



University of Kentucky  
UKnowledge

---

Theses and Dissertations--Physics and  
Astronomy

Physics and Astronomy

---


2022

## MILKY WAY MORPHOLOGY PROBED BY 6D ASTROMETRIC DATA FROM THE GAIA SPACE TELESCOPE

Joshua Taylor Harry

University of Kentucky, jtharry99@gmail.com

Author ORCID Identifier:

 <https://orcid.org/0000-0003-0250-7262>

Digital Object Identifier: <https://doi.org/10.13023/etd.2022.316>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

### Recommended Citation

Harry, Joshua Taylor, "MILKY WAY MORPHOLOGY PROBED BY 6D ASTROMETRIC DATA FROM THE GAIA SPACE TELESCOPE" (2022). *Theses and Dissertations--Physics and Astronomy*. 99.  
[https://uknowledge.uky.edu/physastron\\_etds/99](https://uknowledge.uky.edu/physastron_etds/99)

This Master's Thesis is brought to you for free and open access by the Physics and Astronomy at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Physics and Astronomy by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Joshua Taylor Harry, Student

Dr. Susan Gardner, Major Professor

Dr. Christopher Crawford, Director of Graduate Studies

MILKY WAY MORPHOLOGY PROBED BY 6D ASTROMETRIC DATA FROM  
THE GAIA SPACE TELESCOPE

---

THESIS

---

A thesis submitted in partial  
fulfillment of the requirements for  
the degree of Master of Science in  
the College of Arts and Sciences at  
the University of Kentucky

By  
Joshua T. Harry  
Lexington, Kentucky

Director: Dr. Susan Gardner, Professor of Physics  
Lexington, Kentucky  
2022

Copyright© Joshua T. Harry 2022  
<https://orcid.org/0000-0003-0250-7262>

## ABSTRACT OF THESIS

### MILKY WAY MORPHOLOGY PROBED BY 6D ASTROMETRIC DATA FROM THE GAIA SPACE TELESCOPE

At varying height above and below the plane of the Milky Way, I have used astrometric methods to classify stars of different galactic components of the Milky Way – the thin disk, thick disk, and stellar halo. This work complements prior study of Milky Way sub-structure – notably involving number density and/or pairwise correlations – which demonstrate non-steady state effects in the galaxy, such as axial/north-south symmetry breaking [1, 2] or more complex phenomena like the Gaia snail [3]. This has motivated my exploration of stellar population changes with height about the Milky Way mid-plane, and the study of symmetry in such changes above and below the galactic mid-plane.

Drawing from the second data release (DR2) of the Gaia space telescope, I have chosen a stellar sample with notably small parallax error and a 6D astrometric phase space, totaling 707,772 stars. Prior work in Hinkel 2020 [4] was used as a reference for optimal sampling parameters. Selection cuts have been taken to avoid saturation from the Magellanic Clouds and the mid-plane of the Milky Way, in addition to controlling for *Gaia*'s astrometric uncertainties and sampling biases. After a conversion of astrometric to galactocentric coordinates, I have employed a statistical method to distinguish the populations of stars by their likelihood of belonging to the Milky Way's thin/thick disk and stellar halo components, using stellar kinematic data. Kinematic criteria for galactic populations were chosen to avoid galactic component contamination, where stars are attributed to the incorrect galactic components due to phase space overlap of the different components – a common issue in population separation routines. To assess confidence in this statistical method, I used the bootstrapping method to construct error estimates within sub-slices in galactocentric  $z$ . My methods have yielded a notable, novel result: at low galactic  $z$ , the population fraction variation is asymmetric North and South of the galactic mid-plane.

KEYWORDS: Gaia, Astrometry, Kinematic Probability, Population Separation

---

Joshua T. Harry

---

August 9, 2022

MILKY WAY MORPHOLOGY PROBED BY 6D ASTROMETRIC DATA FROM  
THE GAIA SPACE TELESCOPE

By  
Joshua T. Harry

Dr. Susan Gardner

Director of Thesis

Dr. Christopher Crawford

Director of Graduate Studies

August 9, 2022

Date

## ACKNOWLEDGMENTS

To my advisor Dr. Susan Gardner – thank you for being an exceptionally kind mentor and offering support and guidance time and time again, both in academics and research. Among my research collaborators, I want to thank Dr. Brian Yanny and Dr. Austin Hinkel. Brian has been very receptive to my confusions and questions, and he has always offered insight with great tact. Austin has been by my side for much of my research endeavors in this group, and I am immensely grateful for every bit of consultation and encouragement he has offered – they were well received.

Moreover, I deeply thank my friends and family for their encouragement in my efforts.

## TABLE OF CONTENTS

Acknowledgments . . . . .	iii
List of Tables . . . . .	v
List of Figures . . . . .	vi
Chapter 1 Introduction . . . . .	1
Chapter 2 Data Selection and Processing . . . . .	3
2.1 Gaia Queries with ADQL . . . . .	3
2.2 Stellar Sample . . . . .	3
2.2.1 Selection Cuts and Error Considerations . . . . .	4
2.2.2 Color–Magnitude Distribution over <i>G</i> –Band & Giant Sub–sample . . . . .	6
2.3 Stellar Density Variation . . . . .	9
Chapter 3 Kinematic Analysis . . . . .	13
3.1 Toomre Cutoff Method and Phase Space Behavior . . . . .	13
3.2 Kinematic Probability Method . . . . .	13
3.2.1 Theory and Kinematic Criterion . . . . .	16
3.2.2 Memberships in Phase Space . . . . .	17
3.3 Population Variation & Error Estimation via Bootstrapping . . . . .	20
3.4 Results . . . . .	20
Chapter 4 Study Outcomes and Conclusions . . . . .	23
4.1 Relation to Other Studies . . . . .	23
4.2 Thin Disk Population Asymmetry for Giant and Main Sequence Sub– Samples . . . . .	25
4.3 Future Work . . . . .	26
Appendices . . . . .	27
Appendix A: Galactic Coordinate Conversions . . . . .	27
Galactic Proper Motions . . . . .	27
Galactocentric Velocities . . . . .	27
LSR Frame . . . . .	28
Bibliography . . . . .	30
Vita . . . . .	35

## LIST OF TABLES

2.1	Giant/Main Sequence Proportions for 1 magnitude slices in $G$ -band. . .	8
3.1	Dispersion, $U, V$ asymmetry, and overall population fraction values for the Gaussian velocity signatures of the thin disk, thick disk, and halo. Note that the population fractions do not sum to 1, because Bensby et al. 2014 accounts for the proportion of stars in the Hercules stream, which make up the remaining 6% of their sample [5]. . . . .	16



## LIST OF FIGURES

2.1	<i>Gaia</i> Scanning Law – In the (solarcentric) ICRS coordinate system [6], the number of field transits binned to a given pixel of the sky varies nonuniformly along right ascension ( $x$ -axis) and declination ( $y$ -axis). This references the first 5 years of <i>Gaia</i> data. See also the scanning law movie by Berry Holl [7] . . . . .	4
2.2	Color–magnitude diagram (CMD) for the full $10 < G < 18$ sample. A cutoff has been set so that above both light blue lines, stars are considered giants, and otherwise, are considered main sequence. This yields 203,629 (29.3%) giants and 491,716 (70.7%) main sequence stars. Compared to Yanny & Gardner 2013 which observes only 0.04% giant stars [8], giants are much more appreciable in our sample because we have included more bright–end–magnitude stars ( $10 < G < 18$ ). . . . .	8
2.3	Distribution of stellar number density as a function of galactic $z$ , looking just at the inner $ z  < 1$ kpc for visual clarity. However, the fits are calculated for the full $ z  < 3$ kpc range. The blue points (with error bars) are the stellar densities, the black curve is the single $\text{sech}^2$ fit, while the red curve is the two–component $\text{sech}^2$ fit. Despite appearances, the fits are very poor, taking reduced $\chi^2$ values of $\chi^2_\nu = 1113.43$ (single $\text{sech}^2$ ) and $\chi^2_\nu = 438.95$ (two–component), for 599 degrees of freedom. This is primarily due to the Poisson errors of the points being much smaller than the residuals from the fit lines. It seems that the artificial blending of the thin and thick disk components in the double $\text{sech}^2$ fit form does not fully capture the morphology of our stellar sample. As observed by Widrow et al. 2012 and Yanny & Gardner 2013 [2, 8], the residuals about the fits have wave–like structure as the points oscillate above and below the fit curves (see also Fig. 2.4). . . . .	10
2.4	Plot of normalized residuals ( $\Delta \equiv (\text{data} - \text{fit}) / \text{fit}$ ) as a function of galactic $z$ . For the sake of visualization, the domain in $z$ is limited to $ z  < 2$ kpc and the vertical range is narrowed. The blue points are the residuals for the single $\text{sech}^2$ fit and the orange points are the residuals for the double $\text{sech}^2$ fit. In both cases, the residuals have odd parity at low $ z $ , consistent with the literature [2]. Beyond around $ z  \sim 0.6$ kpc, the single $\text{sech}^2$ fit becomes ill–behaved as the data increasingly deviates from the fit. Not shown here, the single $\text{sech}^2$ fit residuals reach a peak at $ z  \sim 1.5$ kpc then decrease and return to the zero line at about $ z  \sim 2.3$ kpc. The double $\text{sech}^2$ fit residuals appear to agree with the residuals found in Widrow et al. 2012 [2], confirming the wavy north–south asymmetry in the stellar density. . . . .	12

- 3.1 Toomre energy diagram for the stellar sample, plotting  $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$ . The red line traces the simple disk/halo cutoff, where  $\sqrt{U_{LSR}^2 + V_{LSR}^2 + W_{LSR}^2} = 180$  km/s. We have also projected the data by binning along the axes. Each histogram contains  $\lfloor \frac{695345}{200} \rfloor = 3476$  bins. In this case, the cutoff method yields 688609 (99.03%) disk stars and 6736 (0.97%) halo stars. Our Toomre diagram and histogram distributions are consistent with the results of Buder 2019 [9], yet offer better statistics and thus higher resolution. 14
- 3.2 Toomre diagram ( $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$ ), color coded according to different regimes of  $TD/D$ . Light orange is  $TD/D < 0.1$ , orange is  $0.1 < TD/D < 0.5$ , pink is  $0.5 < TD/D < 2$ , purple is  $2 < TD/D < 10$ , and dark purple is  $TD/D > 10$ . The red line is once again the 180 km/s cutoff. Generally, stars closer to the origin of the phase space tend to be dominantly (thin) disk stars. . . . . 18
- 3.3 Toomre diagram ( $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$ ), color coded according to different regimes of  $TD/H$ . Light orange is  $TD/H > 10$ , orange is  $2 < TD/H < 10$ , pink is  $0.5 < TD/H < 2$ , purple is  $0.1 < TD/H < 0.5$ , and dark purple is  $TD/H < 0.1$ . The red line is the 180 km/s cutoff. . . 19
- 3.4 Proportions of membership within the different galactic components, as a function of  $z$ , for stars in the Northern hemisphere of the galaxy ( $z > 0$ ). Red is thin disk, orange is thick disk, purple is thin/thick disk transition, green is thick disk/halo transition, and blue is halo. . . . . 21
- 3.5 Proportions of membership within the different galactic components vs  $z$ , for stars in the Southern hemisphere of the galaxy ( $z < 0$ ). Red is thin disk, orange is thick disk, purple is thin/thick disk transition, green is thick disk/halo transition, and blue is halo. . . . . 21
- 4.1 A plot of the population fraction asymmetry parameter  $\mathcal{A}$  versus galactic  $z$ , for the thin disk. There is surprisingly great agreement between the thin disk asymmetry and the number count asymmetry distributions. Comparing with Fig. 5 in Bennett & Bovy 2018 [10], which uses *Gaia* DR2 data to explore number count asymmetry, the locations in  $z$  of the peaks and dips line up extremely well. For example, in the number count asymmetry, there is a large dip just before 0.5 kpc, a small peak around 0.7 kpc and a small dip around 0.8 kpc, rising up to a small peak near 1.1 kpc, etc. [10]. Yanny & Gardner 2013 give focus to the asymmetry features at 0.8 kpc and 1.5 kpc, using data from SDSS–DR9 [8]. These oscillatory features are also seen in the population fraction asymmetry. . 24
- 4.2 A plot of the population fraction asymmetry parameter  $\mathcal{A}$  versus galactic  $z$ , for the thin disk, using the sub-sample of giant stars. The asymmetry distribution of the giant sample is very similar to that of the combined sample in Fig. 4.1, suggesting similar expression of non-steady state effects. 25

- 4.3 A plot of the population fraction asymmetry parameter  $\mathcal{A}$  versus galactic  $z$ , for the thin disk, using the sub-sample of main sequence stars. The domain and range are restricted to that in Fig. 4.2 for the sake of comparison. Both the main sequence and giant samples agree on the dip near 0.5 kpc and the peak near 0.7 kpc, but the main sequence asymmetry distribution shows unique features beyond  $z \sim 0.8$  kpc. Moreover, due to deficiency in main sequence stars, there are no population fractions and thus no asymmetry values beyond  $z \sim 1.5$  kpc. . . . . 26

## Chapter 1 Introduction

Great data influxes from astronomical surveys have enabled study of astrophysical phenomena in rich detail – notably in observations of the Milky Way. Starting in the early 1990s, the European Space Agency’s *Hipparcos* satellite was the first space-based mission to record astrometric data of stars – their position, distance, motion, colors, and brightness. To distance accuracies of 20%, *Hipparcos* measured  $\sim 2.5$  million objects within 200 pc of Earth [11]. Moreover, several other ground-based surveys, such as the Sloan Digital Sky Survey, Pan-STARRS, etc. have catalogued billions of stars over the past two decades [12,13]. The *Gaia* space telescope, launched in 2013, has yielded high-precision astrometric data for over a billion objects, far outclassing *Hipparcos* [14,15]. A major factor that sets *Gaia* apart from *Hipparcos* is the measurement of radial velocities (starting from *Gaia*’s second data release), allowing *Gaia* a full 6D phase space view of its stars, making a full kinematic analysis possible. In April 2018, the *Gaia* space telescope released preliminary data from its second data collection, containing a total  $\sim 1.7$  billion sources. *Gaia* DR2 yields ample statistics that enable rigorous study of the astrometry of Milky Way stars.

There have been noteworthy developments in the recent study of fine structure in the Milky Way. Most promising have been studies of Milky Way configuration space – using probes of stellar number density (see Widrow 2012 et al., Yanny & Gardner 2013, Ferguson et al. 2017 [2, 8, 16]), pairwise stellar correlations (Hinkel et al. 2021 [17]), stellar kinematics (Williams et al. 2013 [18]), etc. Such work has, for instance, revealed asymmetries in the stellar number counts across the Galactic midplane, as well as nontrivial wave structure in the Milky Way disk [2]. My current focus is to complement the established configuration space point of view with an analysis of the Milky Way’s stellar phase space. Of primary interest are stars of nontrivial velocities in the stellar halo, which may reflect dynamic effects of dark

matter in our galaxy. This thesis is a proof-of-principle demonstration of separating the Milky Way components using a kinematic analysis – for a stellar sample with carefully chosen astrometric parameters – and then studying the component membership variation with height about the Galactic mid-plane.

Chapter 2 details the motivation for and content of my selected stellar sample, along with methodology within the *Gaia* database for drawing astrometric data satisfying the criteria of interest. Moreover, several selections – such as galactic  $R$  and  $\phi$  cuts – are implemented in Python for the drawn sample. Chapter 3 studies the post-processed sample from a kinematic viewpoint. Using only kinematics, two main techniques for separating Milky Way component populations are investigated: (1) a velocity cutoff method suggested by Toomre (cited in Sandage, Fouts et al. 1987 [19] as a private communication with Toomre), and (2) a probability-based method, first implemented by Bensby [20], that classifies stars more flexibly and reliably than the cutoff method, the details and significance of which I will delineate in chapter 3. For the second “kinematic probability” method, whereby classification is based on inequality criteria, I develop an error estimate through bootstrapping – where synthetic data sets are generated from the original set by re-sampling, allowing for repetition [21]. Lastly, chapter 4 is a holistic interpretation of my cumulative results and their place within the greater structure of the Milky Way.

## Chapter 2 Data Selection and Processing

We make use of the online *Gaia* database to access astrometric data from the European Space Agency’s *Gaia* satellite telescope [22]. To conduct an analysis with minimal statistical errors in astrometric quantities of interest, with mitigation of systematic biases in sampling and completeness, the particular selections/criteria must be implemented in the *Gaia* database to yield our stellar sample. The exact criteria imposed are detailed later in the chapter.

### 2.1 Gaia Queries with ADQL

ADQL (Astronomical Data Query Language) is a modification of the SQL language intended for astronomical datasets. Within the context of the *Gaia* database, the ADQL syntax returns indexed columns of the variables you have chosen, in addition to allowing conditional statements on those variables to perform simple selections. Drawing data – or “querying” – from *Gaia* can be done online with the *Gaia* archive [22, 23].

### 2.2 Stellar Sample

Prior work has been done in deducing selection criteria for a *Gaia* DR2 query, so as to simultaneously minimize parallax errors and sampling biases, such as *Gaia*’s scanning pattern of the sky (Hinkel 2021 [17]). The current work imposes some differences, however. For our kinematic study, stars of nontrivial radial velocity are selected. Hinkel 2021 works in the *G*–band magnitude range  $14 < G < 18$  [17], within which the *Gaia* DR2 catalogue is nearly complete (see the *Gaia* collaboration’s DR2 release paper [15]). But because fainter stars consistently lack radial velocity data, astrometric data is sparse in *Gaia* for magnitudes above  $G \sim 13$  (Fig. 1 in *Gaia*’s DR2 paper shows the distribution of objects with respect to *G*–band [15]). Thus, the

$G$ -band selection is expanded to  $10 < G < 18$  to allow for brighter stars and better statistics. There is a slight loss in completeness by including  $10 < G < 12$ , since *Gaia* DR2 has an incomplete source list for bright stars, but this bias primarily becomes significant around  $G < 7$  [15], so that we can neglect it in our case. This yields an “essentially” complete and sizable sample.

### 2.2.1 Selection Cuts and Error Considerations

*Gaia* scans across the sky nonuniformly due to the orbit and rotational behavior of the satellite itself, and certain regions of the sky are sampled more than others. Because relative parallax error  $\sigma_{\varpi}/\varpi$  has dependence on the number of observations for a given star, the relative parallax error is biased with the position in the sky.

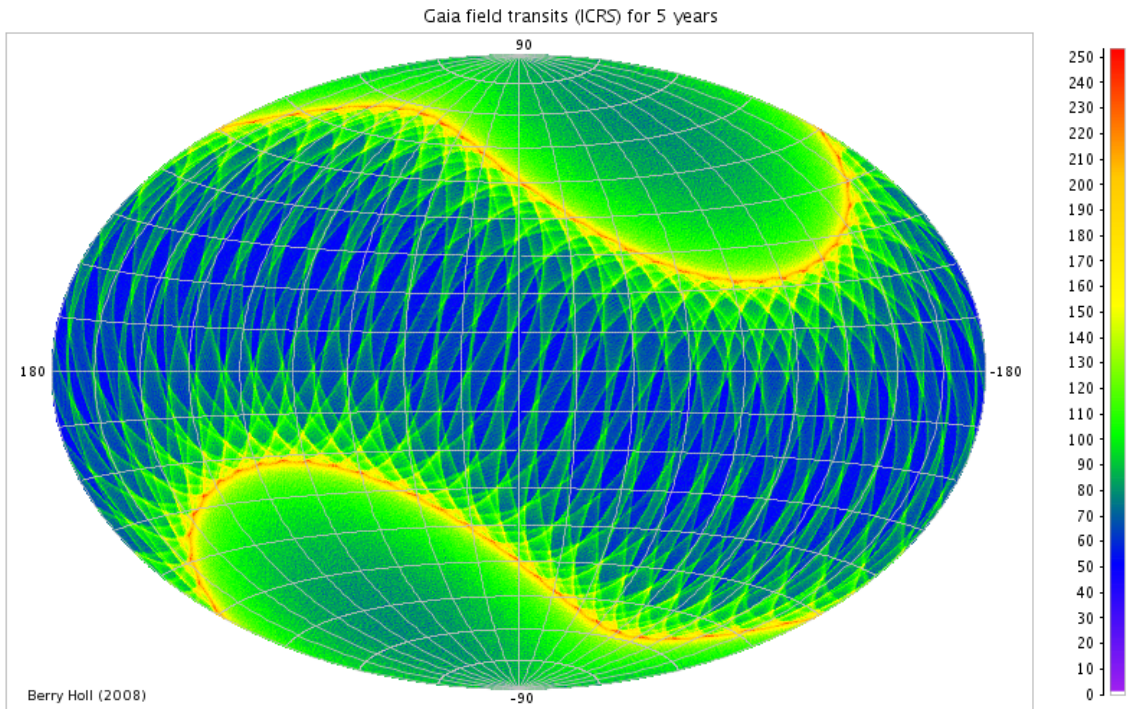


Figure 2.1: *Gaia* Scanning Law – In the (solarcentric) ICRS coordinate system [6], the number of field transits binned to a given pixel of the sky varies nonuniformly along right ascension ( $x$ -axis) and declination ( $y$ -axis). This references the first 5 years of *Gaia* data. See also the scanning law movie by Berry Holl [7]

As determined by Hinkel 2021 [17], the following selections were made to achieve relatively low parallax error with minimal sampling bias: (1) avoiding the Galactic mid-plane in  $b$  by choosing  $|b| > 30^\circ$  (2) G-band magnitude,  $10 < G < 18$  mag (3) BP-RP color,  $0.5 < G_{BP} - G_{RP} < 2.5$  mag (4) positive parallax,  $\varpi > 0$  (5) galactic  $R$ ,  $7 < R < 9$  kpc (6) Galactic  $\phi$ ,  $174 < \phi < 186^\circ$  (7) galactic  $z$ ,  $0.2 < |z| < 3.0$  kpc (8) in-plane distance,  $d \cos b < 1.2$  kpc (9) LMC/SMC excision (also referred to as “box cuts”) – remove data with  $(30 < |b| < 39^\circ) \wedge ((271 < \ell < 287^\circ) \vee (73 < \ell < 89^\circ))$  (LMC), and  $(41 < |b| < 48^\circ) \wedge ((299 < \ell < 307^\circ) \vee (53 < \ell < 61^\circ))$  (SMC)

Note: To prevent sample bias across the Galactic anticenter and across the Galactic mid-plane, due to the LMC/SMC cuts, the data windows to excise have reflection symmetry across the anticenter and mid-plane.

More complicated selections such as the LMC/SMC cuts are difficult to implement with ADQL syntax and increase query time significantly, so that a few selections are done post-query in Python.

Below is the *Gaia* query used for this sample:

```
SELECT 1/gaia_source.parallax AS d, gaia_source.l, gaia_source.b,
gaia_source.ra, gaia_source.dec, gaia_source.ra_error,
gaia_source.dec_error, gaia_source.parallax_error,
gaia_source.pmra_error, gaia_source.pmdec_error,
gaia_source.radial_velocity_error, gaia_source.phot_g_mean_mag as G,
gaia_source.radial_velocity, gaia_source.pmra, gaia_source.pmdec,
gaia_source.teff_val, gaia_source.lum_val, gaia_source.bp_rp,
gaia_source.phot_g_mean_mag+5*log10(gaia_source.parallax)-10 as mg
FROM gaiadr2.gaia_source WHERE (gaia_source.radial_velocity IS NOT
NULL AND ABS(gaia_source.b) > 30 AND gaia_source.phot_g_mean_mag > 10
AND gaia_source.phot_g_mean_mag < 18 AND gaia_source.bp_rp < 2.5 AND
gaia_source.bp_rp > 0.5 AND ABS((1/gaia_source.parallax)*
SIN(RADIANS(gaia_source.b))) < 3.0 AND ABS((1/gaia_source.parallax)*
SIN(RADIANS(gaia_source.b))) > 0.2 AND (1/gaia_source.parallax)*
COS(RADIANS(gaia_source.b)) < 1.2 AND gaia_source.parallax > 0 AND
gaia_source.astrometric_params_solved = 31)
```



This yielded a total of 754,138 stars.

Next, the cuts on  $R$  and  $\phi$  have been implemented, as well as the box cuts, in a Python script. After these selections, our final sample numbers 695,345 stars.

*Gaia* also reports errors for each of its astrometric quantities. Averaged over our entire sample, we find these quantities to have errors

$$\begin{aligned}\bar{\sigma}_{\alpha^*} &= 1.918 \times 10^{-10} \text{ rad} = 0.0396 \text{ mas} \\ \bar{\sigma}_{\delta} &= 1.736 \times 10^{-10} \text{ rad} = 0.0358 \text{ mas} \\ \bar{\sigma}_{\mu_{\alpha}^*} &= 1.211 \times 10^{-17} \text{ rad/s} = 0.0788 \text{ mas/yr} \\ \bar{\sigma}_{\mu_{\delta}} &= 1.102 \times 10^{-17} \text{ rad/s} = 0.0717 \text{ mas/yr} \\ \bar{\sigma}_{v_R} &= 1.934 \text{ km/s} \\ \bar{\sigma}_{\varpi} &= 0.0494 \text{ kpc}^{-1} \quad \& \quad \bar{\sigma}_d = 0.0233 \text{ kpc}\end{aligned}$$

and an average relative parallax error of

$$\overline{\sigma_{\varpi}/\varpi} = 0.0298,$$

an improvement to Hinkel 2021 [17], which quotes an average relative parallax error of 0.086.

Across the full *Gaia* catalogue for DR2, the recorded typical errors for these parameters are of the same order of magnitude for our  $G$ -band of interest [24, 25].

### 2.2.2 Color–Magnitude Distribution over $G$ -Band & Giant Sub-sample

Using the selections from Hinkel 2021 [17], now allowing for nontrivial radial velocities and focusing on the  $14 < G < 18$  range, our stellar sample contained only 489 stars. As mentioned previously, radial velocity data for fainter stars is scarce in *Gaia*, and expanding the  $G$ -band to  $10 < G < 18$  can be done without worry for systematic biases. After our selection cuts, we get 695,345 stars.

Here we explore the stellar sample to study the distribution in the main se-

quence and giant populations. This has little bearing on the kinematic analysis to be conducted later, but it is valuable information nonetheless. To look at the main sequence/giant proportions in the sample, we construct a color–magnitude diagram (absolute magnitude “ $M_G$ ” vs BP–RP color “ $G_{BP} - G_{RP}$ ”) [26].  $M_G$  is found from the mean  $G$ –band magnitude (in mag) and the parallax (in mas) as  $M_G = G + 5 \log_{10}(\varpi) - 10$ , and the  $G_{BP} - G_{RP}$  color is the difference between the mean magnitudes obtained by integrating over the blue (330–680 nm) and red (630–1050 nm) passbands, respectively [15, 26]. One can use more advanced photometric techniques, for example, to distinguish giants and main sequence stars rigorously [27], but to a very rough estimate, giant stars occupy the low  $M_G$ , high  $G_{BP} - G_{RP}$  region of a color–magnitude diagram [28]. In Fig. 2.2, artificial criteria for giants that  $6(G_{BP} - G_{RP}) - 2.5 > M_G$  and  $M_G < 5$  mag are applied for convenience and to make calculations simple. The first condition cuts through the horizontal branch and splits the sample in two, then the second condition sharpens the scope to low  $M_G$ . Again, this is not a rigorous assessment, but a basic estimate of the proportion of giants in our sample.

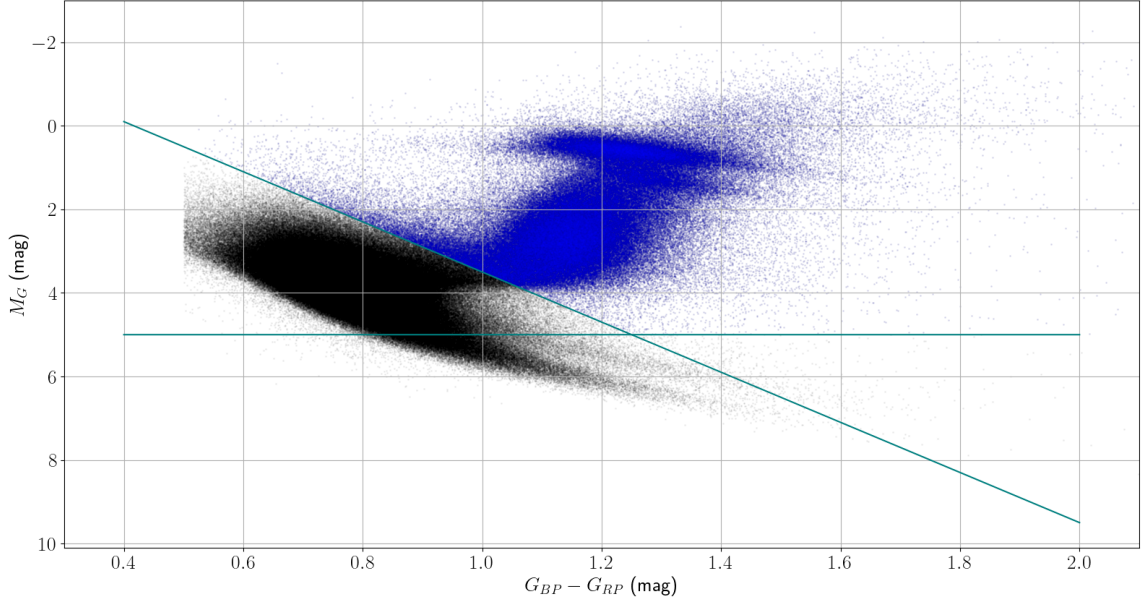


Figure 2.2: Color–magnitude diagram (CMD) for the full  $10 < G < 18$  sample. A cutoff has been set so that above both light blue lines, stars are considered giants, and otherwise, are considered main sequence. This yields 203,629 (29.3%) giants and 491,716 (70.7%) main sequence stars. Compared to Yanny & Gardner 2013 which observes only 0.04% giant stars [8], giants are much more appreciable in our sample because we have included more bright–end–magnitude stars ( $10 < G < 18$ ).

This analysis is repeated in 1 magnitude slices of  $G$  to see how the counts change.

The results are as follows:

Table 2.1: Giant/Main Sequence Proportions for 1 magnitude slices in  $G$ –band.

$G$ –Band	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15
Main Sequence	31215(36.8%)	136465(66.8%)	294502(80.1%)	29075(75.6%)	451(93.6%)
Giant	53545(63.2%)	67713(33.2%)	72949(19.9%)	9391(24.4%)	31(6.4%)

$G$ –Band	15 – 16	16 – 17	17 – 18
Main Sequence	6	1	0
Giant	0	0	0

Notice the statistics drop sharply after  $G \sim 14, 15$ .

### 2.3 Stellar Density Variation

First, we take a histogram of the stellar number counts along  $z$ , divided among 600 bins. From the raw number counts, we divide each bin by the selection function (see Yanny & Gardner 2013 [8])

$$\mathcal{V}(z) = \frac{1}{2}\delta(l_2 - l_1)z^2 \left( \frac{1}{\sin^2 b_1} - \frac{1}{\sin^2 b_2} \right) \quad (2.1)$$

to obtain the number density of stars within each bin ( $n(z) = N(z)/\mathcal{V}(z)$ ). Note that  $\delta$  is the width per bin (kpc) and  $z$  is the midpoint of a given bin (kpc) [8]. For our geometry,  $l_1 = 0$ ,  $l_2 = 2\pi$ ,  $b_1 = \frac{\pi}{6}$ ,  $b_2 = \frac{\pi}{2}$ , and  $\delta = \frac{6.0 \text{ kpc}}{600 \text{ bins}} = 0.01 \text{ kpc/bin}$ .

We use `scipy`'s curve fit method to fit the density data to the functional forms

$$A \operatorname{sech}^2 \left( \frac{x + B}{C} \right) \quad (2.2)$$

$$A \left[ \operatorname{sech}^2 \left( \frac{x + B}{2C} \right) + D \operatorname{sech}^2 \left( \frac{x + B}{2E} \right) \right] \quad (2.3)$$

The first ordinary  $\operatorname{sech}^2$  form (2.2) was derived in Spitzer 1942 [29], wherein an isothermal, axisymmetric thin disk is determined to have a  $\operatorname{sech}^2$  distribution in number density over galactic  $z$ , as the solution to the collisionless Boltzmann equation and the Poisson equation [30].

However, we model the Milky Way with two distinct disk components – the thin and thick disk – and fit the number density to a sum of  $\operatorname{sech}^2$  curves. There has, however, been recent debate about whether the thick disk is physically distinct from the thin disk or whether they form a continuum. Despite this, stars near the Sun show two distinct elemental abundance trends [5]. The second functional form (2.3) is the two-component model used in Widrow et al. 2012 [2].

For each bin of number density, we take Poisson errors from the square root of the number density. In `scipy`, the curve fit method optimizes the free parameters

(using the “Trust Region Reflective” algorithm [31]) to minimize the  $\chi^2$  parameter  $\left(\chi^2 = \sum \left(\frac{\text{ydata} - \text{fit}(\text{xdata}, * \text{popt})}{\sigma}\right)^2\right)$  where \*popt are the fit parameters, “xdata” is our independent variable  $z$ , and  $\sigma^2 = n(z)$  is the stellar number density – our “ydata” in this case. For details, see the scipy documentation of the curve fit method [32].

Fitting the selected stellar number density to the two functional forms, the resulting fit distributions are given in Fig. 2.3.

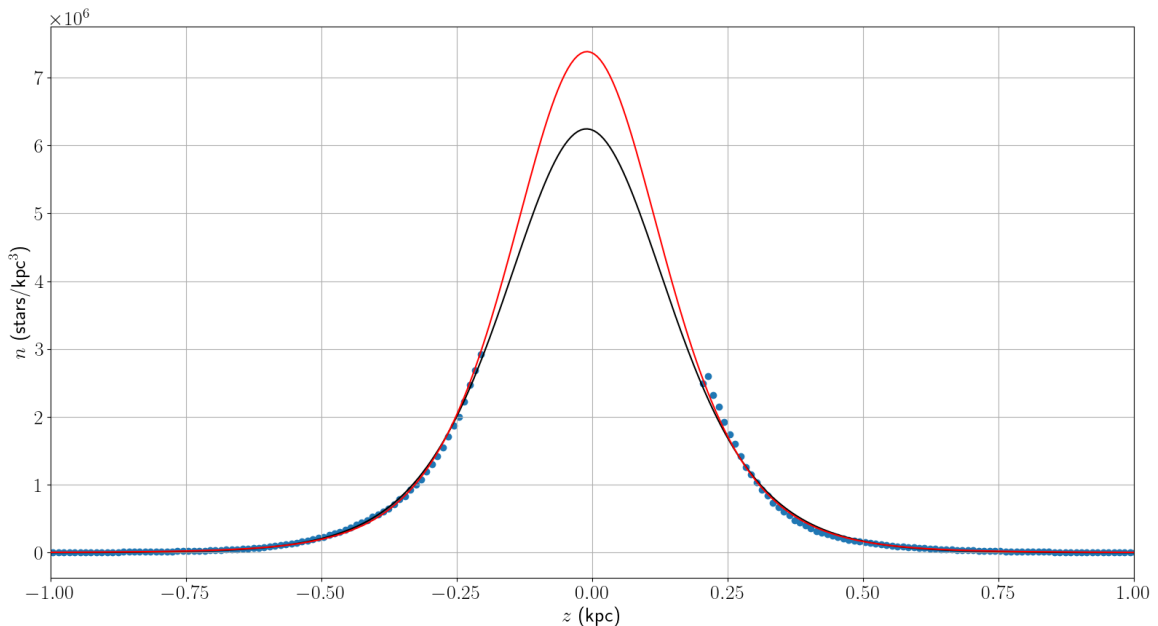


Figure 2.3: Distribution of stellar number density as a function of galactic  $z$ , looking just at the inner  $|z| < 1$  kpc for visual clarity. However, the fits are calculated for the full  $|z| < 3$  kpc range. The blue points (with error bars) are the stellar densities, the black curve is the single  $\text{sech}^2$  fit, while the red curve is the two-component  $\text{sech}^2$  fit. Despite appearances, the fits are very poor, taking reduced  $\chi^2$  values of  $\chi^2_\nu = 1113.43$  (single  $\text{sech}^2$ ) and  $\chi^2_\nu = 438.95$  (two-component), for 599 degrees of freedom. This is primarily due to the Poisson errors of the points being much smaller than the residuals from the fit lines. It seems that the artificial blending of the thin and thick disk components in the double  $\text{sech}^2$  fit form does not fully capture the morphology of our stellar sample. As observed by Widrow et al. 2012 and Yanny & Gardner 2013 [2,8], the residuals about the fits have wave-like structure as the points oscillate above and below the fit curves (see also Fig. 2.4).

For the single  $\text{sech}^2$  fit form (2.2), the calculated fit parameters and their standard errors are:  $A = 6.243 \times 10^6 \pm 2.079 \times 10^3$  stars/kpc<sup>3</sup>,  $B = 1.060 \times 10^{-2} \pm 1.470 \times 10^{-5}$  kpc, and  $C = 2.051 \times 10^{-1} \pm 2.337 \times 10^{-5}$  kpc. For the double  $\text{sech}^2$  fit form (2.3), the calculated fit parameters and their standard errors are:  $A = 7.260 \times 10^6 \pm 3.888 \times 10^3$  stars/kpc<sup>3</sup>,  $B = 9.658 \times 10^{-3} \pm 1.410 \times 10^{-5}$  kpc,  $C = 9.415 \times 10^{-2} \pm 2.883 \times 10^{-5}$  kpc,  $D = 1.677 \times 10^{-2} \pm 1.124 \times 10^{-4}$ , and  $E = 2.116 \times 10^{-1} \pm 3.096 \times 10^{-4}$  kpc. Comparing the double  $\text{sech}^2$  parameters to those in Yanny & Gardner 2013 and Widrow 2012 [2,8], there are marked differences largely owed to studying different populations of stars. The extension of this study’s *Gaia* sample to lower G band means that we measure brighter stars that are characteristically closer to us, at lower galactic  $|z|$ , but because of the  $z$ -dependence of the selection function, this bias should have little consequence.

Returning to the stellar number density fits, consider the normalized residuals of the data from the fits, which we denote  $\Delta \equiv (\text{data} - \text{fit}) / \text{fit}$ . Note that the fit values in the residuals are evaluated at the midpoints of each bin. In Fig. 2.4, we see the distribution of  $\Delta$  over galactic  $z$ .

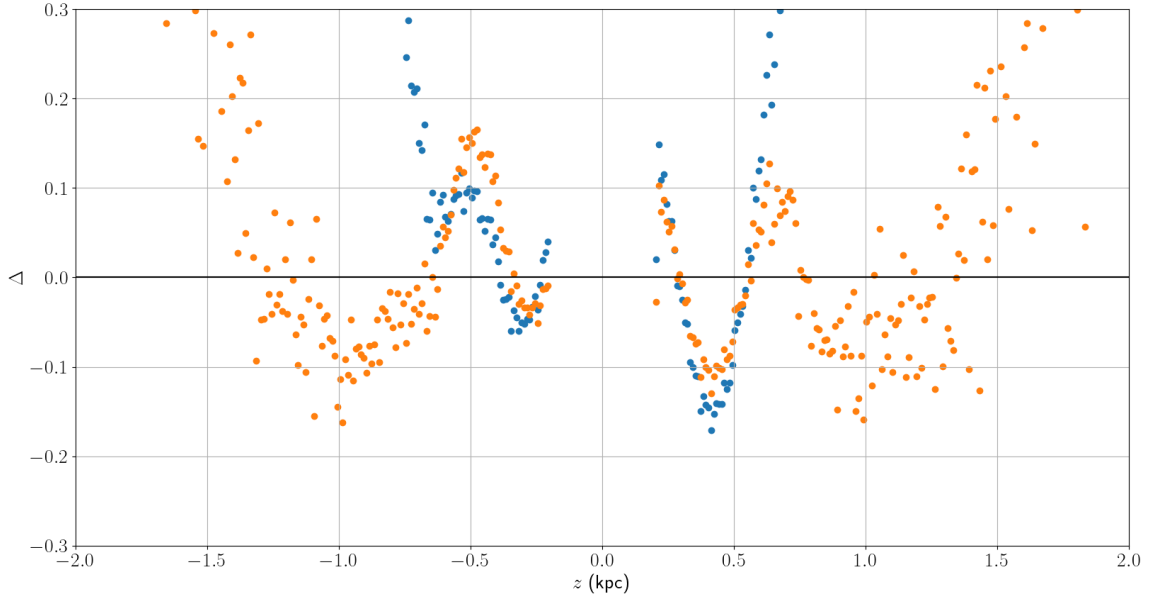


Figure 2.4: Plot of normalized residuals ( $\Delta \equiv (\text{data} - \text{fit}) / \text{fit}$ ) as a function of galactic  $z$ . For the sake of visualization, the domain in  $z$  is limited to  $|z| < 2$  kpc and the vertical range is narrowed. The blue points are the residuals for the single  $\text{sech}^2$  fit and the orange points are the residuals for the double  $\text{sech}^2$  fit. In both cases, the residuals have odd parity at low  $|z|$ , consistent with the literature [2]. Beyond around  $|z| \sim 0.6$  kpc, the single  $\text{sech}^2$  fit becomes ill-behaved as the data increasingly deviates from the fit. Not shown here, the single  $\text{sech}^2$  fit residuals reach a peak at  $|z| \sim 1.5$  kpc then decrease and return to the zero line at about  $|z| \sim 2.3$  kpc. The double  $\text{sech}^2$  fit residuals appear to agree with the residuals found in Widrow et al. 2012 [2], confirming the wavy north-south asymmetry in the stellar density.

## Chapter 3 Kinematic Analysis

### 3.1 Toomre Cutoff Method and Phase Space Behavior

We want to work with galactocentric velocities in the Local Standard of Rest (LSR) frame, the rest frame of a star circularly orbiting the Milky Way galactic center (GC) at the position of the Sun. See Appendix A about the relevant coordinate conversions to this frame. With respect to the LSR frame, the orthogonal velocities in the directions radial to the GC, collinear with the direction of rotation, and perpendicular to the galactic midplane are denoted  $U_{LSR}$ ,  $V_{LSR}$ , and  $W_{LSR}$ , respectively.

A Toomre energy diagram is a plot of the projected velocity in the  $U_{LSR}/W_{LSR}$  plane, that is, of  $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$  [19]. The literature (e.g Buder 2019 [9], Amarsi 2019 [33], Bensby 2014 [5]) agrees that in the LSR frame, a simple cutoff of  $\sqrt{U_{LSR}^2 + V_{LSR}^2 + W_{LSR}^2} < 180$  km/s is an adequate condition for identifying disk stars (no distinction between thin/thick disk). However, there are more accurate, complex methods, like differentiating based on stellar age or by employing more sophisticated kinematic techniques. Some such techniques also allow for the distinction of the thin/thick disk and stellar halo components.

### 3.2 Kinematic Probability Method

A weakness of the Toomre method is the failure to account for the asymmetric velocity drift between the thin and thick disks in  $V_{LSR}$ , as the thin and thick disks have distinct mean velocities along the direction of galactic rotation (see Table 3.1). This is illustrated in Figure 3.1, where binning over  $V_{LSR}$  reveals the convolution of the galactic components, of which the thin/thick disk signatures are markedly different. The Toomre method ignores the complexity of the convolved disk and assumes too simple a model for the Galaxy. It offers a basic proxy for the general population trends in the sample, but fails to account for the richer detail in the phase space, e.g.



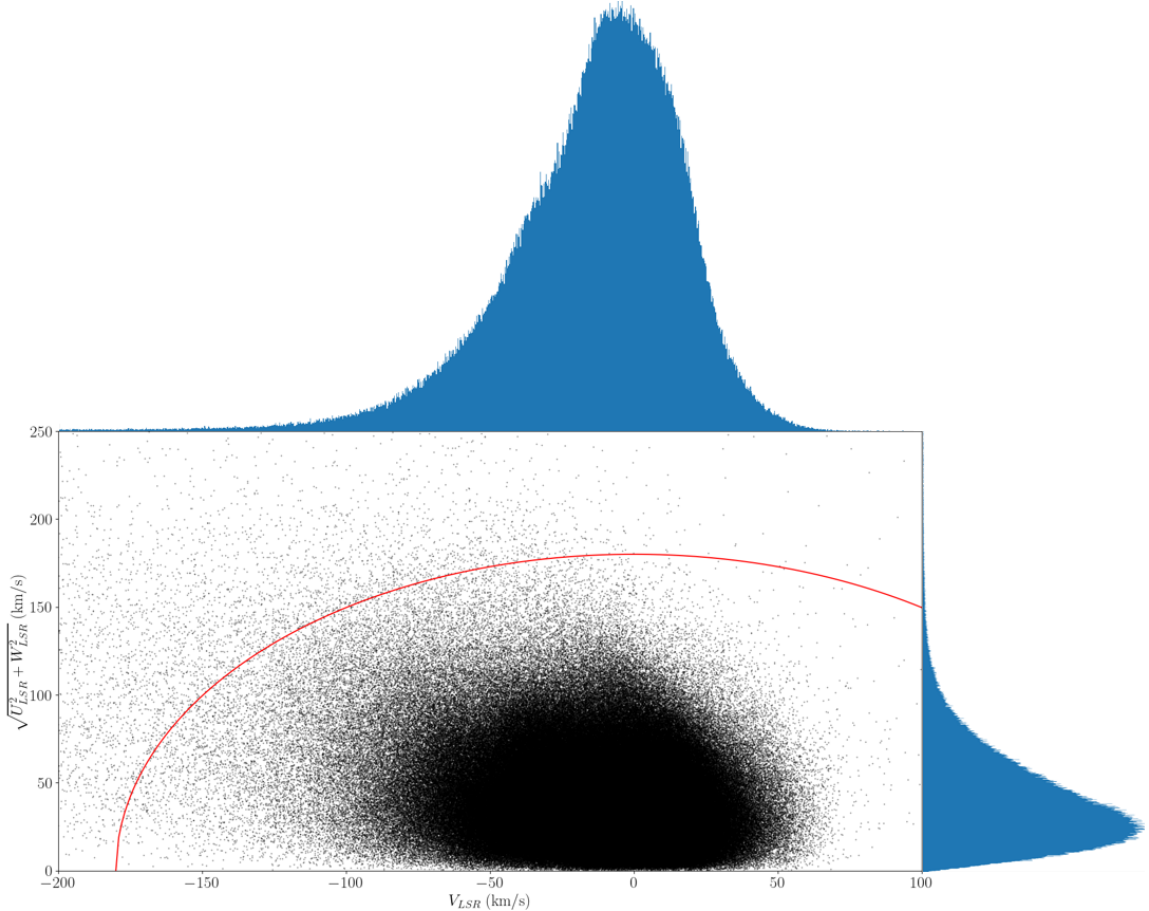


Figure 3.1: Toomre energy diagram for the stellar sample, plotting  $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$ . The red line traces the simple disk/halo cutoff, where  $\sqrt{U_{LSR}^2 + V_{LSR}^2 + W_{LSR}^2} = 180$  km/s. We have also projected the data by binning along the axes. Each histogram contains  $\lfloor \frac{695345}{200} \rfloor = 3476$  bins. In this case, the cutoff method yields 688609 (99.03%) disk stars and 6736 (0.97%) halo stars. Our Toomre diagram and histogram distributions are consistent with the results of Buder 2019 [9], yet offer better statistics and thus higher resolution.

the thin/thick disk, kinematic overlap between the disk and halo, etc.

The more sophisticated kinematic probability method revolves around deconvolution of the Galactic components into distinct Gaussian velocity features in the LSR frame. For each component, the corresponding distribution function takes the form

$$f_i = \frac{1}{(2\pi)^{3/2}\sigma_U\sigma_V\sigma_W} \exp\left(-\frac{(U_{LSR} - U_{asym})^2}{2\sigma_U^2} - \frac{(V_{LSR} - V_{asym})^2}{2\sigma_V^2} - \frac{W_{LSR}^2}{2\sigma_W^2}\right)$$

where  $(\sigma_U, \sigma_V, \sigma_W)$  are the velocity dispersions and  $(U_{asym}, V_{asym})$  are the mean velocity values *unique* to each component (see Table 3.1). It is assumed that  $W_{LSR}$  averages around zero for each of the galactic components [5].

The probabilities of a given star belonging to particular Galactic components are found via

$$P_i = X_i f_i$$

where  $X_i$  is the observed fraction (in the Solar neighborhood) of the component of interest. For the thin disk, thick disk, and stellar halo, we denote the 3 probabilities  $D$ ,  $TD$ , and  $H$  respectively.

Prior studies with deconvolution algorithms (see Soubiran 2003 [34] for example) have deduced the overall population fractions, characteristic dispersions and velocity asymmetries of the different Galactic components. We use the particular values summarized in Bensby et al. 2014 [5], which are tabulated below.

Table 3.1: Dispersion,  $U, V$  asymmetry, and overall population fraction values for the Gaussian velocity signatures of the thin disk, thick disk, and halo. Note that the population fractions do not sum to 1, because Bensby et al. 2014 accounts for the proportion of stars in the Hercules stream, which make up the remaining 6% of their sample [5].

	$\sigma_U$ (km/s)	$\sigma_V$ (km/s)	$\sigma_W$ (km/s)	$U_{asym}$ (km/s)	$V_{asym}$ (km/s)	$X$
Thin Disk	35	20	16	0	-15	0.85
Thick Disk	67	38	35	0	-46	0.09
Halo	160	90	90	0	-220	0.0015

### 3.2.1 Theory and Kinematic Criterion

Among kinematics, metallicity, and stellar age techniques for finding thin/thick disk and halo stars, there is a common issue of “contamination” of one galactic component into another, where the components overlap in phase space and stars in one component can be incorrectly labeled as belonging to another (Carrillo 2019 [35]). To mitigate component contamination, we define transition regions within which stars cannot be confidently assigned to a single population within a galactic component. To decide which population/transition region a star belongs to, we set inequality conditions on the probability ratios  $TD/D$  and  $TD/H$ . The choice of cutoff in the inequalities is subjective – stricter cutoffs can be enforced to be increasingly certain of the galactic components. For the most part, the inequality cutoffs have been chosen so that the probability of being in one population distribution is twice as likely as

being in another distribution. The criteria chosen are:

$$\text{(Thin Disk)} : \frac{TD}{D} < 0.5 \quad (3.1)$$

$$\text{(Thin/Thick Transition)} : 0.5 < \frac{TD}{D} < 2.0 \quad (3.2)$$

$$\text{(Thick Disk)} : \frac{TD}{D} > 2.0 \ \& \ \frac{TD}{H} > 2.0 \quad (3.3)$$

$$\text{(Thick/Halo Transition)} : \frac{TD}{D} > 2.0 \ \& \ 0.5 < \frac{TD}{H} < 2.0 \quad (3.4)$$

$$\text{(Stellar Halo)} : \frac{TD}{D} > 2.0 \ \& \ \frac{TD}{H} < 0.5 \quad (3.5)$$

Two conditions are enforced for the thick disk, halo, and thick disk/halo transition regions as an additional measure for contamination from the thin disk. For example, in identifying thick disk stars, one should check simultaneously that a particular star is likely not a thin disk star ( $TD/D$  condition) and not a halo star ( $TD/H$  condition).

### 3.2.2 Memberships in Phase Space

As a quick point of comparison, examining how probability ratios vary within the phase space of a Toomre diagram gives broad perspective to how the cutoff method and kinematic probability method perform.

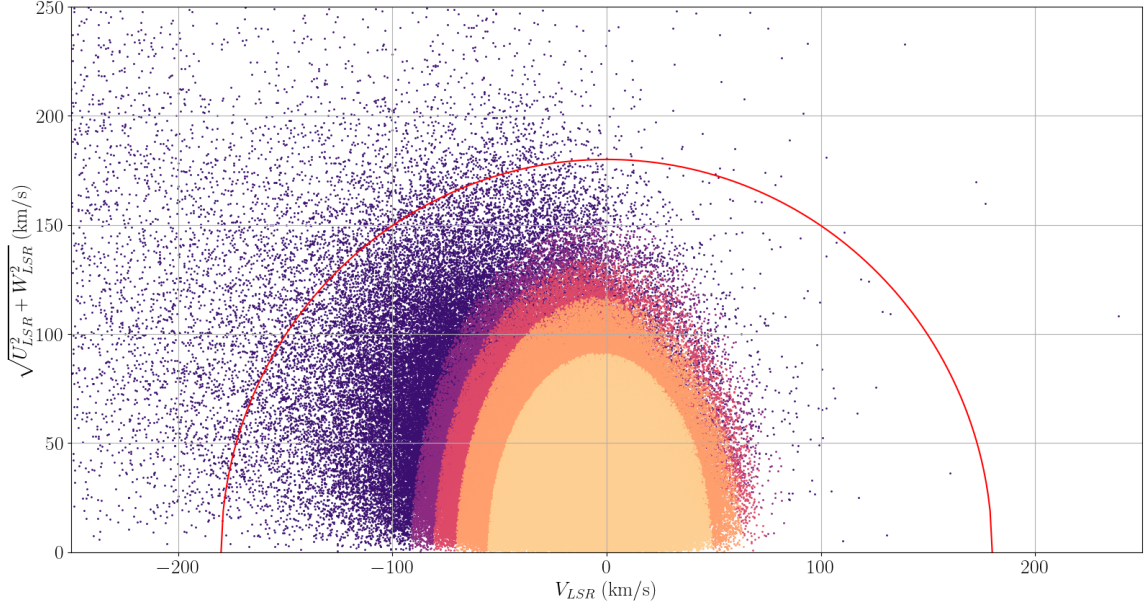


Figure 3.2: Toomre diagram ( $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$ ), color coded according to different regimes of  $TD/D$ . Light orange is  $TD/D < 0.1$ , orange is  $0.1 < TD/D < 0.5$ , pink is  $0.5 < TD/D < 2$ , purple is  $2 < TD/D < 10$ , and dark purple is  $TD/D > 10$ . The red line is once again the 180 km/s cutoff. Generally, stars closer to the origin of the phase space tend to be dominantly (thin) disk stars.

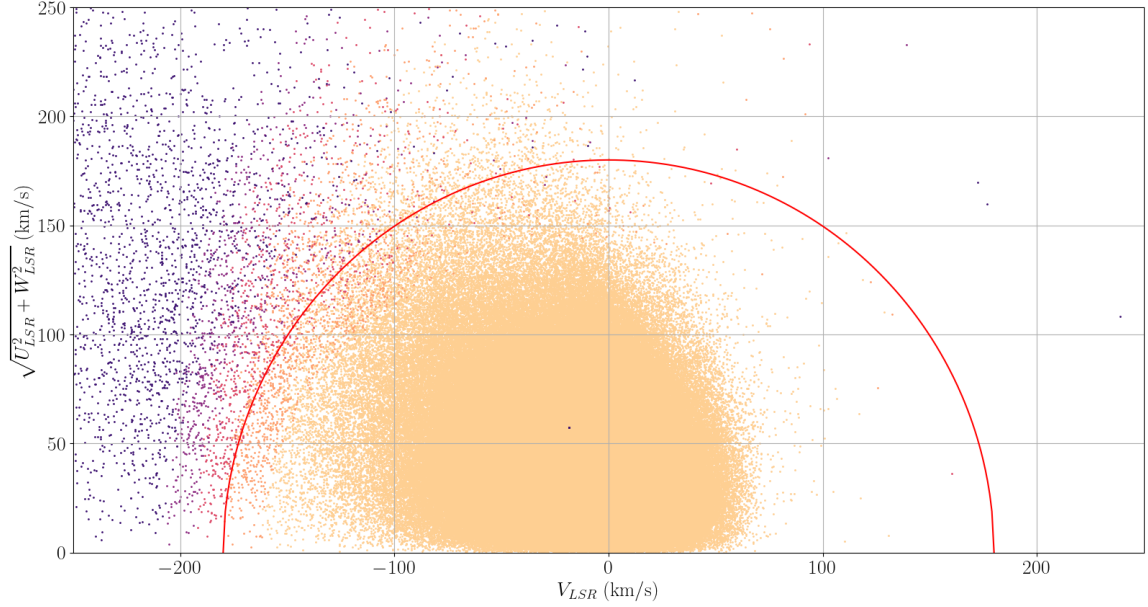


Figure 3.3: Toomre diagram ( $\sqrt{U_{LSR}^2 + W_{LSR}^2}$  vs  $V_{LSR}$ ), color coded according to different regimes of  $TD/H$ . Light orange is  $TD/H > 10$ , orange is  $2 < TD/H < 10$ , pink is  $0.5 < TD/H < 2$ , purple is  $0.1 < TD/H < 0.5$ , and dark purple is  $TD/H < 0.1$ . The red line is the 180 km/s cutoff.

In these figures, stars outside the red semicircle in the Toomre phase space tend to have  $TD/H < 0.5$ , while simultaneously being in the large  $TD/D$  regime from Figure 3.2, which is consistent with the kinematic criterion for halo stars. Thus the Toomre cutoff method captures the same general trends in the phase space of halo stars as with the probability ratio method. There is variation with  $V_{LSR}$ , however, because the thin/thick disk and halo components have different  $V_{LSR}$  asymmetry values. Using kinematic probabilities accounts for the component velocity dispersions and asymmetries that are convolved in the sample, details missed by the Toomre cutoff method.

### 3.3 Population Variation & Error Estimation via Bootstrapping

Thus far, we have looked at the holistic phase space behavior of our sample – now we transition to studying how populations change with galactic  $z$ . For each hemisphere of the galaxy, we split our sample into 50 sub-slices in  $z$ . For the stars in each sub-slice, we discern the number within each galactic component and determine the proportion of stars within each population for the given sub-slice. The choice of 50 sub-slices results in a reasonable balance in binning which yields both small-scale structure and larger trends in the population fractions.

As an error estimation technique, bootstrapping involves taking a particular data set, randomly sampling from it while allowing for *repetition of data points* – as opposed to permutative sampling – and thereby generating a number of synthetic data sets, of the same size as the original set. Then, a quantity of interest from *each* synthetic data set is calculated, yielding an approximate ‘bootstrap distribution’ of values from which a standard deviation can be drawn (see pages 3 & 4 in Efron 1979 [21]).

In context of this study, we seek an error assessment for the population fractions calculated for each sub-slice of our sample. Stars in a given sub-slice are re-sampled with repetition to generate 50 synthetic data sets. From each synthetic data set, we find the fraction (%) of each population. Then, for a population of interest, the 50 fractions are averaged and a standard deviation is drawn from them.

Note that moving away from the galactic mid-plane, there are increasingly less stars per sub-slice, leading to worse statistics and greater error bars.

### 3.4 Results

Figures 3.4 and 3.5 show the distributions of the galactic component populations with  $z$ , North and South.

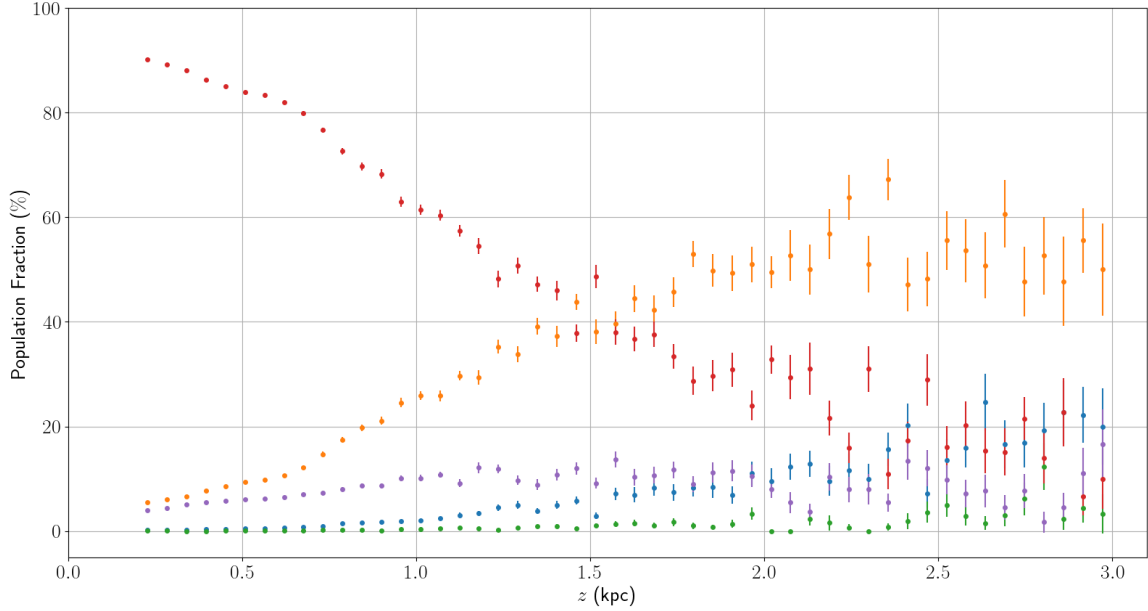


Figure 3.4: Proportions of membership within the different galactic components, as a function of  $z$ , for stars in the Northern hemisphere of the galaxy ( $z > 0$ ). Red is thin disk, orange is thick disk, purple is thin/thick disk transition, green is thick disk/halo transition, and blue is halo.

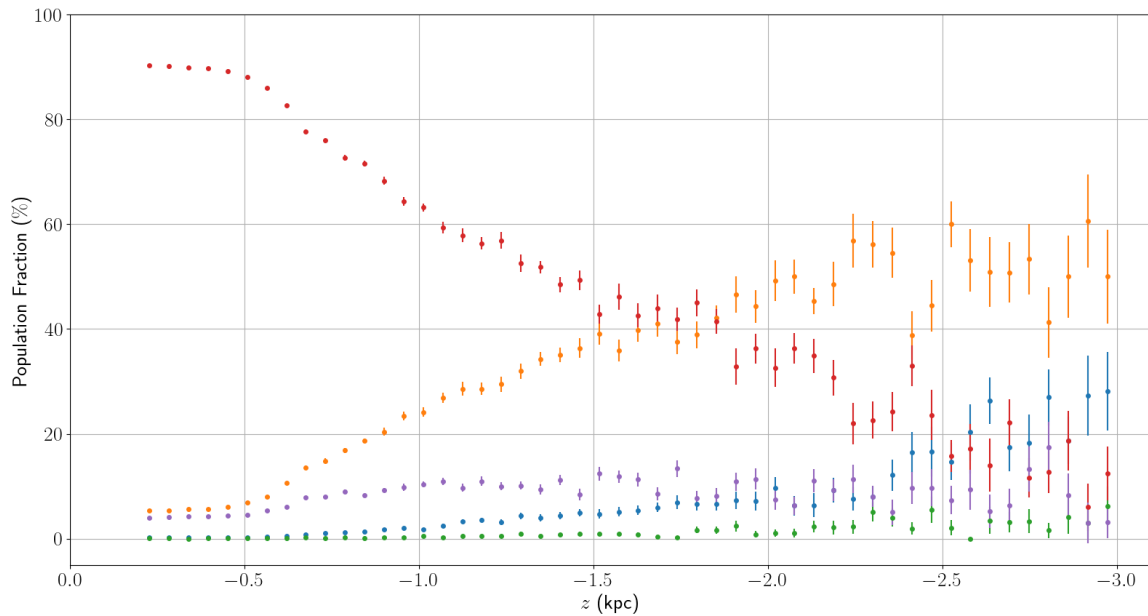


Figure 3.5: Proportions of membership within the different galactic components vs  $z$ , for stars in the Southern hemisphere of the galaxy ( $z < 0$ ). Red is thin disk, orange is thick disk, purple is thin/thick disk transition, green is thick disk/halo transition, and blue is halo.



Some features of interest in the population distributions:

In the north, the intersection point where the thin and thick disk fractions equal each other is around the [1.460,1.572] kpc bin, with midpoint  $1.516 \pm 0.056$  kpc (where half a bin width is taken as the error). In the south, this intersection is at the [1.684,1.796] kpc bin, with midpoint  $1.740 \pm 0.056$  kpc. Thus, the midpoints differ by  $224 \pm 79$  pc, a substantial discrepancy on the galactic scale.

Perhaps most significantly, looking at low  $|z|$ , the thin/thick disk populations remain essentially unchanged in the South for about the leftmost 5 data points in Figs. 3.4, 3.5 (a swath in  $z$  of 280 pc), whereas in the North the populations begin to change immediately as we move away from the mid-plane. Conversely, comparing North and South for the stellar halo, the distribution shape is more-or-less the same.

## Chapter 4 Study Outcomes and Conclusions

Throughout this thesis, we have detailed selection and processing of full 6D astrometric phase space data from stars in the *Gaia* database.

Working in the Local Standard of Rest frame, the phase space signatures of the thin disk, thick disk, and halo components of the Milky Way can be deconvolved. Partitioning the astrometric data along galactocentric  $z$ , the thin and thick disk population fractions are seen to vary asymmetrically North and South, to a significant degree.

### 4.1 Relation to Other Studies

To compare this work with studies of number count asymmetry like Yanny & Gardner 2013 [8] or Bennett & Bovy 2018 [10], I decided to construct an asymmetry parameter comparing population fractions North and South. Analogous to the number count asymmetry parameter, I take

$$\mathcal{A} \equiv \frac{f_i(z) - f_i(-z)}{f_i(z) + f_i(-z)} \quad (4.1)$$

for a fraction  $f$  of population  $i$ . Note that for this data, I chose to bootstrap errors in the population fractions using 100 synthetic data set per sub-slice, rather than 50 as done previously. From these bootstrap errors, the propagated error in the asymmetry is found via

$$\sigma_{\mathcal{A}} = \frac{\sqrt{2 [f_i^2(z) + f_i^2(-z)] [\sigma_{f_i(z)}^2 + \sigma_{f_i(-z)}^2]}}{(f_i(z) + f_i(-z))^2} \quad (4.2)$$

To get a clean asymmetry distribution, I decided to examine the population with the best statistics at low  $z$ , the thin disk. Calculating the asymmetry in each sub-slice, my results are summarized in Fig. 4.1.

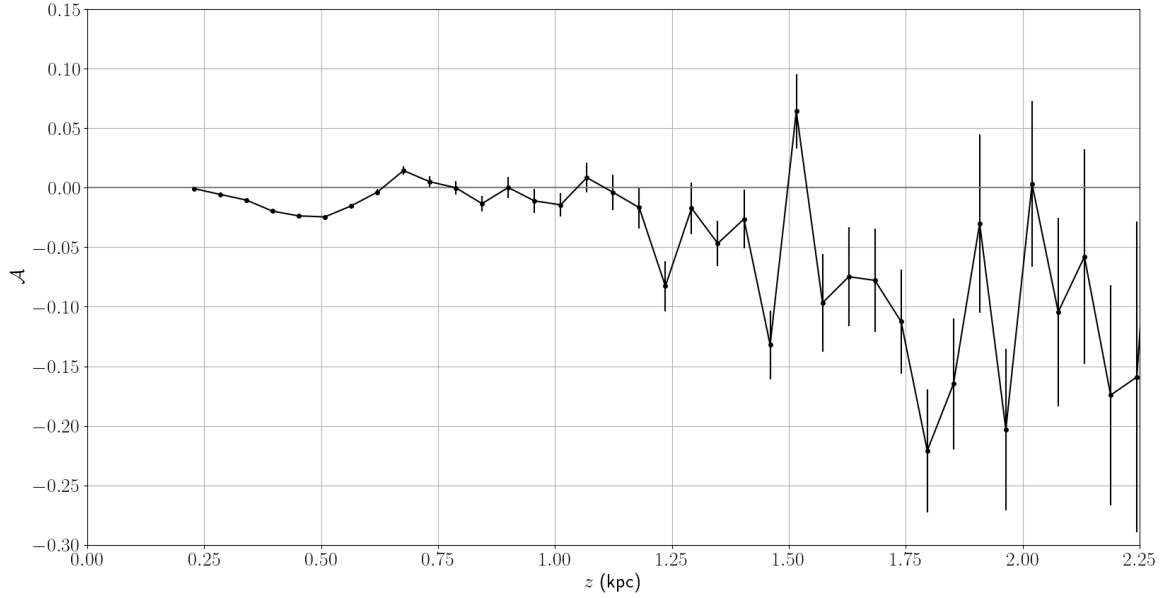


Figure 4.1: A plot of the population fraction asymmetry parameter  $\mathcal{A}$  versus galactic  $z$ , for the thin disk. There is surprisingly great agreement between the thin disk asymmetry and the number count asymmetry distributions. Comparing with Fig. 5 in Bennett & Bovy 2018 [10], which uses *Gaia* DR2 data to explore number count asymmetry, the locations in  $z$  of the peaks and dips line up extremely well. For example, in the number count asymmetry, there is a large dip just before 0.5 kpc, a small peak around 0.7 kpc and a small dip around 0.8 kpc, rising up to a small peak near 1.1 kpc, etc. [10]. Yanny & Gardner 2013 give focus to the asymmetry features at 0.8 kpc and 1.5 kpc, using data from SDSS–DR9 [8]. These oscillatory features are also seen in the population fraction asymmetry.

## 4.2 Thin Disk Population Asymmetry for Giant and Main Sequence Sub-Samples

Using the methodology in chapter 2, I separate my sample into “main sequence” and “giant” stars based on the photometric criteria used in Fig. 2.2. Then, I deduced the population fractions, and thus thin disk asymmetry using the analysis of the previous section, for each of these sub-samples. The distributions are given in Figs. 4.2 and 4.3 below.

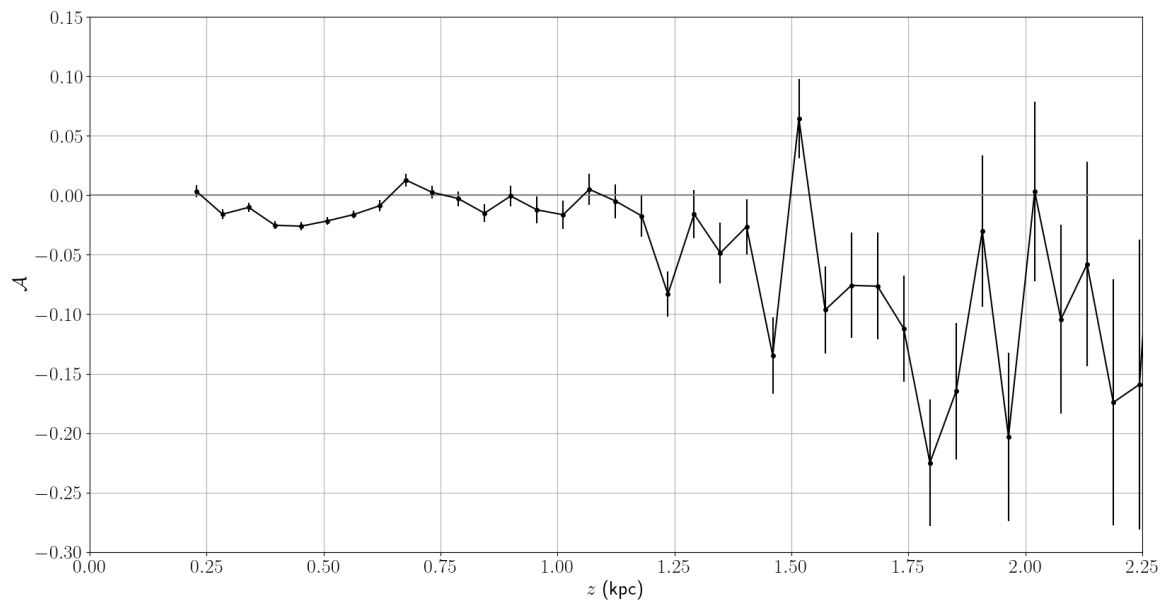


Figure 4.2: A plot of the population fraction asymmetry parameter  $\mathcal{A}$  versus galactic  $z$ , for the thin disk, using the sub-sample of giant stars. The asymmetry distribution of the giant sample is very similar to that of the combined sample in Fig. 4.1, suggesting similar expression of non-steady state effects.

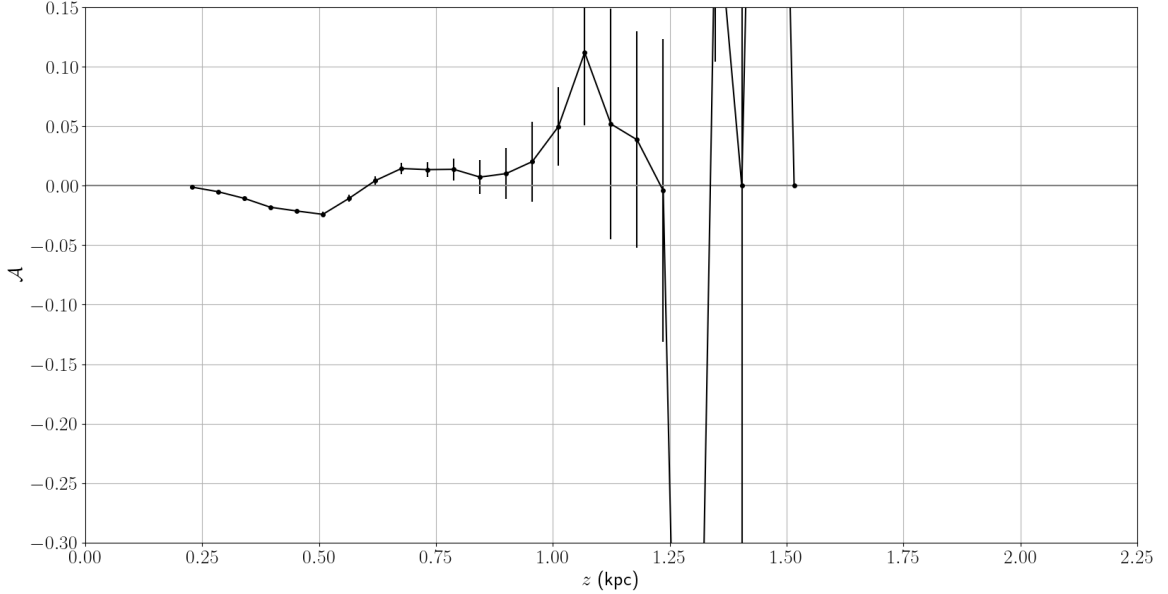


Figure 4.3: A plot of the population fraction asymmetry parameter  $\mathcal{A}$  versus galactic  $z$ , for the thin disk, using the sub-sample of main sequence stars. The domain and range are restricted to that in Fig. 4.2 for the sake of comparison. Both the main sequence and giant samples agree on the dip near 0.5 kpc and the peak near 0.7 kpc, but the main sequence asymmetry distribution shows unique features beyond  $z \sim 0.8$  kpc. Moreover, due to deficiency in main sequence stars, there are no population fractions and thus no asymmetry values beyond  $z \sim 1.5$  kpc.

### 4.3 Future Work

This study is based on data from *Gaia*'s second data release, but recently – in June 2022 – the full third data release (DR3) was published. DR3 is especially wonderful for 6D kinematic studies like mine because the number of sources with radial velocities has increased from DR2/EDR3 by around a factor of 5 [36]. With all the same selections, I have run a query of DR3 to see how my sample compares. Before post-query cuts, DR3 yielded 3,497,716 stars, compared to 754,138 in this study. For a larger sample with – hopefully – smaller measurement errors, the natural next step for this work is to delve into DR3 and repeat the analysis with a new sample.

## Appendices

### Appendix A: Galactic Coordinate Conversions

For each star, the astrometric parameters we require from *Gaia* are equatorial right ascension ( $\alpha$ ) and declination ( $\delta$ ), galactocentric longitude ( $\ell$ ) and latitude ( $b$ ), parallax  $\varpi$ , equatorial proper motions ( $\mu_\alpha^*$  and  $\mu_\delta$ ), and radial velocity ( $v_r$ ).

#### Galactic Proper Motions

There are standard angles  $\alpha_G = 192.85948^\circ$  and  $\delta_G = 27.12825^\circ$  that we consider when constructing the galactocentric coordinate system [37]. Furthermore, we quote the distance to the galactic center (GC) as  $R_0 = 8178 \pm 13_{stat.} \pm 22_{sys.} pc$  (see GRAVITY Collaboration 2019 [38]).

$$\begin{aligned} C_1 &= \sin \delta_G \cos \delta - \cos \delta_G \sin \delta \cos(\alpha - \alpha_G) \\ C_2 &= \cos \delta_G \sin(\alpha - \alpha_G) \\ \mu_\ell &= \frac{1}{\cos^2 b} (C_1 \mu_\alpha^* + C_2 \mu_\delta) \\ \mu_b &= \frac{1}{\cos b} (C_1 \mu_\delta - C_2 \mu_\alpha^*) \end{aligned}$$

#### Galactocentric Velocities

Note that *Gaia* reports its angles in degrees, its parallaxes in *mas*, its proper motions in *mas/yr*, and its radial velocities in *km/s*.

From the parallax angle, the distance in *kpc* is just

$$d = \frac{1}{\varpi}$$

The physical velocities along  $\ell$  and  $b$  are thus

$$v_\ell = 4.74\mu_\ell d \cos b \quad , \quad v_b = 4.74\mu_b d$$

(the factor of 4.74 ensures the proper unit conversion of the velocities to  $km/s$ )

The galactocentric  $x$ ,  $y$ , and  $z$  coordinates are found by a spherical transformation

$$x = d \cos \ell \cos b - R_0, \quad y = d \sin \ell \cos b, \quad z = d \sin b$$

Along  $x$ ,  $y$ , and  $z$ , respectively, the galactocentric velocities  $U$ ,  $V$ , and  $W$  are:

$$U = v_r \cos \ell \cos b - v_\ell \sin \ell - v_b \cos \ell \sin b$$

$$V = v_r \sin \ell \cos b + v_\ell \cos \ell - v_b \sin \ell \sin b$$

$$W = v_r \sin b + v_b \cos b$$

### LSR Frame

The Local Standard of Rest is the rest frame of a star in a circular orbit about the GC at  $R_0$ . With respect to the LSR frame, propagating systematic/statistical errors, the

Sun has an observed relative motion of  $(U, V, W)_\odot = (11.1^{+1.21}_{-1.25}, 12.24^{+2.05}_{-2.05}, 7.25^{+0.62}_{-0.62}) km/s$  (see Schönrich/Binney 2009 [39]).

From  $x$ ,  $y$ , and  $z$ , we define the galactocentric longitude/latitude respectively as

$$\alpha = \arctan\left(-\frac{y}{x}\right) = \arctan\left(\frac{d \sin \ell \cos b}{R_0 - d \cos \ell \cos b}\right)$$

$$\beta = \arctan\left(\frac{z}{\sqrt{y^2 + x^2}}\right) = \arctan\left(\frac{d \sin b}{\sqrt{d^2 \sin^2 \ell \cos^2 b + (R_0 - d \cos \ell \cos b)^2}}\right)$$

Finally, the velocities in the LSR frame are

$$U_{LSR} = \cos \alpha \cos \beta (U + U_{\odot}) - \sin \alpha \cos \beta (V + V_{\odot}) - \sin \beta (W + W_{\odot})$$

$$V_{LSR} = \sin \alpha (U + U_{\odot}) + \cos \alpha (V + V_{\odot})$$

$$W_{LSR} = \cos \alpha \sin \beta (U + U_{\odot}) - \sin \alpha \sin \beta (V + V_{\odot}) + \cos \beta (W + W_{\odot})$$



## Bibliography

- [1] S. Gardner, A. Hinkel, and B. Yanny, “Applying Noether’s Theorem to Matter in the Milky Way: Evidence for External Perturbations and Non-steady-state Effects from Gaia Data Release 2,” *The Astrophysical Journal*, vol. 890, p. 110, February 2020.
- [2] L. M. Widrow, S. Gardner, B. Yanny, S. Dodelson, and H.-Y. Chen, “Galactoseismology: Discovery of Vertical Waves in the Galactic Disk,” *ApJL*, vol. 750, p. L41, May 2012.
- [3] T. Antoja, A. Helmi, M. Romero-Gómez, D. Katz, *et al.*, “A dynamically young and perturbed Milky Way disk,” *Nature*, vol. 561, pp. 360–362, September 2018.
- [4] A. Hinkel, S. Gardner, and B. Yanny, “Probing Axial Symmetry Breaking in the Galaxy with Gaia Data Release 2,” *The Astrophysical Journal*, vol. 893, p. 105, April 2020.
- [5] Bensby, T., Feltzing, S., and Oey, M. S., “Exploring the Milky Way stellar disk - A detailed elemental abundance study of 714 F and G dwarf stars in the solar neighbourhood,” *A&A*, vol. 562, p. A71, 2014.
- [6] P. K. Seidelmann and J. Kovalevsky, “Application of the new concepts and definitions (ICRS, CIP and CEO) in fundamental astronomy,” *A&A*, vol. 392, pp. 341–351, 2002.
- [7] B. Holl, “2D Gaia Nominal Scanning Law (NSL) movie.” <https://www.youtube.com/watch?v=IRhe2grA9wE>, October 2013.

- [8] B. Yanny and S. Gardner, “The Stellar Number Density Distribution in the Local Solar Neighborhood is North-South Asymmetric,” *ApJ*, vol. 777, no. 2, p. 91, 2013.
- [9] Buder, S., Lind, K., Ness, M. K., Asplund, M., Duong, L., Lin, J., Kos, J., Casagrande, L., Casey, A. R., Bland-Hawthorn, J., De Silva, G. M., D’Orazi, V., Freeman, K. C., Martell, S. L., Schlesinger, K. J., Sharma, S., Simpson, J. D., Zucker, D. B., Zwitter, T., Cotar, K., Dotter, A., Hayden, M. R., Hyde, E. A., Kafle, P. R., Lewis, G. F., Nataf, D. M., Nordlander, T., Reid, W., Rix, H.-W., Skúladóttir, Á., Stello, D., Ting, Y.-S., Traven, G., Wyse, R. F. G., and the GALAH collaboration, “The GALAH survey: An abundance, age, and kinematic inventory of the solar neighbourhood made with TGAS,” *A&A*, vol. 624, p. A19, 2019.
- [10] Morgan Bennett and Jo Bovy, “Vertical waves in the solar neighbourhood in *Gaia* DR2,” *Monthly Notices of the Royal Astronomical Society*, vol. 482, pp. 1417–1425, October 2018.
- [11] T. Prusti, “HIPPARCOS FREQUENTLY ASKED QUESTIONS.” <https://www.cosmos.esa.int/web/hipparcos/faqs-facts>.
- [12] Abdurro’uf, K. Accetta, C. Aerts, V. Silva Aguirre, *et al.*, “The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data,” *ApJS*, vol. 259, no. 2, p. 35, 2022.
- [13] E. A. Magnier, W. E. Sweeney, K. C. Chambers, H. A. Flewelling, *et al.*, “Pan-STARRS Pixel Analysis: Source Detection and Characterization,” *The Astrophysical Journal Supplement Series*, vol. 251, p. 5, October 2020.

- [14] Gaia Collaboration, A. Vallenari, A.G.A. Brown, T. Prusti, et al., “Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties,” *A&A*, vol. 595, p. A2, November 2016.
- [15] Gaia Collaboration, A. Vallenari, A.G.A. Brown, T. Prusti, et al., “Gaia Data Release 2. Summary of the contents and survey properties,” *A&A*, vol. 616, p. A1, 2018.
- [16] D. Ferguson, S. Gardner, and B. Yanny, “Milky Way Tomography with K and M Dwarf Stars: The Vertical Structure of the Galactic Disk,” *The Astrophysical Journal*, vol. 843, p. 141, July 2017.
- [17] A. Hinkel, “Axial Symmetry Tests of Milky Way Disk Stars Probe the Galaxy’s Matter Distribution,” *Theses and Dissertations—Physics and Astronomy*, vol. 89, 2021.
- [18] M. E. K. Williams, M. Steinmetz, J. Binney, A. Siebert, *et al.*, “The wobbly Galaxy: kinematics north and south with RAVE red-clump giants,” *Monthly Notices of the Royal Astronomical Society*, vol. 436, pp. 101–121, October 2013.
- [19] A. Sandage and G. Fouts, “New Subdwarfs. VI. Kinematics of 1125 High-Proper-Motion Stars and the Collapse of the Galaxy,” *AJ*, vol. 93, p. 74, 1987.
- [20] T. Bensby, Feltzing, S., and Lundström, I., “Elemental abundance trends in the Galactic thin and thick disks as traced by nearby F and G dwarf stars,” *A&A*, vol. 410, no. 2, pp. 527–551, 2003.
- [21] B. Efron, “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics*, vol. 7, no. 1, pp. 1 – 26, 1979.
- [22] J. Salgado, J. González-Núñez, R. Gutiérrez-Sánchez, J. Segovia, J. Durán, J. Hernández, and C. Arviset, “The ESA Gaia Archive: Data Release 1,” *Astronomy and Computing*, vol. 21, pp. 22–26, October 2017.

- [23] The European Space Agency, “Gaia Archive.” <http://gea.esac.esa.int/archive/>.
- [24] F. van Leeuwen, J. de Bruijne, F. Arenou, *et al.*, “Gaia DR2 documentation.” Gaia DR2 documentation, European Space Agency; Gaia Data Processing and Analysis Consortium. Online at <https://gea.esac.esa.int/archive/documentation/GDR2/index.html>, 2018.
- [25] Lindegren, L., Hernández, J., Bombrun, A., *et al.*, “Gaia Data Release 2 - The astrometric solution,” *A&A*, vol. 616, p. A2, 2018.
- [26] Gaia Collaboration, C. Babusiaux, F. van Leeuwen, M. A. Barstow, *et al.*, “*Gaia* Data Release 2: Observational Hertzsprung-Russell diagrams,” *A&A*, vol. 616, p. A10, August 2018.
- [27] A. Helmi, Z. Ivezić, F. Prada, *et al.*, “Selection of Metal-poor Giant Stars Using the Sloan Digital Sky Survey Photometric System,” *The Astrophysical Journal*, vol. 586, pp. 195–200, March 2003.
- [28] S. U. of Technology, “Hertzsprung-Russell Diagram.” <https://astronomy.swin.edu.au/cosmos/h/hertzsprung-russell+diagram>.
- [29] J. Spitzer, Lyman, “The Dynamics of the Interstellar Medium. III. Galactic Distribution.,” *ApJ*, vol. 95, p. 329, May 1942.
- [30] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*. 2008.
- [31] N. Mayorov, “Trust Region Reflective Algorithm.” <https://nmayorov.wordpress.com/2015/06/19/trust-region-reflective-algorithm/>.
- [32] The SciPy community, “scipy.optimize.curve\_fit.” [https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve\\_fit.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html).

- [33] Amarsi, A. M., Nissen, P. E., and Skúladóttir, Á., “Carbon, oxygen, and iron abundances in disk and halo stars - Implications of 3D non-LTE spectral line formation,” *A&A*, vol. 630, p. A104, 2019.
- [34] C. Soubiran, O. Bienaymé, and A. Siebert, “Vertical distribution of Galactic disk stars,” *A&A*, vol. 398, pp. 141–151, January 2003.
- [35] A. Carrillo, K. Hawkins, B. P. Bowler, W. Cochran, and A. Vanderburg, “Know thy star, know thy planet: chemo-kinematically characterizing TESS targets,” *Monthly Notices of the Royal Astronomical Society*, vol. 491, pp. 4365–4381, November 2019.
- [36] Gaia Collaboration, A. Vallenari, A.G.A. Brown, T. Prusti, et al., “Gaia Data Release 3. Summary of the content and survey properties,” *A&A*, 2022.
- [37] R. Poleski, “Transformation of the equatorial proper motion to the Galactic system,” *arXiv e-prints*, p. arXiv:1306.2945, 2013.
- [38] The GRAVITY Collaboration, Abuter, R., Amorim, A., Bauböck, M., *et al.*, “A geometric distance measurement to the Galactic center black hole with 0.3% uncertainty,” *A&A*, vol. 625, p. L10, 2019.
- [39] R. Schönrich, J. Binney, and W. Dehnen, “Local kinematics and the local standard of rest,” *Monthly Notices of the Royal Astronomical Society*, vol. 403, pp. 1829–1833, April 2010.

## Vita

Joshua Taylor Harry

### Place of Birth:

- West Palm Beach, Florida

### Education:

- B.S. in Physics – University of Kentucky, May 2021
- M.S. in Physics – University of Kentucky, August 2022 (expected)

### Professional Positions:

- Graduate Teaching Assistant, University of Kentucky Fall 2021–Spring 2022

### Awards & Honors

- Patterson Scholarship, University of Kentucky, Fall 2017–Spring 2021
- Lewis Honors College member, University of Kentucky, Fall 2017–Spring 2021
- Secretary of U.K. Society of Physics chapter, University of Kentucky, Fall 2019–Fall 2020
- Outstanding SPS Chapter, University of Kentucky, Spring 2020
- Barry Goldwater Scholarship Nomination, University of Kentucky, Spring 2020
- U. Kentucky Dept. of Physics & Astronomy Outstanding Senior Award, Spring 2021