thick disk, or halo population. We then obtain an overall age-probability distribution for each component by the use of a hierarchical analysis of the BASE-9 age posterior distributions of the individual stars. We have completed a hierarchical analysis for a subset of stars identified as thick disk objects, and initial analysis is discussed within. Initial results show an age posterior distribution which is singly-peaked near 6 Gyr, with a peak FWHM of 0.64 Gyr. This may suggest that the sample of high-velocity white dwarfs in the solar neighbourhood is contaminated by stars from a

later merger event, supernovae within prior binaries, or gravitational scattering due to molecular clouds or Galactic features such as transient spiral structures or a bar. However, another possibility revealed by testing is that the velocity-selection method is unreliable along certain lines of sight and further filtering and refinement of our method is warranted.

Acknowledgments

The author would like to thank advisor, collaborator, and committee chair Ted von Hippel for the now three and a half years of learning, patience, advisement, feedback and opportunities. He also extends gratitude to fellow collaborators Alisa Tiselska and Ally Woodruff who have assisted in running BASE-9 models and in providing ideas and feedback for developing the methods described in this thesis. Further thanks go to Elliot Robinson, for giving advice and clarification on the coding side of things, especially in R with which I was only previously familiar by name. I look forward to refining this process and extracting results from it with all three of you.

The author would also like to thank UT Austin students Bryce Hobbs and Malia Kao for providing encouragement, additional aspects of white dwarf science to consider, and suggesting additional resources and literature at AAS 242 Meeting in Albuquerque. Considering this meeting, thanks also go to the Embry-Riddle Physical Sciences Department for funding the author's attendance. The author also thanks Rachel Schneider for providing ideas for interpreting the results of the V/V_{max} test in §5.2.2. Further study will be necessary to determine if my results imply the fall of the church or the end of the world as she offered.

Finally, a huge thanks to the committee members Drs. Aufdenberg, Caballero-Nieves, and von Hippel for providing their time, feedback, and expertise to refine this thesis document.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

The formation of galaxies is a necessary prediction of any cosmological theory. In the following, I will describe two galactic formation mechanism and the imprints they leave on the ages of galactic structures.

1.1 Galaxy Formation

By studying the age-probability distribution (posterior distribution) of each of the Galactic thick disk and halo, I can see signs of not just their mean and modal age, but of their formation history. This reveals itself in the number of peaks and the width of peaks in the posterior distribution (an example of such a distribution is seen in Figure 4.1). Such an undertaking, based on the targets identified in Section 1.2, is a new method introduced to the field of galactic archaeology.

1.1.1 Top-Down Formation

Generally, a top-down scenario describes the largest-scale structures being formed or analysed first. In galactic astronomy, this is translated to a collapse of a single huge cloud of material to form stars and the galaxy seen today. This form of collapse is similar to star-formation, and can indeed result in the formation of spiral galaxies with a disk and halo structure. When the top-down galactic formation hypothesis was first

suggested by Eggen and Sandage (whose model provides the basis for the remainder of this paragraph), it was shown that the collapse time had to be on the order of 0.1 Gyr or less in order to produce observed elliptical halo orbits [19]. Eccentricities of stars and other matter are minimally affected during the collapse, and so modern halo objects must have begun with significant radial velocities relative to the proto-galactic barycenter as compared to their orbital speed. Ultimately some material would lose vertical velocity by dynamic friction and fall into disk orbits determined by the remaining angular momentum. Hence, the timescale of formation by this mechanism should be significantly less than the rotational period of the galaxy, and is estimated to be on the order of 0.1 Gyr.This model predicts a sharply peaked age posterior distribution for the formation history of the galactic halo.

In the modern understanding of galactic structure which includes the dark matter halo, such a process requires a larger scale matter clump in the early universe. This would in turn imply an overall smoother distribution of matter. Such is favored by predictions of the Hot Dark Matter (HDM) theory, leaving dark matter to be light particles which can travel extremely quickly. Cosmic Microwave Background observations show that the early universe was relatively uniform [5]. However, the best candidate particle fitting the HDM model is the neutrino which tends to be so weakly interacting as to not form any significant inhomogeneities from smooth early matter distributions [23].

1.1.2 Bottom-Up Formation

Bottom-up formation, or hierarchical clustering, is predicted by Λ Cold Dark Matter (ΛCDM) models. In the traditional model of bottom-up formation, a galaxy scale dark matter halo is present to provide gravitational force to begin accreting smaller structures of electromagnetically active (light, as opposed to dark) matter [60]. Importantly, the dark matter halo remains largely unaffected by the infalling baryonic material, neither leading to shock heating nor itself being perturbed from its shell-like distribution [32]. In this model of formation, I would predict the Milky Way's halo age posterior distribution to show many peaks of similar scale from the history of

mergers. The thick disk age distribution appears broader than in the top-down case, and will be made of stars formed from the infalling gas carried by clusters which will build the growing Galaxy.

There are issues with this model as well, first that some mechanism is needed to dramatically slow star formation as the galaxy approaches its modern shape. This has been suggested to be a combination of the ongoing star formation itself with supernovae of the highest mass earliest stars [16]. Thus, while a clean halo sample is expected to show signatures of many merger events, should this theory be reflective of the Milky Way's history, a thick disk posterior distribution might show a singular peak with an asymmetric shape which falls off more steeply in the direction of the current epoch. Another issue is that the formation of galaxies in the ΛCDM model is predicted to be so high that merger events would disrupt many of the spiral galaxies seen today. A merger rate as high as that predicted should leave the universe containing many more irregular galaxies than are observed. An intermediate theory called Warm Dark Matter (WDM) predicts lighter dark matter particles whose speed and mass reduce the mass-clumping predicted by pure ΛCDM (and so reducing the number of galaxies predicted and preserving more as spirals) while its constituent particles are still too massive to be neutrinos [36, 45]. I do not expect to see signatures of this merger-overproduction problem in age posterior distributions for a lone galaxy. The Milky Way's existence as a spiral galaxy already suggests that the density of proto-galactic components in the local Universe was not so great as to disrupt its organization.

1.1.3   Age-Dating in the Context of Galactic Formation

The predictions of low-galactic formation rates from a neutrino HDM, which would likely fail to produce galaxies by 350 Myr after the Big Bang [15], have led to the top-down scenario being largely discounted by the galactic astronomy community. We therefore do not hope to settle the HDM versus CDM debate through this work. Rather, we hope to see signatures of the merger events which constructed the early Milky Way. If my method can successfully identify a significant population of halo stars then we can comment on the apparent galactic merger rate during the birth

of our Galaxy. This would reveal itself in the appearance of a multi-peaked age-distribution for the halo. If I can identify a significant thick disk population, then the formation time reflected in this population gives insights into disk formation processes and may, through gravitational freefall considerations, constrain the relationship between the accreted mass and size of the Galactic halo at the time the disk began to form. Such information helps dark matter theorists to refine their models and so contributes to broader cosmology.

1.2   White Dwarfs

White dwarfs make an ideal target type for such research. Once a low-mass (less than 8 $M_\odot$ [24]) star ends its fusion life and becomes a white dwarf, provided it is allowed to evolve in isolation (I will discuss potentially major issues that arise if this is not done in §4.3.2) it cools over many billions of years. Equipped with models that predict white dwarf aging, it becomes possible to match observed colors and magnitudes to ages.

While I will discuss many limitations of using white dwarfs as a target object, they remain the best target type for a large survey study. Through initial-final mass-relations and cooling models, their ages can be determined and the monotonic nature of their cooling greatly reduces the possibility of degenerate age solutions. Main-sequence stars solve some of the visibility constraints (a sun-like star with main sequence lifetime on the low end of halo estimates has a Gaia detectable range of around 19 kpc), post main-sequence stars can suffer from degenerate ages as they evolve and cross back over the main-sequence. In limited cases, this can be overcome but it requires ultraviolet photometry. Even when UV photometry is available, age determination is seen to be only on the order of ± 1 Gyr precision [27]. No UV photometric survey on the scale of Gaia yet exists.

1.2.1   The Gentile Fusillo Catalog

The Gentile Fusillo catalog (GFC) provides us (the white dwarf community) with a sample of white dwarfs from the Gaia Early Data Release 3 (EDR3) catalog which have passed data quality and white dwarf-validation tests [21]. Using stars which are strongly suspected to be white dwarfs through spectroscopic means [1], the GFC group identify the white dwarf region of the H-R diagram from the EDR3 catalog. Their function used to define the white dwarf region is

$$G_{abs} > 6 + 5(G_{BP} - G_{RP}) \tag{1.1}$$

where G_{abs} refers to the star's absolute magnitude in the bandpass of the Gaia G filter and the quantity in parentheses is a measure of star color C (an analog for temperature), which is calculated as a difference in magnitudes in two filter bands, blue and red Gaia filters (G_{BP} and G_{RP}) in this case. With this region defined and the distribution of stars with spectroscopic observations overlaid, they develop a background probability function which is zero outside of the region of Equation 1.1 and is otherwise contoured according to observations. The probability of an object being a white dwarf is then estimated as the integrated product of this background probability P_{SDSS} with a two-dimensional Gaussian distribution whose width and center are determined by their individual magnitude and color measurements and the associated error. Therefore the white dwarf probability, P_{WD} can be written

$$P_{WD} = k \iint_{HR} P_{SDSS} f(G_{abs}, \sigma_{G_{abs}}) f(C, \sigma_C) dG_{abs} dC \tag{1.2}$$

where the functions f are the two-dimensional Gaussians representing Gaia measurements and k is a normalising factor for those functions. The Gaussians representing these measurements are normalized according to the Gaia signal-to-noise ratio (SNR) for that object, the normalization being set to unity if the SNR is greater than 10, and 0.5 if less. By selecting objects from the GFC with a probability of being a white dwarf greater than 75%, we begin this analysis with approximately 359000 white

dwarfs. Making further cuts to these data based on SNR in the parallax measurement, white dwarf magnitude ranges in other filters, and a maximum value of G_{abs}, we are left with a sample size of 104577 white dwarfs.

1.2.2 BASE-9

Bayesian Analysis for Stellar Evolution with Nine Parameters (BASE-9) is a software which can recover probabilities and posterior distributions for the "age, metallicity, helium abundance, distance modulus, and line-of-sight absorption for a cluster, and the mass, binary mass ratio, and cluster membership probability for every stellar object" from a combination of Markov Chain Monte Carlo (MCMC) and numerical integration methods using photometry of target objects as well as ancillary prior measurements of target parameters such as metallicity and parallax [57, 58]. In the MCMC technique, values are drawn from the parameter space in question (BASE-9 investigates an up to six-dimensional parameter space in this way) and are compared to observation through the use of a model. From the level of agreement between model and observation, the fit quality of a parameter draw is found and that draw retained or rejected accordingly. In this way, better fitting and more probable regions of parameter space will be sampled more often, building the posterior distributions. However, all parameter solutions retained by BASE-9 during a run are valid solutions for the stars under study, within the uncertainties in provided photometry.

BASE-9 is the first application of data from the GFC, with photometry being used to generate possible solutions and age distributions for the white dwarfs under investigation. It has specialized routines which allow it to study white dwarfs and their complexities which differentiate them from main-sequence stars and still perform fitting [39]. It uses white dwarf cooling and interior models [2, 7, 38, 43, 61], as well as initial-final mass relations (IFMRs) [44, 59] and an initial mass function (IMF) [37] to rewind the white dwarf back to its formation and entry onto the main-sequence. These models are integral to this work's goal of finding age posterior distributions for the Galaxy, as the most accurate white dwarf ages will translate into the most accurate Galactic ages.

1.2.3 Detectability of Old White Dwarfs

As I hope to study the population of the Galactic halo, I need detectors to be able to see them. With this in mind, a determination of how far away Gaia can detect these objects is in order. Current estimates for the age of the Milky Way halo typically fall between 10-12 Gyr [29, 33, 49]. The limiting magnitude for Gaia in its G band is 20.7 [25]. With the 10 Gyr and older white dwarfs from the latest version of the Bergeron group white dwarf cooling models [3] and this limiting magnitude, I find a maximum detectable range for white dwarfs of a sufficient age to be halo members (refer to Figure 1.1).
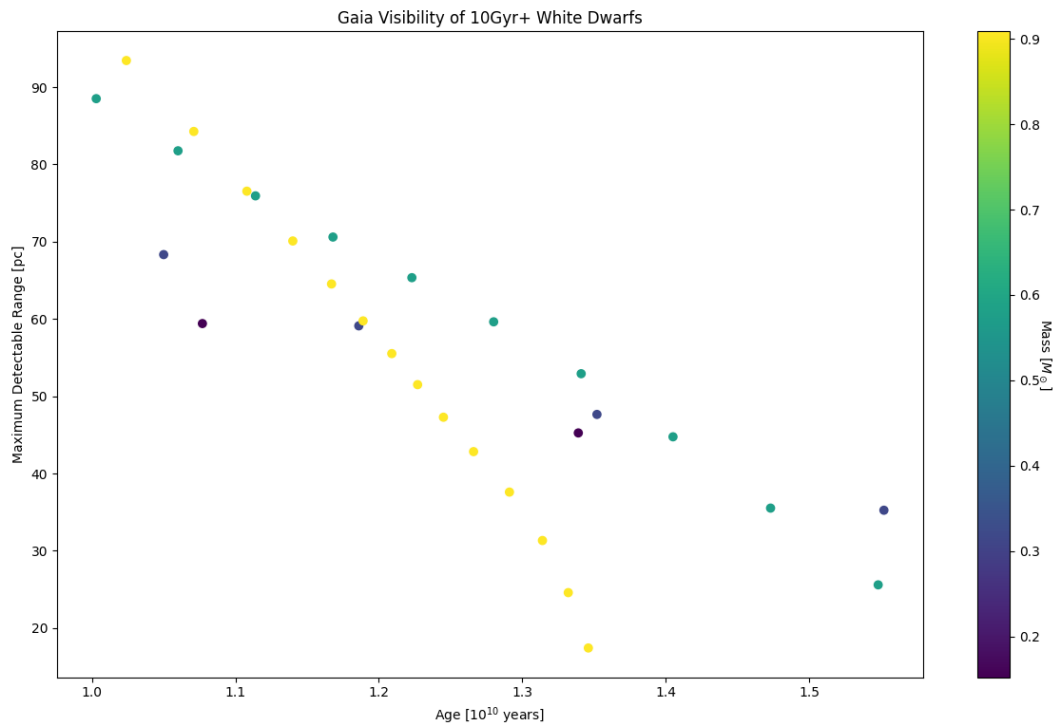
Figure 1.1: Maximum range 10 Gyr or older white dwarfs can be measured with Gaia. Models are from Bergeron group white dwarf evolutionary sequences, University of Montreal, and magnitudes are converted from SDSS according to Tonry et al. [53].

The maximum such range for the brightest white dwarf is just greater than 90 parsecs. With the scale height of the thin disk being estimated between 220-320 parsecs [17], it is impossible to study halo white dwarfs in isolation from thin and thick disk objects

in position space. The limiting detectable range in practice is likely less still due to SNR requirements we make and those in the GFC. Therefore, we need an alternative way to identify halo objects. The situation is slightly better for the thick disk, whose age is estimated at $\leq$10 Gyr [42], but the argument remains. Gaia reports proper motions of targets, so we will make use of kinematics to identify thick disk and halo white dwarfs on high-velocity orbits through the solar neighborhood according to the method laid out in Chapters 2 and 3.

Chapter 2

TRANSFORMATION OF VELOCITY FRAMES

In this section, I will describe the coordinate transformation between the Earth-equatorial frame of Gaia data products and the Galactic frame more suited to this work. I will also justify the use of velocity space for the study of white dwarfs in the older structures of our Galaxy.

2.1 The Radial-Velocity Problem

We live in three spatial dimensions, and so would like to study white dwarfs with the same dimensionality. Gaia reports proper motions of its targets, giving two dimensions of their velocity vector. It is also fitted with the Radial Velocity Spectrometer (RVS) instrument which has been used to measure almost 34 million targets. However, when observing faint objects this instrument is limited to spectral features corresponding to temperatures between 3100 and 6750K, and it has a detection cutoff at $G_{RVS} = 14$ [31]. Even if the dim object temperature range and band of available magnitudes for the RVS was not exclusive of most white dwarfs, they have properties that preclude a systematic radial velocity survey on objects of this type. White dwarfs show significant pressure broadening, gravitational redshift, and in some cases magnetic line splitting. These effects combined with intrinsically low luminosity would

often translate to SNR below criteria of the GFC or of this work. The end result is we do not reliably have radial velocity measurements for stars from the GFC and will have to develop a method to account for this missing information.

2.1.1 Coordinate Systems

Gaia reports measurements in the Earth-equatorial frame, in which the three axes are proper motion in right ascension, proper motion in declination, and radial velocity.
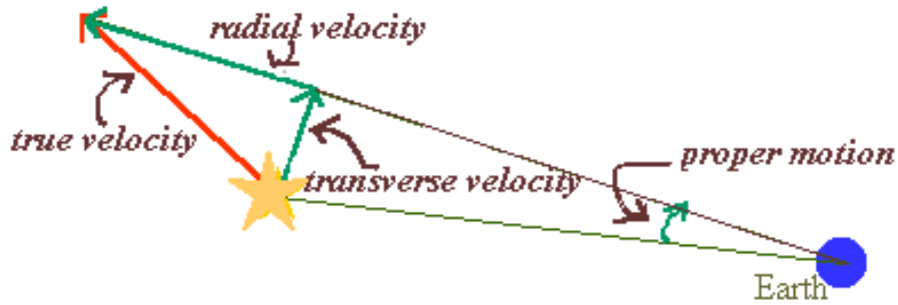
Figure 2.1: Diagram of the Earth-equatorial velocity frame. Figure retrieved from The Distance to the Hyades Cluster, Weber State University.

Gaia proper motions μ are reported in arcseconds per year ("yr^{-1}"), and these can be converted to velocities by the equation

$$v_{\alpha/\delta} = 4.74 \mu_{\alpha/\delta}/\overline{\omega} \tag{2.1}$$

where the subscripts α and δ refer to motion in right ascension and declination directions respectively, $\overline{\omega}$ is the parallax in arcseconds, and 4.74 is a conversion factor so that the result is in km/s.

For any study of Galactic structure, a galactocentric coordinate system is a natural and convenient choice. In this frame, average velocities and their dispersion for each Galactic component will not be line of sight dependent (we will see that the transformation between these frames is dependent on right ascension and declination). The Galactic velocity frame is also known as the UVW frame after the traditional symbols for its coordinates. U is inward (or outward, in certain literature) to the Galactic

center from the Sun, V is in the tangential direction of the Sun's orbit through the Galaxy, and W is the vertical component perpendicular to the plane of the Galactic disk.
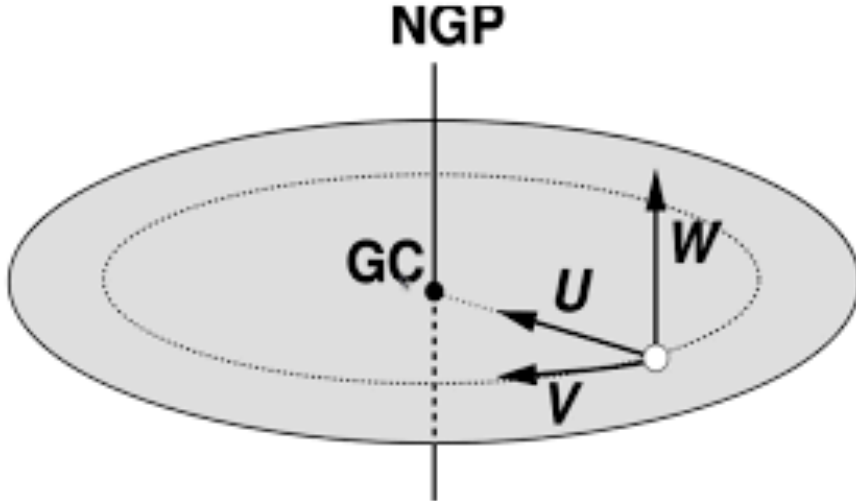
Figure 2.2: Diagram of the Galactic UVW velocity frame. GC is Galactic center, NGP north Galactic pole, and the white circle is the object whose kinematics are under investigation. Figure retrieved from The Milky Way: A structural overview, Case Western Reserve University.

Less critical though still important to this work is the two-dimensional Galactic coordinate system. In this system, Galactic longitude l is defined as zero in the direction from the Sun to the Galactic center and increasing to the east in the Galactic plane as defined by neutral hydrogen tracing [8]. Galactic latitude b is defined as the angle above or below the plane of the Galaxy, with that plane being $b = 0°$.

2.1.2 Estimation, not Measurement

Short of spending the time and money to gather the materials and expertise necessary to build and launch an orbiting observatory which can measure very precise radial velocities of dim objects, we cannot get these values for most stars in the GFC. Therefore, it becomes necessary to generate estimates for the radial velocities of our sample of white dwarfs before they can be studied in UVW-space. We test three

possibilities for the radial velocity of each star: average of the proper motion in right ascension and declination, the negative of that average (a star approaching the Sun), and zero. We then conduct a Monte-Carlo (MC) simulation using thirty velocity ordered-triples drawn from normal distributions with widths given by Gaia uncertainties, centered at proper motion measurements and estimated radial velocity. We perform one of these random draws at each of the radial velocity estimates.
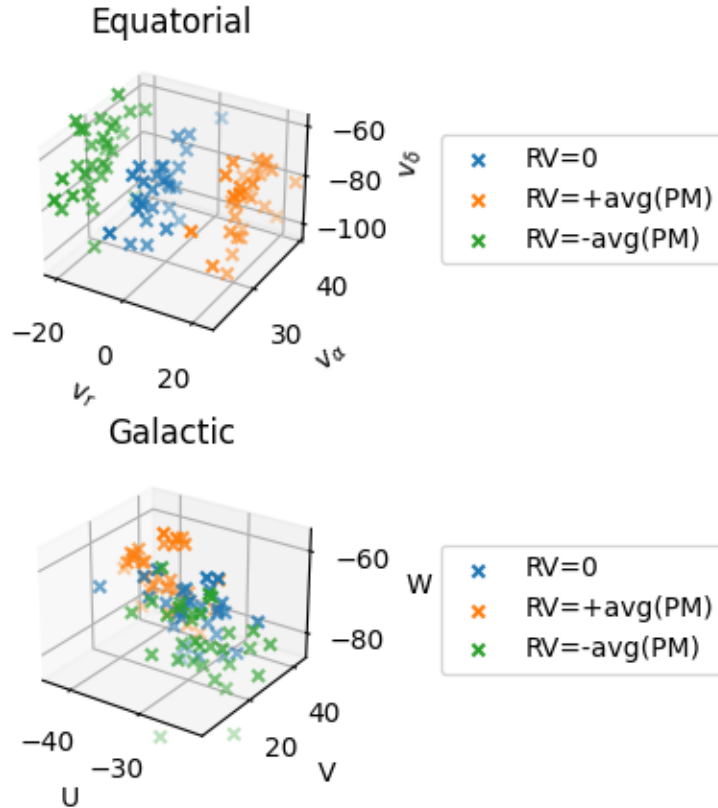
Figure 2.3: The distribution of velocity realizations for one star. This one is taken from the simulated points described in §5.1.1, and shown before and after transformation to the UVW coordinate system. Figure 3.1 shows an example of Galactic velocity points which have been assigned population candidacy.

The MC method allows us to test a range of parameters within some uncertainty condition (radial velocity, in this case), and provides a way to track the distribution of initial parameters to its final state which may change the shape of the distribution and remove symmetry [28]. A total of ninety versions, or possible realizations of

each star's velocity propagate through the transformation and sorting algorithm to be described in the following section and chapter.

2.1.3   Potential Issues with Halo Stars

The Galactic halo is a distribution of stars which lies beyond the disk of the Galaxy spread in a more uniform fashion. Any white dwarf member of that population which is currently close enough to the Sun to be detected by Gaia must then be on a plunging orbit towards (or away from) the Galactic center. Stars far away from the plane of the Galaxy will then have significant radial velocity relative to the Sun. Conversely, on certain near-planar orbits the star's orbit may mimic that of a disk star and show very little radial velocity. In both cases, this will lead to one or more radial velocity estimates and the associated thirty velocity-space points being poor fits to the actual star and reducing the likelihood of it being identified as a potential halo object. Similar issues may occur in thick disk objects, but since these still rotate with the Galaxy with the added vertical velocity, the three radial velocity estimates will be more similar. In that case, it is unlikely that one radial velocity estimate will be significantly worse as a model for the star's true behavior as compared to the halo case.

2.2   Transformation of Equatorial Velocities to the UVW Frame

We make use of the transformation matrix method to relate equatorial velocity to UVW velocity, following the method of Johnson and Soderblom as described in the following [30]. Two matrices T and A are defined (following the Johnson & Soderblom naming convention) which are necessary to make the transformation.

$$T_{i1} = \left\{ \begin{array}{c} -\cos l_{NCP} \sin \delta_{NGP} \cos \alpha_{NGP} - \sin l_{NCP} \sin \alpha_{NGP} \\ -\sin l_{NCP} \sin \delta_{NGP} \cos \alpha_{NGP} + \cos l_{NCP} \sin \alpha_{NGP} \\ \cos \delta_{NGP} \cos \alpha_{NGP} \end{array} \right\}$$

$$T_{i2} = \begin{Bmatrix} -\cos l_{NCP} \sin \delta_{NGP} \sin \alpha_{NGP} + \sin l_{NCP} \cos \alpha_{NGP} \\ -\sin l_{NCP} \sin \delta_{NGP} \sin \alpha_{NGP} - \cos l_{NCP} \cos \alpha_{NGP} \\ \cos \delta_{NGP} \sin \alpha_{NGP} \end{Bmatrix} \tag{2.2}$$

$$T_{i3} = \begin{Bmatrix} \cos l_{NCP} \cos \delta_{NGP} \\ \sin l_{NCP} \cos \delta_{NGP} \\ \sin \delta_{NGP} \end{Bmatrix}$$

where l_{NCP} is the Galactic longitude of the north celestial pole, and α_{NGP} and δ_{NGP} are the right ascension and declination of the north Galactic pole respectively. Equation 2.2 is all that is required to convert equatorial angular positional coordinates to their Galactic counterparts. A second matrix, called A, is required for velocity transformations:

$$A = \begin{bmatrix} \cos \alpha \cos \delta & -\sin \alpha & -\cos \alpha \sin \delta \\ \sin \alpha \cos \delta & \cos \alpha & -\sin \alpha \sin \delta \\ \sin \delta & 0 & \cos \delta \end{bmatrix} \tag{2.3}$$

where in this case the angles α and δ are those of the target object not of a special point in the sky. Defining the product

$$B \equiv TA \tag{2.4}$$

we can find UVW velocities according to

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = B \begin{bmatrix} v_r \\ v_\alpha \\ v_\delta \end{bmatrix} \tag{2.5}$$

and here the first component of the vector on the right is generated by the estimation scheme while the other two come from the GFC.

Chapter 3

POPULATION ASSIGNMENT

Here, I describe the systematic way in which white dwarfs are sorted into Galactic populations.

3.1 Three-Component Probabilities

Each of the ninety realisations of each white dwarf now has a UVW-frame velocity vector. The MC method renders each star as a cloud (or in most cases, a trio of clouds) in UVW space. Different sections of those clouds may fall into kinematic regions belonging to all three of the thin disk, thick disk, and halo components, with the overall distribution representing the probability of the star to belong to each of these. Probabilities will then have to be determined according to velocity components. In this procedure, we will assume that there are not correlations among the different velocity components and assume the shape of the velocity dispersion in each population. This means that probability in each velocity component will be multiplied with that in the other two to compile the overall probability of membership in a particular population. Mathematically this is written that the probability of membership in population i is

$$P_i = \prod_{j=1}^{3} P_{ij} \tag{3.1}$$

with respect to velocity components j. However, existing literature from Skuljan, Hearnshaw, and Cottrell suggests this to be an oversimplification as certain moving groups and clusters are found to occupy non-linear curves in the UV-plane [50]. Finally, for the purposes of analysing probabilities as percentages, we will demand that the probabilities for each realisation of each star are normalised. That is, that

$$\sum_{i=1}^{3} P_i = 1 \tag{3.2}$$

3.1.1 Mean Velocity and Dispersion

Velocity dispersion measurements are crucial to this work, as we will assume that the distribution of stars in velocity space for each Galactic component is approximately a Gaussian whose width is given by that dispersion σ_{ij}. Here lies another oversimplification, in nearby stars U and V velocity dispersions have been observed to be non-Gaussian [48]. Still, we will treat the probability that a realisation of a star belongs to the ith Galactic population according to the jth component of its velocity vector as

$$P_{ij} \propto \exp\left(-\frac{(v_j - \bar{v}_{ij})^2}{\sigma_{ij}^2}\right) \tag{3.3}$$

where $\bar{v}_{ij}$ is the mean of the jth velocity component in the ith population. This becomes the center of the Gaussian for each velocity component of each population.

Population	$\bar{v}_U$	σ_U	$\bar{v}_V$	σ_V	$\bar{v}_W$	σ_W
Thin Disk	0	39	-12	20	0	16
Thick Disk	0	67	-51	38	0	35
Halo	0	160	196	90	0	90

Table 3.1: Mean velocities and velocity dispersions used in this work shown by Galactic population. Note that the mean V velocity of the halo is nearly that of the orbit of the local standard of rest ($V_{LSR} \approx 202 - 241$ km/s [35]) indicating these stars tend to not carry much orbital velocity in the plane of the Galaxy. All values shown are measured in km/s. Mean velocities are from [51] and velocity dispersions from [6].

3.1.2   Stellar Number Density

In the thin disk, the most numerous stars are native thin disk stars. Thus, if any nearby star is randomly selected, the probability is high this is a thin disk star. Since the probability of making a (truly) random selection of a given object is proportional to the ratio of objects of those types to the total possible selection, the probability of selecting a star of a given type in the solar neighborhood is proportional to the local number density of those stars. As this is also independent of the velocity of each star, it becomes another multiplicative factor in the probability

$$P_i \propto n_i \prod_{j=1}^{3} \exp\left(-\frac{(v_j - \bar{v}_{ij})^2}{\sigma_{ij}^2}\right) \tag{3.4}$$

with the rest of Equation 3.4 coming from Equations 3.1 and 3.3. The remaining conversion from proportionality to equality comes from normalisation.

Population	Number Density
Thin Disk	0.124
Thick Disk	0.0156
Halo	2.65×10^{-4}

Table 3.2: Number densities of stars from each population in the stellar neighborhood as used here. All have units of stars/pc^3, and are calculated from [18, 34, 56].

Currently, the number densities used represent the total of the respective population in the solar neighborhood. As we study white dwarfs, we are more interested in the number density of white dwarfs from each population in nearby space. Our group is in the process of generating this correction, and since more stars will have evolved to a white dwarf stage in an older population we expect this may result in identifying more stars as thick disk and halo objects.

3.2   Halo and Thick Disk Membership

We make the requirement that, across all realisations of each star, the probability is greater than 50% that a star is a thick disk or halo member candidate. Due to the

methods of estimating the radial velocities (see §2.1.2) this effectively introduces the requirement that stars either have such high proper motions they fit in all estimations or that one estimate fits so strongly in one population as to dominate the selection process.
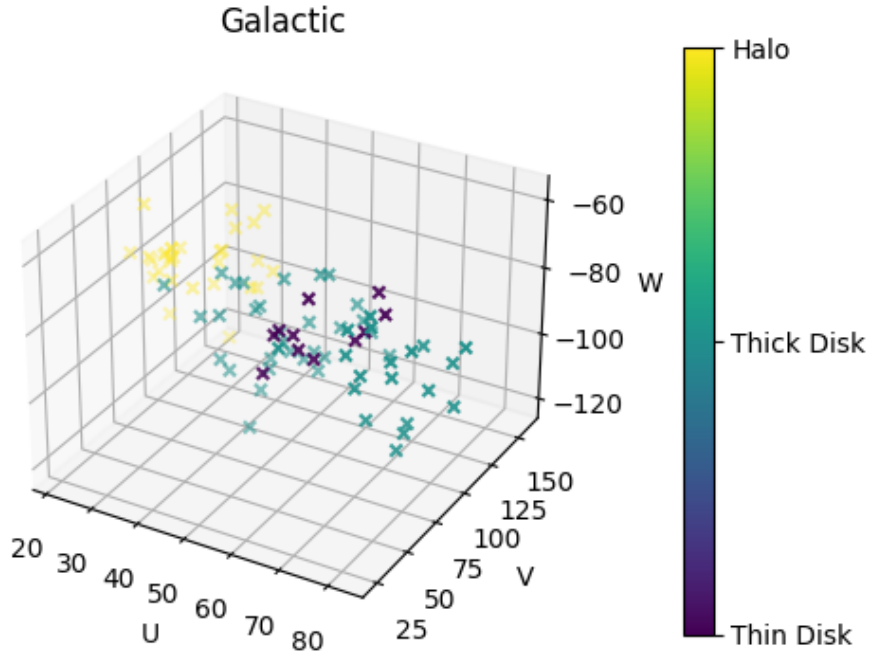
Figure 3.1: Velocity realizations for one star color-coded by candidate population. This is not the same star as shown in Figure 2.3, it was chosen as its possible velocities can correspond to all three Galactic populations. This star is also simulated and is a halo object, though it was assigned thick disk candidacy (see Section 5.1). The presence of realizations with thin disk candidacy is illustrative of the strong bias towards thin disk measurement due to stellar number densities in Table 3.2.

Any star that fails to meet one of these criteria we assume to be a thin disk star. By erring towards under-counting halo and thick disk objects, I hope to not introduce further sources of contamination and uncertainty in the resulting age estimates.

From the data quality-filtered GFC sample, I find 4446 thick disk candidate stars and six halo candidates. In UVW space we see these two populations forming shells (Figure 3.2) around the lower velocity thin disk population. Admittedly, with such a small sample significant imagination is necessary to see the halo shell though these

stars clearly lie on the edges of the population.
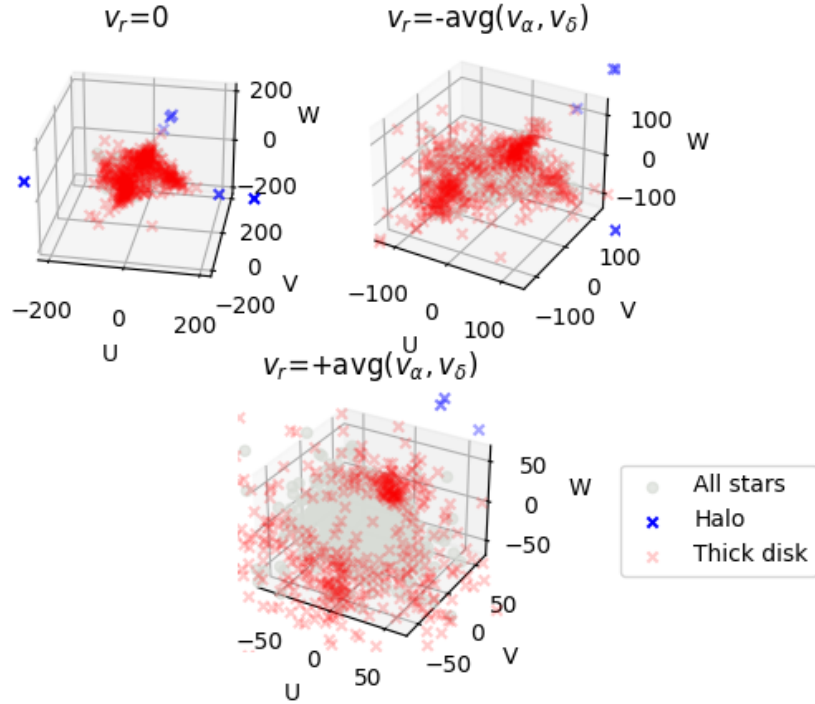
Figure 3.2: Three viewing widths of the studied white dwarf population in UVW space. Red crosses are thick disk candidates, blue crosses are halo candidates, while grey points are thin disk objects. Note that while the three plots correspond to different radial velocity estimates, the general shell morphology of these distributions is common to all three.

We have completed BASE-9 analyses of subset of the white dwarfs which are sorted into thick disk and halo populations. This includes only four of the six identified halo candidates while 294 of the identified thick disk candidates have results, and these will be the primary subjects of the analysis process I describe in the following section.

Chapter 4

HIERARCHICAL ANALYSIS

Using the method developed by Si et al. and implemented in the `HierarchicalWDs` routine [49], I describe how the population sorting leads to an age and formation history for the Galactic thick disk and halo. Technical details of the hierarchical analysis that follow are taken from this paper.

4.1   The MCMC Age-Sampling Technique

Before any sampling is done, it is necessary to gather a population of stars that share a common formation history. We have accomplished this by kinematically sorting white dwarfs into their different populations. Under this condition, there is commonality within the individual age distributions of the stars that can be extracted. We perform this analysis using a similar method to that described in §1.2.2. Ages are sampled from the distributions of each star in the population, and the "model" in this case is agreement between the stars on this age. Ages with higher levels of agreement are retained more often, leading to higher probabilities associated with these in the final population posterior distribution. Any age drawn which is not possible for all stars is rejected, but draws are kept for possible ages with low probability for certain stars. The sampler will move away from this region of age space and sample it less, translating to lower probability in the population age distribution.

4.2 Posterior Distributions and Shrinkage Plots

Upon reaching this step in the analysis, I have significantly whittled down the available white dwarf population from the GFC due to data quality requirements. Further reductions from this sample into the stars used to study each population occur due to my selection process and time constraints in running a further sub-group in BASE-9.

GFC Sample	Data Quality Cuts	Assigned Population		BASE-9 Results
359,000	104,577	Thick Disk	4446	294
		Halo	6	4

Table 4.1: Progressive reduction in the number of white dwarfs available to study the old Galactic populations.

4.2.1 Thick Disk

In literature, ages for the thick disk range between 8-10 Gyr, from a roughly 1 Gyr burst of star formation [22, 42]. My thick disk posterior distribution is shown below in Figure 4.1. We see a single, nearly symmetric peak in this distribution, representing a single burst of star formation. The shape also tells us about the length of the formation period, this appears as the width. We see a full-width at half-maximum (FWHM) of around 0.43 Gyr. The most-probable standard deviation is 0.3175 Gyr, as this represents the mean excursion of stars from the most probable age and can be positive or negative (so the population age is 5.95 ± 0.32), I find a formation time of 0.64 Gyr. My formation time implies a shorter and therefore more active starburst episode when the thick disk formed, but is not dramatically different from existing 1 Gyr measurements. More worrisome is the fact that my most probable age estimate is only 5.95 Gyr as compared to the expected 8-10 Gyr. Unfortunately for these results, but expected in the scientific process, this suggests there are significant further refinements and sources of confusion which need to be explored and addressed.The poor age agreement also implies my formation time is not likely to be valid. For reasons to be discussed in Section 5.1, I do not believe the white dwarfs used here are

entirely thick disk stars.
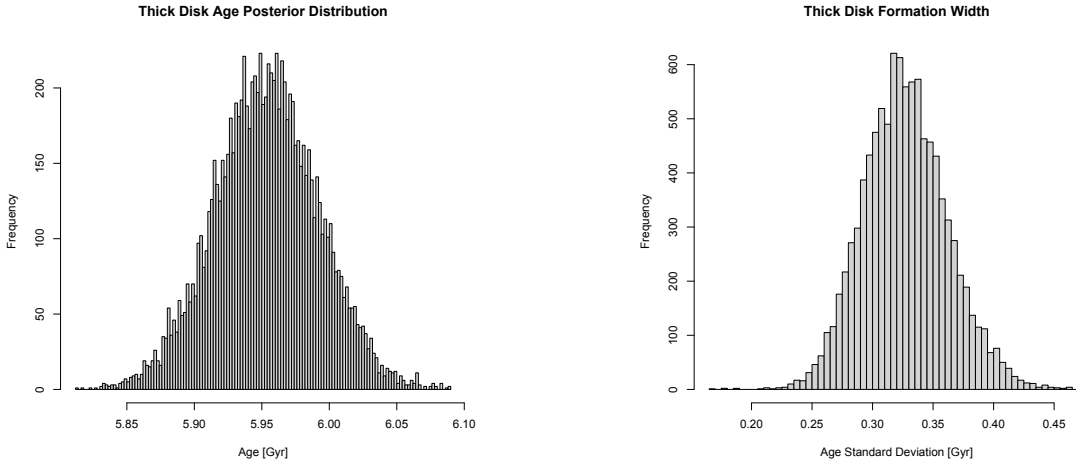
Figure 4.1: Thick disk age distribution and standard deviation in linear scale, compiled from stars with completed BASE-9 analysis (see Table 4.1). The peak of the left distribution represents the most probable age of a thick disk star, while the peak of the right distribution represents the most probable standard deviation in thick disk age.

4.2.2   Shrinkage Estimations

Hierarchical analysis further reveals more about individual stars comprising each population, limiting their allowed parameter spaces by requiring group agreement.

BASE-9 analysis results in a distribution of possible ages for each star. However, taken in the context of the greater population some of those possible ages may not be consistent with other stars which should have similar ages. The age distributions of the individual stars are then made narrower by this hierarchical analysis. This type of reduction of possible parameter space is known as a shrinkage estimation [10, 52]. While outside the goals of this work, a systematic analysis of shrinkage for thick disk stars can be performed to refine parameters of thick disk candidate stars such as age, distance, and metallicity.
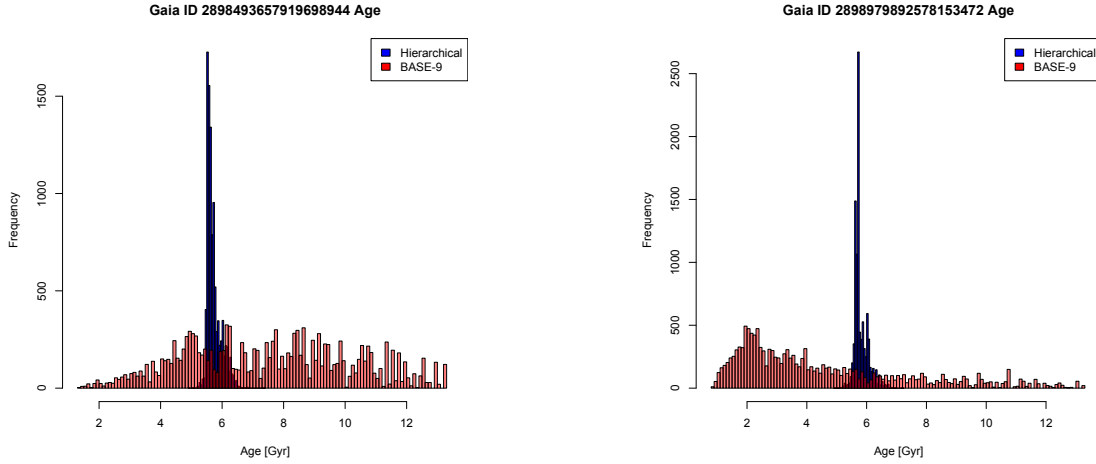
Figure 4.2: Two example shrinkage plots showing the original BASE-9 ages in pale red, and the ages allowed by the hierarchical analysis in blue.

4.2.3   A Brief Note Anticipating Analysis of the Halo

With only four successfully run white dwarfs, the results of the hierarchical analysis are going to be very sensitive to the factors discussed in Section 4.3 as well as the effects and performance discussed throughout Chapter 5. We suspect the failure of the remaining two halo stars is a sign of poor fitting of white dwarf models to these stars. This situation could arise if these are unresolved binaries which would place them outside of the white dwarf region of the color-magnitude diagram by inflating the apparent luminosity.

4.3   Factors Affecting Age Results

Unfortunately, I have thus far found a thick disk age which is inconsistent with literature. There are multiple factors which can affect this result, both within my analysis and in the star sample.

4.3.1 Initial-Final Mass Relation

White dwarfs experience crystallisation during cooling at a rate which is dependent on their mass. Higher mass stars crystallise at higher temperatures, or sooner in their evolution. The crystallisation process releases heat which can give the appearance of a younger white dwarf [55]. Furthermore, stars with an zero-age main-sequence (ZAMS) mass around 1 $M_\odot$ spend around 10 Gyr on the main sequence, or a majority of the lifetime of any such stars which have already become white dwarfs. The situation is less severe for stars with a higher ZAMS mass, but nevertheless an accurate mapping back to this mass and its associated main sequence lifetime is crucial if we hope to study Galactic ages using white dwarfs. These two factors mean that an IFMR which accounts for as much of white dwarf physics as possible is a requirement for this work.
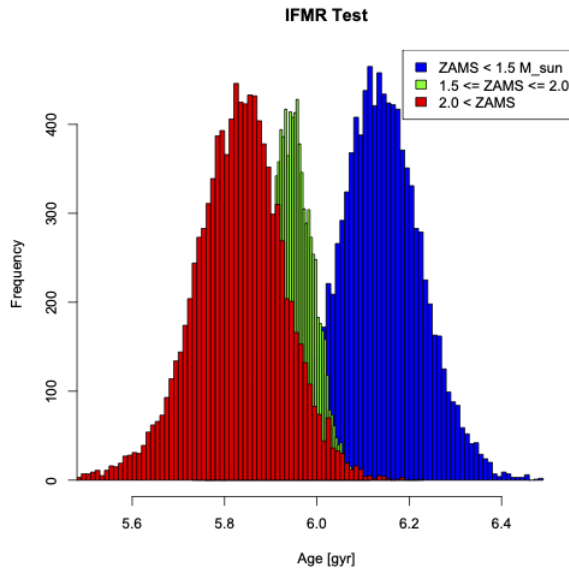
Figure 4.3: Age posterior distributions for mass-binned thick disk stars, showing a definite mass-age trend.

In order to test the sensitivity of this method to different IFMRs, I sort the thick disk candidates into three different ZAMS mass bins $0 \leq M \leq 1.5 M_\odot$, $1.5 < M \leq 2 M_\odot$, and $M > 2 M_\odot$. If the subset of white dwarfs identified as thick disk are truly members of the same population, then any subset of them should show a similar age

posterior distribution. Higher mass stars will have spent more time as white dwarfs, but overall age should be the same. If these different mass bins give different age results, that is suggestive that the choice of IFMR will affect the age results.

This is exactly what we see in Figure 4.3, with lower mass stars showing the oldest population and the higher mass bins trending towards younger populations. This is potentially indicative of incomplete treatment of the crystallisation of higher mass stars which slows their cooling, or that mass-loss during the atmosphere-shedding phase of the star is incorrect. Should this latter option be the case, the trend I find corresponds to an overestimation of evolutionary mass loss which maps the star to higher ZAMS mass and shorter main sequence lifetime.

4.3.2 White Dwarfs with Companions: Past and Present

From surveys and binary star evolution simulations, around 32% of white dwarfs occur in multiple star systems, while around half of those are expected to be in a binary system with two white dwarfs [13, 26]. Further models suggest that between 9 and 16% of white dwarfs could be the result of mergers, while 1 to 3% of white dwarfs are in unresolved binaries [54]. Among white dwarfs with masses greater than $0.8M_\odot$, the proportion which are the products of mergers may be closer to 20% [14].

Either case listed above, unresolved binary or a merger product, will cause issues with BASE-9 fitting. An unresolved binary will appear brighter and so younger and/or closer than in reality, thus BASE-9 fits are likely to sample this region of parameter space and pollute the hierarchical age. A merger product has higher mass than would be otherwise appropriate. This results in a hotter white dwarf, sooner crystallisation, and a shorter apparent main-sequence lifetime. This effect is demonstrated by the young-ward trend in higher mass bins shown in Figure 4.3.

4.3.3 Velocity Pollution

Multiple methods exist by which stars can be heated onto high-velocity orbits. Perhaps the simplest is a modern continuation of the formation processes we wish to study. Satellite galaxies to the Milky Way are still suffering tidal disruption and

young (<1 Gyr) stars from these galaxies are observed accreting into the Galactic halo [4]. Assuming the modern epoch is not unique in the occurrence of that process, then this will have resulted in stars being deposited into the halo which have had the time to evolve into white dwarfs while still being much younger than the original halo population.

White dwarf progenitors that were once in a binary with a high mass companion can obtain significant velocities after the companion experiences a supernova. Simulations show that in this scenario, upwards of 80% of systems will become unbound with the supernova remnant entering onto a high velocity orbit [9]. Conservation of momentum suggests the low-mass member of the system should also be diverted onto a high velocity orbit. Should this system have been originally in the disk and ejected to the halo, this will affect the halo age, similarly for a thin to thick disk transfer. This effect will be magnified if significant accretion onto the white dwarf progenitor also occurs, as this would result in a smaller scale version of the merger scenario.

In a recent study with Gaia data, Fleury, Caiazzo, and Heyl identify a population of local massive white dwarfs with young apparent ages and large V dispersion [20]. They suggest that the ages found for these stars from photometry are not consistent with these stars' V velocities, that they appear too young. In the high mass regime of these targets, if there are cooling and crystallisation effects which are poorly treated then these targets are potential examples of stars whose ages are being underestimated. It is also possible that these stars are remnants of a more recent merger and represent pollution in the high end of velocity space by that mechanism, though Fleury also propose a triple-system origin for these targets. Whichever of these cases this group may exhibit, they represent an example of the type of confounding white dwarf detection which we will need to further examine in refining this selection process.

Chapter 5

CONCLUSION AND SIGNIFICANCE

In this final chapter, I perform two tests designed to quantify the performance of this thesis' method. I also discuss an apparent alignment with our results and recent revelations about younger halo stars.

5.1 Analysis of Simulated Stars with Assumed Population Membership

With the aim of quantifying the reliability of the sorting algorithm, I have run tests along different lines of sight with simulated stars whose population membership is assumed to be known. They pass through the sorting process, and I record its accuracy in determining proper population membership and into which other population each of them is likely to fail.

5.1.1 Generating the Sample

Here, I make use of a variation of the MC method. Using the mean velocities and velocity dispersions from Table 3.1, 500 stars are generated belonging to each of the thin disk, thick disk, and halo. Since the simulated populations will be of identical

size, I must adjust the population number densities appropriately to reflect this. In the code listed in Appendix A.5, I have set these to unity. The probability normalization ensures any number density will work, provided it is the same across all three populations. These simulated stars fall in UVW-space on Gaussians centered at the mean velocity for their population with widths determined by the respective measured velocity dispersions. We invert Equation 2.5

$$\begin{bmatrix} v_r \\ v_\alpha \\ v_\delta \end{bmatrix} = B^{-1} \begin{bmatrix} U \\ V \\ W \end{bmatrix} \tag{5.1}$$

where B^{-1} is the matrix inverse of B (defined by Equation 2.4), to generate the corresponding equatorial velocity vectors for each star. We then remove the radial velocity from these vectors and convert the remaining right ascension and declination velocities into proper motions according to Equation 2.1. These proper motions are combined with assumed equatorial sky position (right ascension and declination), distance, and signal-to-noise ratio (SNR, a measurement of how significant measurement uncertainty is to the measured value) to simulate an input of measurements of 1500 stars that would come from Gaia. For tracking purposes, each star is given an ID number whose thousands digit; 1, 2, or 3, indicates assumed membership in the thin disk, thick disk, or halo populations respectively. No other information is saved and passed to the sorting algorithm which identifies the assumed membership.

5.1.2 Conditions

As stated above, the white dwarf sample used to test the performance of the sorting algorithm has parameters set by us and held constant across all stars in each line of sight sample. We choose multiple lines of sight to test. In this initial investigation, these are nine significant locations on the sky in the equatorial and Galactic coordinate systems. These are listed with locations in both systems in Table 5.2. We also set a distance or parallax for the sample stars, using 90 pc for the results to follow. Finally, we choose the SNR for parallax and proper motions to be ten for all stars. While

the direction on the sky is varied, on each pointing all parameters are held invariant across the simulated stars so that all variations in performance are due to the different velocities and measurements of the standards used for each population.

Location	α	δ	l	b
North Celestial Pole (NCP)	0^h	$90°$	$122.93°$	$27.13°$
North Galactic Pole (NGP)	$12^h51^m26.27^s$	$27.13°$	$0°$	$90°$
South Galactic Pole (SGP)	$0^h51^m26.27^s$	$-27.13°$	$0°$	$-90°$
Galactic Center (GC)	$17^h45^m37.20^s$	$-28.94°$	$0°$	$0°$
Galactic Anticenter (GAC)	$5^h45^m37.20^s$	$28.94°$	$180°$	$0°$
$90°$ east of GC[1] (GCE)	$21^h12^m1.05^s$	$48.33°$	$90°$	$0°$
$45°$ east of GC (GCNE)	$19^h13^m42.62^s$	$10.72°$	$45°$	$0°$
$45°$ west of GC (GCNW)	$14^h31^m55.23^s$	-60.50	$-45°$	$0°$
$90°$ west of GC (GCW)	$9^h12^m1.05^s$	$-48.33°$	$-90°$	$0°$

Table 5.2: Chosen significant locations on the sky to test the sorting algorithm.

5.1.3 Performance Results

Performance is reported as the percentage of stars the sorting algorithm assigns correctly. We also report the percentage of stars which are sorted into the incorrect population. These are calculated by counting the number of stars (which are marked as noted in §5.1.2) and dividing by the number of simulated stars placed into each population. Each percentage listed right of a population refers to the portion of that assumed population which gets assigned to each of the possible populations. Of particular interest will be lines of sight in which a significant portion of stars are either sorted correctly, or, more alarmingly, sorted into a particular incorrect population. I identify these as green (if greater than 60% correct) and red (if greater than 40% incorrect) respectively.

---

[1]Here east refers to the direction of increasing Galactic longitude l

Line of Sight	Actual Population	Assigned Population		
		Thin Disk	Thick Disk	Halo
NCP	Thin Disk	38.6	51.8	9.6
	Thick Disk	13.4	71.4	15.2
	Halo	10.8	45.4	43.8
NGP	Thin Disk	36.0	48.4	15.6
	Thick Disk	26.0	50.6	23.4
	Halo	11.2	26.6	62.2
SGP	Thin Disk	3.2	32.4	64.4
	Thick Disk	6.4	18.4	75.2
	Halo	11.6	63.8	24.6
GC	Thin Disk	6.6	51.2	42.2
	Thick Disk	1.6	5.8	92.6
	Halo	15.2	44.4	40.4
GAC	Thin Disk	34.8	54.2	11.0
	Thick Disk	14.6	63.0	22.4
	Halo	9.0	87.6	3.4
GCE	Thin Disk	50.2	49.4	0.4
	Thick Disk	3.0	37.6	59.4
	Halo	20.6	65.2	14.2
GCNE	Thin Disk	35.0	61.4	3.6
	Thick Disk	3.4	14.2	82.4
	Halo	13.8	41.2	45.0
GCNW	Thin Disk	6.6	37.8	55.6
	Thick Disk	3.0	22.0	75.0
	Halo	5.0	37.4	57.6
GCW	Thin Disk	76.8	0.4	22.8
	Thick Disk	93.8	0.8	5.4
	Halo	7.4	3.4	89.2

Table 5.3: Performance results of the velocity-derived sorting algorithm presented. Green highlighted cells identify populations which were identified correctly at a rate greater than 60%. Red cells identify populations identified incorrectly at a rate greater than 40%.

There are many examined lines of sight where our sorting method appears to perform unreliably. Of particular concern are directions in which younger disk population stars are highly likely to be sorted into the halo candidate bin. However, the

relatively superior performance of the directions tangential to the solar orbit (GCE and GCW) suggests that there may be sectors in which we will be more successful in determining the correct population. The direct anticenter direction (GAC) did not perform impressively though, so it is not likely that the entire semicircle in the Galactic plane outside of the solar orbit is useful. Further testing in this region, and beyond the Galactic plane is therefore necessary for us to determine where we can use this sorting method and so to generate additional filtering criteria to screen only for white dwarfs in those regions of the sky.

5.2   Uniformity Tests

The importance of testing the distribution of target objects is revealed when we consider how the surveys which we use to conduct Galactic archaeology might be improved and expanded. If we can say that a great number of target objects lie very near their detection limit, then it may be inspiring to conduct a deeper and more sensitive survey to improve measurements on distant faint objects and to bring more potential targets into the light. Thus, while we have shown that the sorting and aging method presented here has its shortcomings, by testing the nearby white dwarf distribution we can estimate whether or not the successor to Gaia EDR3 will improve the aging results we have found.

5.2.1   The V/V_{max} Test

The V/V_{max} test was first developed in 1968 to characterize the distribution of radio sources. The ratio is defined as the volume inside the radius of the object of interest divided by that within its maximum detectable range as measured from our instrumentation [46]. While in the case of one source, this is little more than a number, its value across a population is revealing about the distribution of that population. It is known that for a population, if the source distribution is uniform then the average value of V/V_{max} is $1/2$ [41]. In fact, it can be shown this value holds for any

accurate model of source distribution. Thus, when I say V/V_{max} test, this will actually refer to the value of that measurement's average over all stars in a sample population, $\langle V/V_{max} \rangle$. Using the formalism of Qin and Xie ([41]), we can show that if white dwarfs in the GFC are concentrated close to the Sun then the V/V_{max} test yields a value less than $1/2$. If they are found more at the edge of their visibility to Gaia then the V/V_{max} test yields a value greater than $1/2$. Following Qin and Xie, let $\Phi(M, z)$ be the "luminosity function" which describes the spatial distribution in distance from the observed, z, of sources with absolute magnitude M. Then the total number $N(M, z)$ of sources with magnitude M within a distance z is

$$N(M, z) = \int_0^z \Phi(M, z') dV(z') \tag{5.2}$$

However, for the purposes of illuminating the V/V_{max} test, we can regard the luminosity function to be replaced by the normal number density $n(z)$ with no magnitude dependence. We then construct an expectation value equation for the fraction of stars across all magnitudes within a distance z. This will be a sum of the ratios of stars of a given magnitude within a given distance to the total number of stars of that magnitude within the maximum distance of interest, weighted by the stellar density out to z, and finally divided by the total number of stars. With discrete sums now being continuous,

$$\langle N/N(z) \rangle = \frac{\int_{(M)} \int_0^z n(z) N(M, z')/N(M, z) dV(z') dM}{\int_{(M)} \int_0^z n(z') dV(z') dM} \tag{5.3}$$

Without specifying the form of $n(z)$ (or even requiring Φ not be a luminosity function) it is trivial to show that the value of Equation 5.3 is indeed $1/2$, *provided* that Equation 5.2 is true. In the case of uniform stellar density however, the total number of stars within a volume is simply proportional to that volume. Hence,

$$\langle N/N(z) \rangle = \langle V/V(z) \rangle \tag{5.4}$$

and we need merely extend the limits of the volume integral to the maximum detectable distance for a star, z_{max}, to transform $N(z)$ and $V(z)$ in the denominators into N_{max} and V_{max} (this final note is true even for nonuniform distributions).

If Equation 5.2 is not true (meaning the assumed source distribution is incorrect), then the value of Equation 5.3 is not $1/2$, and how it differs depends on how the source distribution was misidentified. If Equation 5.2 under-counts sources at low z', then the value of Equation 5.3 will be less than $1/2$ since $N(z')/N(z)$ is also necessarily small at low z'. The opposite is true if under-counting sources at high z', assuming of course that the correct total number of objects is still found out to z_{max} as will be the case when discretely counting this measurement for individual stars. Crucially, we are testing the white dwarf *detections*, which may differ from the true distribution. If we find they are clustered near to the Earth, that may imply that quality cuts made by us and in the GFC systematically reduce the limiting magnitude of Gaia and so shrink the sphere of detectable white dwarfs.

5.2.2 Thick Disk Candidates from the Gentille-Fusillo Catalog

We have constructed a grid of different evolutionary tracks for cooling white dwarfs. From this grid, we have a mapping from mass and age to absolute magnitude in the Gaia G band. Using the most probable mass and age from allowed BASE-9 runs as target points, we conduct a two-dimensional cubic$-$spline interpolation on each star's coordinates in this space to estimate their G absolute magnitudes. As an experiment, I also ran this test using the average age and mass values. This resulted in a value for the V/V_{max} test which was ten percent lower than that given later in this section. The large discrepancy serves to highlight the asymmetric nature of posterior distributions coming from MCMC analysis. I will only discuss results here which use the peak of the posterior distributions for mass and age, as these represent the most probable value of that parameter respectively. These are calculated as the centerpoint of the most frequent of 100 bins evenly dividing the range of each parameter. Continuing, we again use the limiting magnitude for Gaia detections in the G band, $G_{lim} = 20.7$ [25], we construct the maximum detectable distance d_{max} from the distance modulus

equation

$$\log_{10} d_{max} \text{ [pc]} = 1 + \frac{G_{lim} - G}{5} \tag{5.5}$$

We take the true distance to each star as the parallax listed in the GFC. Then the value of V/V_{max} for an individual star is calculated according to

$$V/V_{max} = \left(\frac{1}{\overline{\omega} d_{max}}\right)^3 \tag{5.6}$$

while the average is computed from the whole sample of stars used. For the thick disk candidate sample with BASE-9 analysis, we find $V/V_{max} = 0.17976$. Qin and Xie suggest a criterion of

$$\left|\langle V/V_{max}\rangle - \frac{1}{2}\right| < \frac{1}{\sqrt{12 N_{total}}} \tag{5.7}$$

for a density function to fit acceptably. In our case, we find

$$\left|\langle V/V_{max}\rangle - \frac{1}{2}\right| = 0.32024 \quad \text{and} \quad \frac{1}{\sqrt{12 N_{total}}} = 0.01710 \tag{5.8}$$

This does indeed suggest that our quality requirements are limiting the region of space in which we can find white dwarfs with acceptable precision in their photometry. Said differently, our useful sources are not evenly distributed within the theoretical region of space in which Gaia can detect them. Let us define two quantities; α a number which describes the degree to which our white dwarf distribution is nonuniform, and d'_{max} the limiting distance for these stars which gives $V/V_{max} = 1/2$. These are

$$\alpha = 2\langle V/V_{max}\rangle \quad \text{and} \quad \langle (d/d'_{max})^3\rangle = 1/2 \tag{5.9}$$

Let us assume this failure to fit is due to a systematic effect in making the measurement for all stars. Then we can consider any star, since the correction will be the same. We have

$$d'_{max} = \alpha^{1/3} d_{max} = \alpha^{1/3} 10^\wedge \left[1 + \frac{G_{lim} - G}{5}\right] \tag{5.10}$$

from which we are inspired to define the effective limiting magnitude

$$10^\wedge \left[\frac{G'_{lim}}{5}\right] = \alpha^{1/3} 10^\wedge \left[\frac{G_{lim}}{5}\right] \tag{5.11}$$

from which we find the effective magnitude limit to be

$$G'_{lim} = G_{lim} + \frac{5}{3}\log_{10}\alpha = 19.96 \tag{5.12}$$

for this study, indicating that we are not eliminating a significant part of the dimmest end of Gaia's detection where the oldest white dwarfs lie. While data quality varies between stars, we can regard this new limit as gathering our data quality requirements so as to predict their effect on the available white dwarf population from Gaia.

5.2.3 V/V_{max} for 10 Gyr White Dwarfs or: How Likely are we to Detect the Oldest Stars in our Galaxy?

Unfortunately, there are no >10 Gyr stars in the sample that produced the V/V_{max} test result above, so the question in the section title cannot be answered directly at this time. However, under the assumption that our data quality requirements will similarly affect these stars (or, rather, in the absence of evidence to the contrary), we can borrow this result and make representative estimates of the number of visible halo stars nearby.

Under the above assumption, the V/V_{max} test should carry on from the thick disk to the oldest stars, the halo stars. We have the halo stellar density in Table 3.2. All that remains is to choose a suitable volume. For this, we consult Figure 1.1 and the calculations underlying it. In a spherical region (a reasonable approximation as we must stay well inside one thin disk scale height), the estimated number of detectable halo stars within our chosen distance d is estimated as

$$N_{halo} \approx \langle V/V_{max}\rangle \left(\frac{4\pi}{3}d^3\right) n_{halo} \tag{5.13}$$

Choosing $d = 60$ pc as including many younger model stars while not including

the unnecessary > 14 Gyr models, we estimate there should be approximately 43 halo stars within this radius with acceptable photometry. However, there are two important caveats to this estimate. First, these stars may exist according to the statistic, but they may be in a part of the sky where their velocities are not well selected as halo objects by our method (see Table 5.3). Second, it has been reported that radio sources, for which the V/V_{max} test was first devised, show an increasing trend of that statistic with increasing luminosity of the selected group of targets [47]. White dwarfs may follow that pattern, and thus I will have overestimated by using the measurement from a younger and more luminous population.

5.3 Young Stars on Hot Orbits

As more and more extensive photometric and astrometric surveys are performed (*e.g.* Gaia [25], SDSS [1], LAMOST [62]), more studies are being performed on high-velocity (or kinematically hot) stars including determinations of ages. Some of these studies have provided evidence that the processes discussed in Section 1.1 are still ongoing in the Milky Way. Nearby satellites of our Galaxy continue to experience disruption and they, or parts of them are now observed to be falling into the Milky Way [4]. Such stars falling from regions of much higher gravitational potential will have high-velocity (hot) orbits. These galactic disruption events are even hypothesised to be responsible for recent star-formation episodes in the halo [40].

Evidence that the Galactic halo is not exclusively old stars goes beyond direct observations in our Galaxy. Astrometric information from earlier Gaia releases and spectroscopy reveals some halo stars are not as metal-poor relative to the Sun as would be expected for 10-12 Gyr stars. Age estimates for such stars in that study put some at <7 Gyr [12]. Similar ages are seen when looking at our nearest spiral neighbor, M31. Brown *et al.* [11] used Hubble Space Telescope images to construct a color-magnitude diagram for the halo of M31. Isochrone fitting reveals multiple populations defined by metallicity. Brown *et al.* estimate that on the order of 25% of the halo of M31 is made up of stars which are younger than 8 Gyr. Such a population, if present in the Milky Way, should be apparent in our sample of white

dwarfs. By virtue of being younger and therefore brighter, those stars would be more likely to meet data quality selection criteria and could come to dominate our samples of hot-orbit stars, even if our selection algorithm is refined to better performance than indicated in Table 5.3. With reference to this table and Section 5.1, and also addressing the anomalously young thick disk age we find, despite these indicators discussed of younger stars being found in the outer "older" regions of the Galaxy we cannot at this time say the latter is responsible for the former. Systematic trends in metallicity are a recurring theme in identifying young stars in unexpected places, however there is no initial-final metallicity-relation which would allow my method to also select by ZAMS metallicity.

5.4 Summary

I have presented the development of a white dwarf velocity-selection algorithm. Upon testing it, I have found that it does not yet produce reliable results. These tests now constructed suggest ways to refine this method and will continue to provide a means of checking its performance. I present below a summary of the motivation, results obtained at this stage of research, performance, and goals for further development.

1. Galactic formation is most likely a complicated series of merger events which continue at a smaller relative scale into the present day. Evidence from ever-growing databases of large survey missions make it easier to reveal signs of these mergers. However, in the early universe more radical changes would occur in the Milky Way with each event, and we need to be able to identify the oldest stars in the halo and the intermediate-age stars that make up the thick disk. Along certain lines of sight, the sorting method discussed here performs well at distinguishing these populations. While we are disappointed it does not perform well across the whole sky, these tests have revealed where we can rely upon it and inspires the development of new filters to apply with the goal to refine and expand beyond included results for the thick disk.

2. We find a most probable age for the thick disk of 5.95 Gyr, with most of its star

formation happening within less than 1 Gyr. The formation time is consistent with literature, but the age is not. Combined with the results of performance tests of our method at the time these thick disk stars were analysed, we do not at this time believe these results are valid. They still serve to demonstrate my population-selection process.

3. We have performed tests on the accuracy of our selection process in identifying correct population membership and on the distribution and detectability of older white dwarfs. Current accuracy is not adequate for uncovering new results at this time. Results of these tests to date, and those to come will suggest new filtering of which GFC white dwarfs we can analyse. I further predict the best application of sky-position filters will vary with each target population. We also estimate that there are less than 50 true halo stars within 60 pc which meet our current selection criteria. This number is more still than we will ultimately find useful once additional filters are implemented. If we can refine and validate this sorting method, then when additional Gaia data releases and potentially more sensitive instruments become available and expand our sphere of white dwarf visibility, then the development we have performed for this technique positions us well for quick application to the analysis of new data.

4. The primary area of our future work will be in designing and implementing further positional filters for white dwarfs. Additional pointings to those listed in Table 5.2 as well as more distances will be trialed to examine the performance of our selection method. These tests will include more simulated stars.

   We will also revisit the transformation and population probability assignment steps in detail to search for increases in efficiency and any sources of ambiguity or loss of accuracy. Ultimately, we hope to implement these refinements in order to get reliable results for the Milky Way thick disk. We will then be ready to examine additional white dwarf observations to perform the sought after analysis on the Galactic halo.

Bibliography

[1] Ahumada, R. et al. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. *The Astrophysical Journal Supplement Series*, 249(1):3, July 2020. doi: 10.3847/1538-4365/ab929e.

[2] Althaus, L.G. and Benvenuto, O.G. Evolution of DA white dwarfs in the context of a new theory of convection. *Monthly Notices of the Royal Astronomical Society*, 296(1):206–216, May 1998. doi: 10.1046/j.1365-8711.1998.01332.x.

[3] Bédard, A. et al. On the Spectral Evolution of Hot White Dwarf Stars. I. A Detailed Model Atmosphere Analysis of Hot White Dwarfs from SDSS DR12. *The Astrophysical Journal*, 901(2):93, Oct. 2020. doi: 10.3847/1538-4357/abafbe.

[4] Bellazzini, M. et al. Young stars raining through the galactic halo: the nature and orbit of Price-Whelan 1. *Monthly Notices of the Royal Astronomical Society*, 490(2):2588–2598, Dec. 2019. doi: 10.1093/mnras/stz2788.

[5] Bennett, C.L. et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Are There Cosmic Microwave Background Anomalies? *The Astrophysical Journal Supplement Series*, 192(2):17, Feb. 2011. doi: 10.1088/0067-0049/192/2/17.

[6] Bensby, T. and Feltzing, S. The origin and chemical evolution of carbon in the Galactic thin and thick discs*. *Monthly Notices of the Royal Astronomical Society*, 367(3):1181–1193, Apr. 2006. doi: 10.1111/j.1365-2966.2006.10037.x.

[7] Bergeron, P., Wesemael, F. and Beauchamp, A. Photometric Calibration of Hydrogen- and Helium-Rich White Dwarf Models. *Publications of the Astronomical Society of the Pacific*, 107:1047, Nov. 1995. doi: 10.1086/133661.

[8] Blaauw, A. et al. The new I. A. U. system of galactic coordinates (1958 revision). *Monthly Notices of the Royal Astronomical Society*, 121:123, Jan. 1960. doi: 10.1093/mnras/121.2.123.

[9] Brandt, N. and Podsiadlowski, P. The effects of high-velocity supernova kicks on the orbital properties and sky distributions of neutron-star binaries. *Monthly Notices of the Royal Astronomical Society*, 274(2):461–484, May 1995. doi: 10.1093/mnras/274.2.461.

[10] Brown, L.D. and Zhao, L.H. A Geometrical Explanation of Stein Shrinkage. *Statistical Science*, 21(1):24–30, 2012. doi: 10.1214/11-STS382.

[11] Brown, T.M. et al. Evidence of a Significant Intermediate-Age Population in the M31 Halo from Main-Sequence Photometry. *The Astrophysical Journal Letters*, 592(1):L17–L20, July 2003. doi: 10.1086/376935.

[12] Caffau, E. et al. High-speed stars: Galactic hitchhikers. *Astronomy & Astrophysics*, 638:A122, June 2020. doi: 10.1051/0004-6361/202038057.

[13] Camacho, J., Torres, S. and García-Berro, E. Monte Carlo simulations of the binary white dwarf population: A progress report. In *Journal of Physics Conference Series*, volume 172 of *Journal of Physics Conference Series*, page 012030, June 2009. doi: 10.1088/1742-6596/172/1/012030.

[14] Cheng, S. et al. Double White Dwarf Merger Products among High-mass White Dwarfs. *The Astrophysical Journal*, 891(2):160, Mar. 2020. doi: 10.3847/1538-4357/ab733c.

[15] Curtis-Lake, E. et al. Spectroscopic confirmation of four metal-poor galaxies at z = 10.3-13.2. *Nature Astronomy*, 7:622–632, May 2023. doi: 10.1038/s41550-023-01918-w.

[16] Diao, J. et al. The impact of the dark matter on galaxy formation. In *Journal of Physics Conference Series*, volume 2441 of *Journal of Physics Conference Series*, page 012025, Mar. 2023. doi: 10.1088/1742-6596/2441/1/012025.

[17] Du, C. et al. Galactic structure studies from the Beijing-Arizona-Taiwan-Connecticut survey. *Monthly Notices of the Royal Astronomical Society*, 372 (3):1304–1314, Nov. 2006. doi: 10.1111/j.1365-2966.2006.10940.x.

[18] Du, C. et al. The High-velocity Stars in the Local Stellar Halo from Gaia and LAMOST. *The Astrophysical Journal*, 863(1):87, Aug. 2018. doi: 10.3847/1538-4357/aad088.

[19] Eggen, O.J., Lynden-Bell, D. and Sandage, A.R. Evidence from the motions of old stars that the Galaxy collapsed. *The Astrophysical Journal*, 136:748, Nov. 1962. doi: 10.1086/147433.

[20] Fleury, L., Caiazzo, I. and Heyl, J. The origin of ultramassive white dwarfs: hints from Gaia EDR3. *Monthly Notices of the Royal Astronomical Society*, 520 (1):364–374, Mar. 2023. doi: 10.1093/mnras/stad068.

[21] Gentile Fusillo, N.P. et al. A catalogue of white dwarfs in Gaia EDR3. *Monthly Notices of the Royal Astronomical Society*, 508(3):3877–3896, Dec. 2021. doi: 10.1093/mnras/stab2672.

[22] Gratton, R.G. et al. Abundances of light elements in metal-poor stars. IV. [Fe/O] and [Fe/Mg] ratios and the history of star formation in the solar neighborhood. *Astronomy & Astrophysics*, 358:671–681, June 2000. doi: 10.48550/arXiv.astro-ph/0004157.

[23] Hannestad, S. et al. Neutrino and axion hot dark matter bounds after WMAP-7. *Journal of Cosmology and Astrophysics*, 2010(8):001, Aug. 2010. doi: 10.1088/1475-7516/2010/08/001.

[24] Hillebrandt, W., Nomoto, K. and Wolff, R.G. Supernova explosions of massive stars - The mass range 8 to 10 solar masses. *Astronomy & Astrophysics*, 133(1): 175–184, Apr. 1984.

[25] Hodgkin, S.T. et al. Gaia Early Data Release 3. Gaia photometric science alerts. *Astronomy & Astrophysics*, 652:A76, Aug. 2021. doi: 10.1051/0004-6361/202140735.

[26] Holberg, J.B. What fraction of white dwarfs are members of binary systems? In *Journal of Physics Conference Series*, volume 172 of *Journal of Physics Conference Series*, page 012022, June 2009. doi: 10.1088/1742-6596/172/1/012022.

[27] Howes, L.M. et al. Estimating stellar ages and metallicities from parallaxes and broadband photometry: successes and shortcomings. *Astronomy & Astrophysics*, 622:A27, Feb. 2019. doi: 10.1051/0004-6361/201833280.

[28] IBM. What is monte carlo simulation? URL https://www.ibm.com/topics/monte-carlo-simulation.

[29] Jofré, P. and Weiss, A. The age of the Milky Way halo stars from the Sloan Digital Sky Survey. *Astronomy & Astrophysics*, 533:A59, Sept. 2011. doi: 10.1051/0004-6361/201117131.

[30] Johnson, D.R.H. and Soderblom, D.R. Calculating Galactic Space Velocities and Their Uncertainties, with an Application to the Ursa Major Group. *Astronomical Journal*, 93:864, Apr. 1987. doi: 10.1086/114370.

[31] Katz, D. et al. Gaia Data Release 3. Properties and validation of the radial velocities. *Astronomy & Astrophysics*, 674:A5, June 2023. doi: 10.1051/0004-6361/202244220.

[32] Kereš, D. et al. How do galaxies get their gas? *Monthly Notices of the Royal Astronomical Society*, 363(1):2–28, Oct. 2005. doi: 10.1111/j.1365-2966.2005.09451.x.

[33] Kilic, M. et al. The age of the Galactic stellar halo from Gaia white dwarfs. *Monthly Notices of the Royal Astronomical Society*, 482(1):965–979, Jan. 2019. doi: 10.1093/mnras/sty2755.

[34] Latyshev, I.N. Star density in the solar neighborhood. *Soviet Astronomy*, 22: 186–188, Apr. 1978.

[35] Majewski, S.R. Precision astrometry, galactic mergers, halo substructure and local dark matter. In Jin, W.J., Platais, I. and Perryman, M.A.C., editors, *A Giant Step: from Milli- to Micro-arcsecond Astrometry*, volume 248, pages 450–457, July 2008. doi: 10.1017/S1743921308019790.

[36] Menci, N. et al. A Stringent Limit on the Warm Dark Matter Particle Masses from the Abundance of z = 6 Galaxies in the Hubble Frontier Fields. *The Astrophysical Journal Letters*, 825(1):L1, July 2016. doi: 10.3847/2041-8205/825/1/L1.

[37] Miller, G.E. and Scalo, J.M. The Initial Mass Function and Stellar Birthrate in the Solar Neighborhood. *The Astrophysical Journal Supplement Series*, 41:513, Nov. 1979. doi: 10.1086/190629.

[38] Montgomery, M.H. et al. Evolutionary Calculations of Phase Separation in Crystallizing White Dwarf Stars. *The Astrophysical Journal*, 525(1):482–491, Nov. 1999. doi: 10.1086/307871.

[39] O'Malley, E.M., von Hippel, T. and van Dyk, D.A. A Bayesian Approach to Deriving Ages of Individual Field White Dwarfs. *The Astrophysical Journal*, 775(1):1, Sept. 2013. doi: 10.1088/0004-637X/775/1/1.

[40] Price-Whelan, A.M. et al. Discovery of a Disrupting Open Cluster Far into the Milky Way Halo: A Recent Star Formation Event in the Leading Arm of the Magellanic Stream? *The Astrophysical Journal*, 887(1):19, Dec. 2019. doi: 10.3847/1538-4357/ab4bdd.

[41] Qin, Y.P. and Xie, G.Z. A Generalization of the V/V_{max} Test. *The Astrophysical Journal*, 486(1):100–101, Sept. 1997. doi: 10.1086/304525.

[42] Reid, I.N. High-Velocity White Dwarfs and Galactic Structure. *Annual Review of Astronomy & Astrophysics*, 43(1):247–292, Sept. 2005. doi: 10.1146/annurev.astro.43.072103.150623.

[43] Renedo, I. et al. New Cooling Sequences for Old White Dwarfs. *The Astrophysical Journal*, 717(1):183–195, July 2010. doi: 10.1088/0004-637X/717/1/183.

[44] Salaris, M. et al. Semi-empirical White Dwarf Initial-Final Mass Relationships: A Thorough Analysis of Systematic Uncertainties Due to Stellar Evolution Models. *The Astrophysical Journal*, 692(2):1013–1032, Feb. 2009. doi: 10.1088/0004-637X/692/2/1013.

[45] Schive, H.Y. et al. Contrasting Galaxy Formation from Quantum Wave Dark Matter, ψDM, with ΛCDM, using Planck and Hubble Data. *The Astrophysical Journal*, 818(1):89, Feb. 2016. doi: 10.3847/0004-637X/818/1/89.

[46] Schmidt, M. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *The Astrophysical Journal*, 151:393, Feb. 1968. doi: 10.1086/149446.

[47] Schmidt, M. and Green, R.F. Quasar evolution derived from the Palomar bright quasar survey and other complete quasar surveys. *The Astrophysical Journal*, 269:352–374, June 1983. doi: 10.1086/161048.

[48] Seifahrt, A. et al. On the kinematic age of brown dwarfs: radial velocities and space motions of 43 nearby L dwarfs. *Astronomy & Astrophysics*, 512:A37, Mar. 2010. doi: 10.1051/0004-6361/200913368.

[49] Si, S. et al. A hierarchical model for the ages of Galactic halo white dwarfs. *Monthly Notices of the Royal Astronomical Society*, 468(4):4374–4388, July 2017. doi: 10.1093/mnras/stx765.

[50] Skuljan, J., Hearnshaw, J.B. and Cottrell, P.L. Velocity distribution of stars in the solar neighbourhood. *Monthly Notices of the Royal Astronomical Society*, 308(3):731–740, Sept. 1999. doi: 10.1046/j.1365-8711.1999.02736.x.

[51] Soubiran, C. and Girard, P. Abundance trends in kinematical groups of the Milky Way's disk. *Astronomy & Astrophysics*, 438(1):139–151, July 2005. doi: 10.1051/0004-6361:20042390.

[52] Stein, C. Inadmissability of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 3:197–206, Jan. 1956.

[53] Tonry, J.L. et al. The Pan-STARRS1 Photometric System. *The Astrophysical Journal*, 750(2):99, May 2012. doi: 10.1088/0004-637X/750/2/99.

[54] Torres, S. et al. A population synthesis fitting of the Gaia resolved white dwarf binary population within 100 pc. *Monthly Notices of the Royal Astronomical Society*, 511(4):5462–5474, Apr. 2022. doi: 10.1093/mnras/stac374.

[55] van Horn, H.M. Crystallization of White Dwarfs. *The Astrophysical Journal*, 151:227, Jan. 1968. doi: 10.1086/149432.

[56] Vieira, K. et al. Milky Way Thin and Thick Disk Kinematics with Gaia EDR3 and RAVE DR5. *The Astrophysical Journal*, 932(1):28, June 2022. doi: 10.3847/1538-4357/ac6b9b.

[57] von Hippel, T. et al. Inverting Color-Magnitude Diagrams to Access Precise Star Cluster Parameters: A Bayesian Approach. *The Astrophysical Journal*, 645(2):1436–1447, July 2006. doi: 10.1086/504369.

[58] von Hippel, T. et al. Bayesian Analysis for Stellar Evolution with Nine Parameters (BASE-9): User's Manual. *arXiv e-prints*, art. arXiv:1411.3786, Nov. 2014. doi: 10.48550/arXiv.1411.3786.

[59] Weidemann, V. Revision of the initial-to-final mass relation. *Astronomy & Astrophysics*, 363:647–656, Nov. 2000.

[60] White, S.D.M. and Rees, M.J. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183:341–358, May 1978. doi: 10.1093/mnras/183.3.341.

[61] Wood, M.A. Constraints on the Age and Evolution of the Galaxy from the White Dwarf Luminosity Function. *The Astrophysical Journal*, 386:539, Feb. 1992. doi: 10.1086/171038.

[62] Yan, H. et al. Overview of the LAMOST survey in the first decade. *The Innovation*, 3:100224, Mar. 2022. doi: 10.1016/j.xinn.2022.100224.

Appendix A

Important Codes Developed and/or Used for Analysis in this Thesis

In this appendix, I provide the main codes used to perform analysis described in this thesis. Some, which perform simple tasks such as sorting files are omitted. All included here are written by me with the exception of that in Appendix A.1. Contributions and inspirations are listed in the text at the beginning of each appendix section. Comments are indicated in green text, strings in purple, and code keywords in magenta. Red arrows at the beginning of each line indicate where a line of code has wrapped from the line above.

A.1 Hierarchical Analysis

This code is written in R and is the only one included not written in Python. It is a slightly modified version of the one provided to me through the BASE-9 GitHub, this done with the advice of Elliot Robinson, to take command line arguments. This code performs hierarchical analysis using the MCMC technique described in Chapter 4 when provided a list of BASE-9 white dwarf results.

HierarchicalWDs.R

```
1  library(foreach)
2  library(pscl)
```

```r
3  library(truncnorm)
4
5  # monte carlo EM algorithm to fit hierarchical model of logAge
6  MHFB = function(starAges, steps, chainDepth) {
7    minLogAge = 9.5
8    maxLogAge = 10.176
9
10   nStars = ncol(starAges)
11
12   sampledAges      = matrix(0, steps, nStars)
13   sampledAges[1,] = starAges[1,] # sampledAges starts with a random draw
       ↪   from the ages. Row 1 is randomly chosen.
14
15   tauSquared.scale = sum((sampledAges[1,] - mean(sampledAges[1,])) ^ 2)
16
17   # Mean of inverse gamma is b / (a - 1) for a > 1
18   tauSquared = tauSquared.scale / (nStars - 1)
19   gamma        = rtruncnorm(1, minLogAge, maxLogAge, mean(sampledAges[1,])
       ↪   , sqrt(tauSquared[1] / nStars)) # optimal gamma
20
21   for(step in 2:steps) {
22     last.sampledAge = sampledAges[step - 1,]
23     last.gamma      = gamma[step - 1]
24     last.tau        = sqrt(tauSquared[step - 1])
25
26     for(iter in 1:chainDepth) {
27       proposedAge = NULL
28
29       # Randomly draw one age for each star
30       for(star in 1:nStars) {
31         proposedAge[star] = sample(starAges[,star], 1)
32       }
33
34       probabilityRatio = dnorm(proposedAge,      last.gamma, last.tau) /
35                          # ————————————————————————————————
36                          dnorm(last.sampledAge, last.gamma, last.tau)
37
```

```r
38        uniformDraws = runif(nStars)
39        accept       = uniformDraws <= probabilityRatio
40
41        last.sampledAge = ifelse(accept, proposedAge, last.sampledAge)
42     }
43
44     sampledAges[step,] = last.sampledAge
45
46     tauSquared.scale = sum((sampledAges[step,] - last.gamma) ^ 2)
47
48     tauSquared[step] = rigamma(1, (nStars - 1) / 2, tauSquared.scale /
     ↪ 2)
49     gamma[step] = rtruncnorm(1, minLogAge, maxLogAge, mean(sampledAges[
     ↪ step,]), sqrt(tauSquared[step] / nStars))
50   }
51
52   list("sampledAges" = sampledAges,
53        "gamma"       = gamma,
54        "tauSquared"  = tauSquared)
55 }
56
57 main = function() {
58 #resultf=commandArgs(trailingOnly=TRUE)
59 #flist=resultf[2]
60   resultFiles=commandArgs(trailingOnly=T)
61   nStars = length(resultFiles)
62
63   print(resultFiles)
64
65   starAges = foreach(i = 1:nStars, .combine = "cbind") %do% {
66     singlePopMcmcAges = read.table(resultFiles[i], header = T)$logAge
67   }
68
69   # chainDepth the chain
70   fit = MHFB(starAges, steps = 10000, chainDepth = 400)
71
72   fb.age    = fit$sampledAges
```

```
73    fb.gamma = fit$gamma
74    fb.tauSquared = fit$tauSquared
75
76    save(fb.age, fb.gamma, fb.tauSquared, file = "ApproxFBsimulation1.
      ↪ Rdata")
77 }
78
79 main()
```

A.2 GFC Filtering

This code analyzes Gaia data for White Dwarfs in the GFC. It is developed from and partly inspired by an existing `Matlab` code written by Ally Woodruff, Dr. Ted von Hippel, and Alisa Tiselska which has a similar function but uses different methods. Here, I make further requirements on parallax SNR and narrows our allowed region of the Gaia color-magnitude diagram from which we draw white dwarfs. The astrometry and photometry from surviving stars are then saved to a text file which can be processed by the code in Appendix A.3.

`filter_dr3.py`

```python
#Calculate UWW velocities of white dwarves
#Proper motions from Gaia DR3 (new column order -> new code because I'm
    not advanced enough to make one that adapts)
#Began: 16 Jan 2023
#Updated: 9 Aug 2023

import numpy as np
import astropy as astro
import matplotlib.pyplot as plt
import time
import pandas as pd

def calc_main():
#Read in parameters file
    #sample_merge1=pd.read_csv('GF_PS1_topcat2.csv',delimiter=',',
    skiprows=0)
    sample_merge1=pd.read_csv('GaiaEDR3_WD_main_PS1.txt',delimiter=' ',
    skiprows=0)
    #ll: _RAJ2000 _DEJ2000 Source RA_ICRS e_RA_ICRS DE_ICRS e_DE_ICRS
    Plx e_Plx pmRA
    #e_pmRA pmDE e_pmDE epsi amax FG e_FG Gmag FBP e_FBP
    #BPmag FRP e_FRP RPmag E(BR/RP) GLON GLAT Density AG umag
    #e_umag gmag e_gmag rmag e_rmag imag e_imag zmag e_zmag Pwd
    #f_Pwd TeffH e_TeffH loggH e_loggH MassH e_MassH chi2H TeffHe
    e_TeffHe
```

```
21    #loggHe e_loggHe MassHe e_MassHe chisqHe _RA.icrs _DE.icrs recno
  ↪ objID RAJ2000
22    #DEJ2000 errHalfMaj errHalfMin errPosAng f_objID Qual Epoch Ns Nd
  ↪ gmag_x
23    #e_gmag_x gKmag e_gKmag gFlags rmag_x e_rmag_x rKmag e_rKmag rFlags
  ↪ imag_x
24    #e_imag_x iKmag e_iKmag iFlags zmag_x e_zmag_x zKmag e_zKmag zFlags
  ↪ ymag
25    #e_ymag yKmag e_yKmag yFlags angDist
26
27    sample_merge=sample_merge1.to_numpy()
28
29    ra=sample_merge[:,2] #Python is exclusive counting at upper index
30    dec=sample_merge[:,4] #and starts from 0 like normal lol
31    source_id=sample_merge[:,0]
32    prlx=sample_merge[:,6]
33    prlx_e=sample_merge[:,7]
34    pmra=sample_merge[:,11]
35    pmra_e=sample_merge[:,12]
36    pmdec=sample_merge[:,13]
37    pmdec_e=sample_merge[:,14]
38    FG=sample_merge[:,15] #Gaia G-band mag
39    e_FG=sample_merge[:,16] #error
40    #U_est=sample_merge[:,17] #G-band mag (scale to Vega)
41    U_est=sample_merge[:,23] #G-band mag (scale to Vega)
42    FBP=sample_merge[:,19] #BP mean flux
43    e_FBP=sample_merge[:,20]
44    #BP=sample_merge[:,20]
45    FRP=sample_merge[:,21] #RP band
46    e_FRP=sample_merge[:,22]
47    #RP=sample_merge[:,23]
48    Pwd=sample_merge[:,10] #prob white dwarf
49    g_ps1=sample_merge[:,70]
50    r_ps1=sample_merge[:,72]
51    i_ps1=sample_merge[:,74]
52    z_ps1=sample_merge[:,76]
53    y_ps1=sample_merge[:,78]
```

```python
54
55        """
56        e_Gmag=1.086*e_FG/FG
57        e_BPmag=1.086*e_FBP/FBP
58        e_RPmag=1.086*e_FRP/FRP
59        """
60
61
62        e_Gmag=e_FG
63        e_BPmag=e_FBP
64        e_RPmag=e_FRP
65
66        SNR_prlx=prlx/prlx_e
67
68        #Colour filtering
69        #G_RP=U_est-RP
70        #BP_G=BP-U_est
71        G_RP=U_est-FRP
72        BP_G=FBP-U_est
73        M_g=g_ps1+5*np.log10(prlx/1000)+5
74        np.savetxt('M_g_dr3.txt',M_g)
75
76        print(M_g[0:10]-sample_merge[0:10,23])
77        print(G_RP-sample_merge[:,26])
78        print(BP_G-sample_merge[:,25])
79
80        M_g=sample_merge[:,23]
81        G_RP=sample_merge[:,26]
82        BP_G=sample_merge[:,25]
83
84        gbprp_cut=0
85        grmg=0
86        snr=0
87
88        keep_ra=[]#np.array()
89        keep_dec=[]#np.array()
90        keep_prlx=[]#np.array()
```

```
91        keep_prlx_e =[]#np.array()
92        keep_pmra =[]#np.array()
93        keep_pmra_e =[]#np.array()
94        keep_pmdec =[]#np.array()
95        keep_pmdec_e =[]#np.array()
96        keep_G =[]#np.array()
97        keep_BP =[]
98        keep_RP =[]
99        keep_source_id =[]
100       keep_g_ps1 =[]
101       keep_r_ps1 =[]
102       keep_y_ps1 =[]
103       new_id =[]
104       line_no =[]
105       k=0
106
107       for i in range(len(ra)):
108           if G_RP[i] > 0.5*(BP_G[i]-1.0)+1.1+0.1:
109           #if G_RP[i] < 0.5*(BP_G[i]-1.0)+1.1+0.1:
110               gbprp_cut+=1
111           elif (M_g[i] <= 14 and g_ps1[i]-r_ps1[i]>0.1*(M_g[i]-10)) or (
    ↪ M_g[i]>14 and g_ps1[i]-r_ps1[i]>(1/3)*(M_g[i]-12.8)):
112               grmg+=1
113           elif SNR_prlx[i] < 5:
114               snr+=1
115           else:
116               keep_ra=np.append(keep_ra,ra[i])
117               keep_dec=np.append(keep_dec,dec[i])
118               keep_prlx=np.append(keep_prlx,prlx[i])
119               keep_prlx_e=np.append(keep_prlx_e,prlx_e[i])
120               keep_pmra=np.append(keep_pmra,pmra[i])
121               keep_pmra_e=np.append(keep_pmra_e,pmra_e[i])
122               keep_pmdec=np.append(keep_pmdec,pmdec[i])
123               keep_pmdec_e=np.append(keep_pmdec_e,pmdec_e[i])
124               keep_G=np.append(keep_G,U_est[i])
125               #keep_BP=np.append(keep_BP,BP[i])
126               #keep_RP=np.append(keep_RP,RP[i])
```

```
127              keep_BP=np.append(keep_BP,FBP[i])
128              keep_RP=np.append(keep_RP,FRP[i])
129              keep_source_id=np.append(keep_source_id,source_id[i])
130              keep_g_ps1=np.append(keep_g_ps1,g_ps1[i])
131              keep_r_ps1=np.append(keep_r_ps1,r_ps1[i])
132              keep_y_ps1=np.append(keep_y_ps1,y_ps1[i])
133              new_id=np.append(new_id,1e6+i)
134              line_no=np.append(line_no,i)
135              k+=1
136
137
138      #Important elements for stars that make the final cut
139      print('Rejecting {} objects based on Gaia G,BP,RP locus\n'.format(
     ↪ gbprp_cut))
140      print('Rejecting {} objects based on WD cut in g-r vs. M_g\n'.format
     ↪ (grmg))
141      print('Rejecting {} objects based on SNR of parallax > 5\n'.format(
     ↪ snr))
142      print('Keeping {} of {} objects'.format(k,i))
143
144      np.savetxt('Keep_stars.txt',np.hstack((new_id.reshape((k,1)),keep_ra
     ↪ .reshape((k,1)),keep_dec.reshape((k,1)),keep_prlx.reshape((k,1)),
     ↪ keep_prlx_e.reshape((k,1)),keep_pmra.reshape((k,1)),keep_pmra_e.
     ↪ reshape((k,1)),keep_pmdec.reshape((k,1)),keep_pmdec_e.reshape((k
     ↪ ,1)),keep_G.reshape((k,1)),keep_BP.reshape((k,1)),keep_RP.reshape
     ↪ ((k,1)),keep_source_id.reshape((k,1)),keep_g_ps1.reshape((k,1)),
     ↪ keep_r_ps1.reshape((k,1)),keep_y_ps1.reshape((k,1)),line_no.
     ↪ reshape((k,1)))))
145
146 calc_main()
```

A.3 Determination of Population Candidacy

Using astrometry from the processed white dwarf sample, this code generates the radial velocity estimates and then the MC velocity cloud points around the three options. It is in this step that I rotate the stars' velocity realizations into UVW space and probabilistically assign population candidacy.

`transform_UVW.py`

```python
# Version 15 July 2023

import numpy as np
import astropy as astro
import matplotlib.pyplot as plt
#import emcee as mc
from time import perf_counter# as perf_counter
from datetime import timedelta
#import timeit

def main():
    start_time=perf_counter()
    filtered=np.loadtxt('Kept_stars.txt')
    #LL our ID, RA, dec, prlx, prlx_e, pmRA, pmRA_e, pmdec, pmdec_e,
    ↪ photometry stuff
    ra=np.radians(filtered[:,1])
    dec=np.radians(filtered[:,2])
    gaia_vels=filtered[:,3:8]*1e-3
    n_stars=filtered.shape[0]
    n_sim=30 #Number of stars to simulate (yeet) through whatever approx
    ↪ to MC this is
    T=np.array([[-0.06699, -0.87276, -0.48354],[0.49273, -0.45035,
    ↪ 0.74458],[-0.86760, -0.18837, 0.46020]]) #First part of rotation
    ↪ matrix from Johnson & Soderblom 1987
    A=np.zeros([3,3,n_stars])
    k=4.74057 #some conversion factor #astornomy lul
    bary_vels=np.zeros([n_stars,3,3,30]) #star #, v component, rv
    ↪ estimate, MC estimates
    gal_vels=np.zeros([n_stars,3,3,30]) #star, v component, rv est, MC
```

```
25
26      #U x3, V x3, W x3
27      mu=np.array([[0,0,0],[-12,-51,196],[0,0,0]])
28      sig=np.array([[39,67,160],[20,38,90],[16,35,90]])
29      n_den=np.array([0.124,0.0156,2.65e-4]) #stars per cubic pc
30      norms=[1/sig[0,:],1/sig[1,:],1/sig[2,:]]/(np.sqrt(2*np.pi)) #
↪       normalising factors for gaussians
31      P=np.zeros([n_stars,3,3,n_sim]) #star #, rv estimate, population, mc
32      totprobs_rv=np.zeros([n_stars,3,3]) #star,pop, rv estimate
33      totprobs=np.zeros([n_stars,3]) #star, pop
34      stdevs=np.zeros([n_stars,3,3]) #star, pop, rv
35      P_thrsh=np.array([0,0.5,0.5])
36
37      for i in range(n_stars):
38          A[:,:,i]=np.array([[np.cos(ra[i])*np.sin(dec[i]), -np.sin(ra[i])
↪           , -np.cos(ra[i])*np.sin(dec[i])],[np.sin(ra[i])*np.cos(dec[i]), np
↪           .cos(ra[i]), -np.sin(ra[i])*np.cos(dec[i])],[np.sin(dec[i]), 0, np
↪           .cos(dec[i])]])
39          bary_vels[i,0,0,:]=(1/2)*k*(filtered[i,6]+filtered[i,8])*np.
↪           random.rand(1,n_sim)/filtered[i,3] #Basic MC using vals scattered
↪           around rv=0 by average of pmRA and pmdec error
40          for j in range(1,3):
41              bary_vels[i,0,j,:]=((-1)**j)*k*(filtered[i,5]+filtered[i,7])
↪               /(2*filtered[i,3])+(1/2)*k*(filtered[i,6]+filtered[i,8])*np.random
↪               .randn(1,n_sim)/filtered[i,3] #Same thing but centred around +/-
↪               avg pmRA and pmdec
42
43          bary_vels[i,1,:,:]=k*(filtered[i,5]+filtered[i,6]*np.random.
↪           randn(3,n_sim))/(filtered[i,3]+filtered[i,4]*np.random.randn(3,30)
↪           )
44          bary_vels[i,2,:,:]=k*(filtered[i,7]+filtered[i,8]*np.random.
↪           randn(3,n_sim))/(filtered[i,3]+filtered[i,4]*np.random.randn(3,30)
↪           )
45
46          #print(T*A[:,:,i])
47          B=np.matmul(T,A[:,:,i])
48          for l in range(3):
```

```
49              gal_vels[i,:,l,:]=np.matmul(B,bary_vels[i,:,l,:])
50              for m in range(3):
51                  #print(norms[:,m]*np.exp(-0.5*((gal_vels[i,:,l,:]-mu[:,m
    ↪  ].reshape((3,1)))/sig[:,m])**2).shape)
52                  #print(np.exp(-0.5*((gal_vels[i,:,l,:]-mu[:,m].reshape
    ↪  ((3,1))))**2).shape)
53                  P[i,l,m,:]=n_den[m]*np.prod(norms[:,m].reshape((3,1))*np
    ↪  .exp(-0.5*((gal_vels[i,:,l,:]-mu[:,m].reshape((3,1)))/sig[:,m].
    ↪  reshape((3,1)))**2),axis=0)
54              P[i,l,:,:]=P[i,l,:,:]/np.sum(P[i,l,:,:],axis=0)
55          totprobs_rv[i,:,:]=np.average(P[i,:,:,:],axis=2)
56          totprobs[i,:]=np.average(totprobs_rv[i,:,:],axis=0)
57          stdevs[i,:,:]=np.std(P[i,:,:,:],axis=2)
58
59      #Save each of the three (NOT! lol) probability arrays, each with
    ↪  increasingly granular data. Hopefully only the simplest will be
    ↪  necessary but who tf knows
60      #gal_vels=gal_vels.reshape((n_stars,3,3,n_sim))
61      np.savetxt('everything.txt',np.hstack((filtered[:,0].reshape((
    ↪  n_stars,1)),P.reshape((n_stars,3*3*n_sim)))))
62      #resultant 2D array will look like... 3 Us, 3Vs, 3 Ws
63      np.savetxt('avg_probs.txt',np.hstack((filtered[:,0].reshape((n_stars
    ↪  ,1)),totprobs_rv.reshape((n_stars,3*3)),stdevs.reshape((n_stars
    ↪  ,3*3)),np.average(gal_vels[:,:,:,:],axis=3).reshape((n_stars,3*3))
    ↪  , np.std(gal_vels[:,:,:,:],axis=3).reshape((n_stars,3*3)))))
64
65      #Determine which stars are likely thick disk or halo objects (based
    ↪  on total average probabilities), save those to another file
66      thicc_cand=np.ravel(np.argwhere(totprobs[:,1]>=P_thrsh[1]))
67      halo_cand=np.ravel(np.argwhere(totprobs[:,2]>=P_thrsh[2]))
68      np.savetxt('halo_p.txt',totprobs[halo_cand,2])
69      #plt.plot(halo)
70      #print(gal_vels[thicc_cand,:,:,:].shape)
71      if len(thicc_cand) > 0:
```

```
72          np.savetxt('td_poss.txt',np.hstack((filtered[thicc_cand,0].
   ↪ reshape((len(thicc_cand),1)),totprobs_rv[thicc_cand,:,:].reshape((
   ↪ len(thicc_cand),3*3)),stdevs[thicc_cand,:,:].reshape((len(
   ↪ thicc_cand),3*3)), np.average(gal_vels[thicc_cand,:,:,:],axis=3).
   ↪ reshape((len(thicc_cand),3*3)), np.std(gal_vels[thicc_cand,:,:,:],
   ↪ axis=3).reshape((len(thicc_cand),3*3))))))
73       if len(halo_cand) > 0:
74          np.savetxt('halo_poss.txt',np.hstack((filtered[halo_cand,0].
   ↪ reshape((len(halo_cand),1)),totprobs_rv[halo_cand,:,:].reshape((
   ↪ len(halo_cand),3*3)),stdevs[halo_cand,:,:].reshape((len(halo_cand)
   ↪ ,3*3)), np.average(gal_vels[halo_cand,:,:,:],axis=3).reshape((len(
   ↪ halo_cand),3*3)), np.std(gal_vels[halo_cand,:,:,:],axis=3).reshape
   ↪ ((len(halo_cand),3*3))))))
75          np.savetxt('halo_dist.txt',1/((1e-3)*filtered[halo_cand,3]))
76
77       print('Calculations complete, files written. Elapsed time =',(
   ↪ perf_counter()-start_time),' [s]')
78
79       fig0,ax0=plt.subplots(2,2)
80       fig0.tight_layout(pad=2.5)
81       ax0[0,0].hist(P[:,:,0,:].reshape((n_stars*3*n_sim,1)),bins=100)
82       ax0[0,0].set_title('Thin disk')
83       ax0[0,1].hist(P[:,:,1,:].reshape((n_stars*3*n_sim,1)),bins=100)
84       ax0[0,1].set_title('Thick disk')
85       ax0[1,0].hist(P[:,:,2,:].reshape((n_stars*3*n_sim,1)),bins=100)
86       ax0[1,0].set_title('Halo')
87       ax0[1,1].plot(np.arange(0,n_stars*3*n_sim,1),np.sum(P[:,:,:,:],axis
   ↪ =2).reshape((n_stars*3*n_sim,1)))
88       usetex=True
89       ax0[1,1].set_title("test to make sure normalising worked""\n""
   ↪ because i'm paranoid",fontsize=10)
90       plt.show()
91
92 main()
```

A.4 Simulation of White Dwarfs for Testing

Using a list of lines of sight and Galactic population mean velocity and dispersions, I use this code to generate a star sample and simulate their astrometries. Using each line of sight to develop a different rotation matrix, I send these stars into the equatorial velocity frame. This code saves these into text files without radial velocities and calls the code in Appendix A.5 to analyze each group. I have also added a printout of the ratio of radial velocity to proper motion for each Galactic population in each line of sight.

`filter_sim.py`

```python
import numpy as np
import pandas as pd
from astropy import units as u
from astropy.coordinates import SkyCoord
import subprocess as sp
import os


mu=np.array([[-10,0,0],[-20,-40,196],[-7,0,0]])
sig=np.array([[35,50,141],[18,45,75],[25,50,85]])


ngp=SkyCoord(l=0*u.degree,b=90*u.degree,frame='galactic')
gc=SkyCoord(l=0*u.degree,b=0*u.degree,frame='galactic')
gac=SkyCoord(l=180*u.degree,b=0*u.degree,frame='galactic')
sgp=SkyCoord(l=0*u.degree,b=-90*u.degree,frame='galactic')
gce=SkyCoord(l=90*u.degree,b=0*u.degree,frame='galactic')
gcw=SkyCoord(l=-90*u.degree,b=0*u.degree,frame='galactic')
gcne=SkyCoord(l=45*u.degree,b=0*u.degree,frame='galactic')
gcnw=SkyCoord(l=-45*u.degree,b=0*u.degree,frame='galactic')

los=np.vstack(((0,np.radians(90)),(ngp.icrs.ra.radian,ngp.icrs.dec.
    radian),(gc.icrs.ra.radian,gc.icrs.dec.radian),(gac.icrs.ra.radian
    ,gac.icrs.dec.radian),(sgp.icrs.ra.radian,sgp.icrs.dec.radian),(
    gce.icrs.ra.radian,gce.icrs.dec.radian),(gcw.icrs.ra.radian,gcw.
    icrs.dec.radian),(gcne.icrs.ra.radian,gcne.icrs.dec.radian),(gcnw.
    icrs.ra.radian,gcnw.icrs.dec.radian)))
nl=los.shape[0]
```

```
22
23 T=np.array([[-0.06699, -0.87276, -0.48354],[0.49273, -0.45035,
     ↪ 0.74458],[-0.86760, -0.18837, 0.46020]]) #First part of rotation
     ↪ matrix from J & S 1987
24 rotmats=np.zeros((3,3,nl))
25
26 d=90
27 snr=10
28
29 for i in range(nl):
30     A=np.array([[np.cos(los[i,0])*np.sin(los[i,1]), -np.sin(los[i,0]), -
     ↪ np.cos(los[i,0])*np.sin(los[i,1])],[np.sin(los[i,0])*np.cos(los[i
     ↪ ,1]), np.cos(los[i,0]), -np.sin(los[i,0])*np.cos(los[i,1])],[np.
     ↪ sin(los[i,1]), 0, np.cos(los[i,1])]])
31     rotmats[:,:,i]=np.linalg.inv(np.matmul(T,A))
32
33 n_sim=500
34 star_vels_g=np.zeros((n_sim,3,3)) #scatter, line of sight, vel component
     ↪ , pop
35 star_vels_e=np.zeros((nl,n_sim,3,3))
36 star_name=np.vstack((1000*np.ones(n_sim),2000*np.ones(n_sim),3000*np.
     ↪ ones(n_sim))).T
37
38 for j in range(n_sim):
39     star_vels_g[j,:,:]=mu[:,:]+np.random.randn(3,3)*sig[:,:]
40     star_name[j,:]=star_name[j,:]+j
41 fstar_name=np.zeros((nl,n_sim,3))
42 for k in range(nl):
43     star_vels_e[k,:,:,:]=np.matmul(rotmats[:,:,k],star_vels_g[:,:,:])
44     fstar_name[k,:,:]=k*1e4+star_name[:,:]
45
46 pms=star_vels_e[:,:,0:2,:]/(4.74057*d)
47 pm_e=pms/snr
48 vrats=star_vels_e[:,:,2,:]/np.average(star_vels_e[:,:,0:2,:],axis=2)
49 prlx=(1/d)*np.ones((nl,n_sim,3))
50
51 labeldict={
```

```python
52        "los":["NCP","NGP","GC","GaC","SGP","GC+90","GC-90","GC+45","GC-45"]
53  }
54
55  for l in range(nl):
56      fname=labeldict["los"][l]+"_prlx"+str(np.round(1/d,2))+"snr_"+str(
        ↪ snr)+".txt"
57      np.savetxt(fname,np.hstack((fstar_name[l,:,:].reshape(n_sim*3,1),np.
        ↪ degrees(los[l,:])*np.ones((n_sim*3,2)),1/d*np.ones((n_sim*3,1))
        ↪ ,1/(d*snr)*np.ones((n_sim*3,1)),pms[l,:,:,:].reshape((n_sim*3,2)),
        ↪ pm_e[l,:,:,:].reshape((n_sim*3,2)),vrats[l,:,:].reshape(n_sim*3,1)
        ↪ )),header=labeldict["los"][l]+'\nID\tRA\tDec\tprlx\tprlx_e\tPM_ra\
        ↪ tPM_dec\tPM_ra_e\tPM_dec_e\trv_ratio',comments='')
58      this_loc=SkyCoord(ra=los[l,0]*u.radian, dec=los[l,1]*u.radian, frame
        ↪ ='icrs')
59      print(labeldict["los"][l]+'\nequatorial: \n',this_loc.ra.hms,
        ↪ this_loc.dec.degree,'\ngalactic: \n',this_loc.galactic.l.degree,
        ↪ this_loc.galactic.b.degree,'\nthin avg RV/pm: ',np.average(vrats[l
        ↪ ,:,0]),'\nthick avg RV/pm: ',np.average(vrats[l,:,1]),'\nhalo avg
        ↪ RV/pm: ',np.average(vrats[l,:,2]))
60      os.system('python3 transform_UVW.py'+' -f '+fname)
```

A.5 Population Determination of Sample Stars

This code is a modified version of that in Appendix A.3. It no longer looks for photometry in its input files (as my simulated stars have none, they are all assumed to be white dwarfs) and produces different outputs. These outputs are sorted based on patterns in the listing of the input simulated stars and give the correct and incorrect candidacy fraction for each Galactic population. The comment block at the bottom is left included as it produces the plots shown in Figures 2.3 and 3.1.

transform_UVW.py

```python
# Version 29 July 2023

import numpy as np
import astropy as astro
import matplotlib.pyplot as plt
#import emcee as mc
from time import perf_counter# as perf_counter
from datetime import timedelta
import pandas as pd
import argparse
#import timeit
from mpl_toolkits import mplot3d

parser = argparse.ArgumentParser()
parser.add_argument("-f", "--Filename", help = "enter -f then name of WD
    ↪  data file")
popdict={
    "pop":["Thin Disk","Thick Disk","Halo"]
}

def main():
    args=parser.parse_args()
    fname=args.Filename
    start_time=perf_counter()
    #filtered=np.loadtxt('Kept_stars.txt')
    filtered=pd.read_csv(fname, delim_whitespace=True, header=1)
    fbase=fname.split('.')[0]
```

```
27    #LL ID, RA, Dec, prlx, prlx_e, PM_ra, PM_dec, PM_ra_e, PM_dec_e,
28    ra=np.radians(filtered.RA)
29    dec=np.radians(filtered.Dec)
30    #gaia_vels=filtered[:,3:8]*1e-3
31    prlx=filtered.prlx
32    prlx_e=filtered.prlx_e
33    pmra=filtered.PM_ra
34    pmra_e=filtered.PM_ra_e
35    pmdec=filtered.PM_dec
36    pmdec_e=filtered.PM_dec_e
37    n_stars=len(filtered.ID)
38    n_sim=30 #Number of stars to simulate (yeet) through whatever approx
   ↪    to MC this is
39    T=np.array([[-0.06699, -0.87276, -0.48354],[0.49273, -0.45035,
   ↪    0.74458],[-0.86760, -0.18837, 0.46020]]) #First part of rotation
   ↪    matrix from Johnson & Soderblom 1987
40    A=np.zeros([3,3,n_stars])
41    k=4.74057 #some conversion factor #astornomy lul
42    bary_vels=np.zeros([n_stars,3,3,30]) #star #, v component, rv
   ↪    estimate, MC estimates
43    gal_vels=np.zeros([n_stars,3,3,30]) #star, v component, rv est, MC
44
45    #U x3, V x3, W x3
46    mu=np.array([[0,0,0],[-12,-51,196],[0,0,0]])
47    sig=np.array([[39,67,160],[20,38,90],[16,35,90]])
48    n_den=np.array([1,1,1]) #stars per cubic pc
49    norms=[1/sig[0,:],1/sig[1,:],1/sig[2,:]]/(np.sqrt(2*np.pi)) #
   ↪    normalising factors for gaussians
50    P=np.zeros([n_stars,3,3,n_sim]) #star #, rv estimate, population, mc
51    #relprobs=np.zeros([n_stars,3,n_sim]) #star #, pop, mc
52    totprobs_rv=np.zeros([n_stars,3,3]) #star,pop, rv estimate
53    totprobs=np.zeros([n_stars,3]) #star, pop
54    stdevs=np.zeros([n_stars,3,3]) #star, pop, rv
55    P_thrsh=np.array([0,0.5,0.5])
56
57    for i in range(n_stars):
```

```
58          A[:,:,i]=np.array([[np.cos(ra[i])*np.sin(dec[i]), -np.sin(ra[i])
   ↪ , -np.cos(ra[i])*np.sin(dec[i])],[np.sin(ra[i])*np.cos(dec[i]), np
   ↪ .cos(ra[i]), -np.sin(ra[i])*np.cos(dec[i])],[np.sin(dec[i]), 0, np
   ↪ .cos(dec[i])]])
59          bary_vels[i,0,0,:]=(1/2)*k*(pmra_e[i]+pmdec_e[i])*np.random.rand
   ↪ (1,n_sim)/prlx[i] #Basic MC using vals scattered around rv=0 by
   ↪ average of pmRA and pmdec error
60          for j in range(1,3):
61              bary_vels[i,0,j,:]=((-1)**j)*k*(pmra[i]+pmdec[i])/(2*prlx[i
   ↪ ])+(1/2)*k*(pmra_e[i]+pmdec_e[i])*np.random.randn(1,n_sim)/prlx[i]
   ↪  #Same thing but centred around +/- avg pmRA and pmdec

62
63          bary_vels[i,1,:,:]=k*(pmra[i]+pmra_e[i]*np.random.randn(3,n_sim)
   ↪ )/(prlx[i]+prlx_e[i]*np.random.randn(3,n_sim))
64          bary_vels[i,2,:,:]=k*(pmdec[i]+pmdec_e[i]*np.random.randn(3,
   ↪ n_sim))/(prlx[i]+prlx_e[i]*np.random.randn(3,n_sim))

65
66          #print(T*A[:,:,i])
67          B=np.matmul(T,A[:,:,i])
68          for l in range(3):
69              gal_vels[i,:,l,:]=np.matmul(B,bary_vels[i,:,l,:])
70              for m in range(3):
71                  #print(norms[:,m]*np.exp(-0.5*((gal_vels[i,:,l,:]-mu[:,m
   ↪ ].reshape((3,1)))/sig[:,m])**2).shape)
72                  #print(np.exp(-0.5*((gal_vels[i,:,l,:]-mu[:,m].reshape
   ↪ ((3,1))))**2).shape)
73                  P[i,l,m,:]=n_den[m]*np.prod(norms[:,m].reshape((3,1))*np
   ↪ .exp(-0.5*((gal_vels[i,:,l,:]-mu[:,m].reshape((3,1)))/sig[:,m].
   ↪ reshape((3,1)))**2),axis=0)
74              P[i,l,:,:]=P[i,l,:,:]/np.sum(P[i,l,:,:],axis=0)
75          totprobs_rv[i,:,:]=np.average(P[i,:,:,:],axis=2)
76          totprobs[i,:]=np.average(totprobs_rv[i,:,:],axis=0)
77          stdevs[i,:,:]=np.std(P[i,:,:,:],axis=2)
78 #          if np.amax(np.argmax(P[i,:,:,:],axis=1))-np.amin(np.argmax(P[i
   ↪ ,:,:,:],axis=1)) == 2:
79 #              print(i)
```

```
80      #Save each of the three (NOT! lol) probability arrays, each with
    ↪    increasingly granular data. Hopefully only the simplest will be
    ↪    necessary but who tf knows
81      #gal_vels=gal_vels.reshape((n_stars,3,3,n_sim))
82      ##np.savetxt('everything.txt',np.hstack((filtered[:,0].reshape((
    ↪    n_stars,1)),P.reshape((n_stars,3*3*n_sim)))))
83      #resultant 2D array will look like... 3 Us, 3Vs, 3 Ws
84      ##np.savetxt('avg_probs.txt',np.hstack((filtered[:,0].reshape((
    ↪    n_stars,1)),totprobs_rv.reshape((n_stars,3*3)),stdevs.reshape((
    ↪    n_stars,3*3)),np.average(gal_vels[:,:,:,:],axis=3).reshape((
    ↪    n_stars,3*3)), np.std(gal_vels[:,:,:,:],axis=3).reshape((n_stars
    ↪    ,3*3)))))
85

86      #Determine which stars are likely thick disk or halo objects (based
    ↪    on total average probabilities), save those to another file
87      thicc_cand=np.ravel(np.argwhere(totprobs[:,1]>=P_thrsh[1]))
88      halo_cand=np.ravel(np.argwhere(totprobs[:,2]>=P_thrsh[2]))
89      ##np.savetxt('halo_p.txt',totprobs[halo_cand,2])
90      #plt.plot(halo)
91      #print(gal_vels[thicc_cand,:,:,:].shape)
92      """
93      if len(thicc_cand) > 0:
94          np.savetxt('td_poss.txt',np.hstack((filtered[thicc_cand,0].
    ↪    reshape((len(thicc_cand),1)),totprobs_rv[thicc_cand,:,:].reshape((
    ↪    len(thicc_cand),3*3)),stdevs[thicc_cand,:,:].reshape((len(
    ↪    thicc_cand),3*3)), np.average(gal_vels[thicc_cand,:,:,:],axis=3).
    ↪    reshape((len(thicc_cand),3*3)), np.std(gal_vels[thicc_cand,:,:,:],
    ↪    axis=3).reshape((len(thicc_cand),3*3)))))
95      if len(halo_cand) > 0:
96          np.savetxt('halo_poss.txt',np.hstack((filtered[halo_cand,0].
    ↪    reshape((len(halo_cand),1)),totprobs_rv[halo_cand,:,:].reshape((
    ↪    len(halo_cand),3*3)),stdevs[halo_cand,:,:].reshape((len(halo_cand)
    ↪    ,3*3)), np.average(gal_vels[halo_cand,:,:,:],axis=3).reshape((len(
    ↪    halo_cand),3*3)), np.std(gal_vels[halo_cand,:,:,:],axis=3).reshape
    ↪    ((len(halo_cand),3*3)))))
97          np.savetxt('halo_dist.txt',1/((1e-3)*filtered[halo_cand,3]))
98      """
```

```
99      np.savetxt(fbase+'_results.txt',np.hstack((filtered.ID.to_numpy().
    ↪ reshape(n_stars,1),totprobs)))
100     thinlist=np.arange(0,n_stars,3,dtype=int).reshape((int(n_stars/3),1)
    ↪ )
101     thicklist=np.arange(1,n_stars,3,dtype=int).reshape((int(n_stars/3)
    ↪ ,1))
102     halolist=np.arange(2,n_stars,3,dtype=int).reshape((int(n_stars/3),1)
    ↪ )
103     totlist=np.hstack((thinlist,thicklist,halolist))
104     perflist=np.zeros((3,3))
105     for i in range(int(n_stars/3)):
106         for j in range(3):
107             if np.argmax(totprobs[totlist[i,j],:]) == 1:
108                 perflist[j,1]+=1
109             elif np.argmax(totprobs[totlist[i,j],:]) == 2:
110                 perflist[j,2]+=1
111             else:
112                 perflist[j,0]+=1
113     perflist=perflist*3/n_stars
114     np.savetxt(fbase+'_perf.txt',np.hstack((np.arange(1,4,1,dtype=int).
    ↪ reshape((3,1)),perflist)),header=fbase+'\n\t1_ThinDisk\
    ↪ t2_ThickDisk\t3_Halo',comments='',fmt="%-12s")
115     print('Calculations complete, files written. Elapsed time =',(
    ↪ perf_counter()-start_time),' [s]')
116 """
117     sample_pops=np.argmax(P[8,:,:,:],axis=1).reshape((3*n_sim,1))
118     print(sample_pops)
119     axis_labels=['U','V','W']
120     labels=['RV=0','RV=+avg(PM)','RV=-avg(PM)']
121     fig0=plt.figure()
122     fig0.tight_layout(pad=10)
123     ax0=fig0.add_subplot(2,1,1,projection='3d')
124     ax1=fig0.add_subplot(2,1,2,projection='3d')
125     #for k in range(3):
126     #    ax0.scatter3D(bary_vels[0,0,k,:],bary_vels[0,1,k,:],bary_vels
    ↪ [0,2,k,:],marker='x',label=labels[k])
```

```
127        p=ax1.scatter3D(gal_vels[8,0,:,:].reshape((3*n_sim,1)),gal_vels
    ↪    [8,1,:,:].reshape((3*n_sim,1)),gal_vels[8,2,:,:].reshape((3*n_sim
    ↪    ,1)),marker='x',c=sample_pops)#label=labels[k])
128        ax0.set_title('Equatorial')
129        ax1.set_title('Galactic')
130        ax0.legend(loc='center left', bbox_to_anchor=(1.35, 0.5))
131 #     ax1.legend(loc='center left', bbox_to_anchor=(1.35, 0.5))
132        ax0.set_xlabel(r'$v_r$')
133        ax0.set_ylabel(r'$v_\alpha$')
134        ax0.set_zlabel(r'$v_\delta$')
135        ax1.set_xlabel(axis_labels[0])
136        ax1.set_ylabel(axis_labels[1])
137        ax1.set_zlabel(axis_labels[2])
138        cbar=fig0.colorbar(p,ax=ax1,ticks=[0,1,2])
139        cbar.ax.set_yticklabels(['Thin Disk','Thick Disk','Halo'])
140        plt.show()
141 """
142
143 main()
```

A.6 The V/V_{max} Test

Using a series of tabulated white dwarf evolutionary tracks using the models in BASE-9 (compiled and provided by Alisa Tiselska), I use this code to perform an interpolation on age and mass of the white dwarf, generating a function which outputs Gaia G band magnitude. As described in Section 5.2.2, this produces a maximum detectable range and therefore an enclosed detectable volume for each star. Thus, I calculate V/V_{max} for each individual input star and find the average of this value for the input population. This code also outputs the effective limiting magnitude which would give $\langle V/V_{max}\rangle = 1/2$ and the two sides of the fit criterion Equation 5.7 as seen in the same Section.

vmax.py

```python
1  import numpy as np
2  import matplotlib.pyplot as plt
3  import pandas as pd
4  from scipy.interpolate import bisplrep as rep
5  from scipy.interpolate import bisplev as ev
6  import argparse
7
8  parser=argparse.ArgumentParser()
9  parser.add_argument("-f","--filenames",nargs='+',help="Filename or
       ↪ metanames for BASE9 files")
10 """
11 def get_stars():
12     fname='Kept_stars.txt'
13     data=np.loadtxt(fname)
14     return(data)
15
16 def get_halo():
17     fname='halo_obj.txt'
18     data=np.loadtxt(fname)
19     return(data)
20 def get_td():
21     fname='td_obj.txt'
22     data=np.loadtxt(fname)
```

```
23        return(data)
24 """
25
26 args=parser.parse_args()
27 fnames=args.filenames
28 nstars=len(fnames)
29 stardata=np.zeros((nstars,5)) #Gaia ID, age, mass (both avgd), prlx,
   ↪ dmax
30 gfc=np.loadtxt('Keep_stars.txt')
31 g_plx=(1e-3)*gfc[:,3]
32 g_id=gfc[:,12]
33 #g_G=gfc[:,9]
34 ndel=0
35 for i in range(nstars):
36     b9data=pd.read_csv(fnames[i],delim_whitespace=True,header=0)
37     #print(b9data.stage)
38     #print(np.argwhere(b9data.stage.to_numpy() >= 3))
39     postbi=np.ravel(np.argwhere(b9data.stage.to_numpy() >= 3))
40     starname=float(fnames[i].split('/')[-1].split('.')[0])
41     hage=np.histogram(10**(b9data.logAge[postbi]-9),bins=100)
42     hmass=np.histogram(b9data.mass[postbi])
43     nage=np.argmax(hage[0])
44     nmass=np.argmax(hmass[0])
45     avage=np.average(hage[1][nage:nage+2])
46     avmass=np.average(hmass[1][nmass:nmass+2])
47     gfc_loc=np.ravel(np.argwhere(g_id==starname))
48     if len(gfc_loc) == 0:
49         stardata=np.delete(stardata,i,axis=0)
50         nstars-=1
51         ndel+=1
52         print(stardata.shape)
53     else:
54         plx=g_plx[gfc_loc].astype('float')
55         stardata[i-ndel,0:4]=np.array((starname, avage, avmass, plx[-1])
   ↪ )
56 wd=pd.read_csv('wd_evol.grid',delim_whitespace=True,header=0)
57 #mass=DA.loc[:,['M/Mo']].to_numpy()
```

```python
58  mass=(wd.Mass.to_numpy()).reshape((10,13))
59  mag=(wd.G.to_numpy()).reshape((10,13))
60  age=(np.linspace(3,12,13)*np.ones((10,1)))#.reshape((len(wd.Mass),1))
61  plt.figure()
62  for j in range(10):
63      plt.scatter(np.linspace(12,3,13),mass[j,:],c=mag[j,:])
64  plt.set_cmap('viridis_r')
65  usetex=True
66  plt.xlabel('Age [Gyr]')
67  plt.ylabel(r'Mass [$M_\odot$]')
68  cbar=plt.colorbar(label=r'$G$ Magnitude')
69  cbar.ax.invert_yaxis()
70  plt.show()
71  #plt.figure(1)
72  #plt.plot(mag)
73  #plt.show()
74  mag_tck=rep(age.reshape(130),mass.reshape(130),mag.reshape(130))
75  abs_mag=np.zeros((nstars,1))
76  for k in range(nstars):
77      abs_mag[k]=ev(stardata[k,1],stardata[k,2],mag_tck)
78
79  lim_mag_G=20.7
80
81  stardata[:,-1]=np.ravel(10**(1+(lim_mag_G-abs_mag)/5))
82  vvmax=(1/(stardata[:,4]*stardata[:,3]))**3
83
84  np.savetxt('vvmax.txt',np.hstack((stardata,abs_mag.reshape((nstars,1)),
        ↪ vvmax.reshape((nstars,1)))),comments='',header='ID\tavg_Age\
        ↪ tavg_mass\tprlx\td_max\tG\tv/v_max')
85  vvmaxtest=np.average(vvmax)
86  alpha=vvmaxtest/0.5
87  G_lim_p=lim_mag_G+(5/3)*np.log10(alpha)
88  print(vvmaxtest,G_lim_p, np.abs(vvmaxtest-0.5),1/np.sqrt(12*nstars))
89
90  """
91  halo=get_halo()
92  td=get_td()
```

```
93  gaia_td=np.zeros(len(td))
94  gaia_h=np.zeros(len(halo))
95  """
```