Research Progress Based on Gaia DR2 and the Comparison of the Gaia Proper Motions to the Galaxia Model of the Milky Way

A Ritter¹ and Jiaju Li²

¹ Laboratory for Space Research, Hong Kong University, Pokfulam Rd, Hong Kong (SAR China)
² Aidi International School, Louzizhuang No. 7, Chaoyang, Beijing, China
E-mail: azuri.ritter@gmail.com

Abstract. The background of the Gaia Data Release 2 (DR2) is introduced and research progress based on Gaia DR2 is reviewed. Gaia DR2 coordinates and proper motions are transformed to Galactic coordinates to allow for a comparison of the Gaia data to models of the Galaxy. A comparison of the absolute predicted and measured numbers of stars per unit area on the sky as well as the corresponding proper motions in Galactic X, Y, and Z are made. It is found that there is a very big difference between the measurements taken from the Gaia Survey and the predictions made by the Galaxia Model of the Galaxy for the proper motions. A further in depth analysis, taking into account the proper motions as a function of distances, is required to fully understand the discovered discrepancies.

1. Introduction

The Gaia mission is a space telescope of the European Space Agency (ESA), designed to chart a three-dimensional map of more than one billion stars in our Galaxy, the Milky Way, in the process revealing the composition, formation and evolution of the Galaxy (http://sci.esa.int/gaia/). This massive stellar census will provide the basic observational data to analyze a wide range of important questions related to the origin, structure, and evolutionary history of our Galaxy which has not been possible before. The Gaia Data Release 2 [1] contains astrometric measurements (3D positions, radial velocities, proper motions) and key stellar parameters (e.g. spectral type, surface gravity, effective temperature, metallicity) of nearly 1.7 billion stars. The magnitude limit of Gaia is $G_{BP} = 25$ mag, which gives a maximum distance to a (bright) star that Gaia can measure of 25 kpc. Given that the position of Earth is $\approx 8.5$ kpc from the Galactic center, and the diameter of the Galaxy is 32.4 kpc, Gaia is able to measure at least every bright star in the Galaxy which is not hidden by dust clouds or overcrowding towards the Galactic bulge. During its anticipated lifetime of five years, Gaia will observe each of its more than one billion sources about 70 times, resulting in a record of the brightness and position of each source over time, continuously reducing the uncertainties in the measurements.

Galactic models like Galaxia [2] or the Besançon Model of the Galaxy [3] are powerful tools to test different scenarios of galaxy formation and evolution, theories of stellar formation and evolution, models of stellar atmospheres as well as dynamical constraints. They are also the key to predictions of future photometric, astrometric, and spectroscopic surveys. Most Galactic models incorporate analytical functions for density distributions, the age/metallicity relation and the Initial Mass Function (IMF) provided for each population of stars the Galaxy consists of. These populations are the thin and thick discs, the bar, the bulge, and the stellar halo with its globular clusters (Fig. 1). More advanced models
like *Galaxia* also incorporate N-body models for the known substructures like the Sagittarius stellar stream which may constitute a large fraction of the present halo [4, 5].

![Figure 1](image1.png)

**Figure 1.** Edge-on schematic of the main components of the Galaxy.

Combining *Gaia* DR2 with the *Galaxia* model of the Galaxy allows for a detailed comparison of the measured data to the predicted parameter distributions. Subtracting the smooth mean model parameters from the mean of the observed values [6] then reveals systematic differences which can be attributed to either new substructures of the Galaxy or real inconsistencies in the model. As an example, Fig. 2 shows the comparison of the mean radial velocity per pixel measured by the RAdial Velocity Experiment (RAVE) Survey [7] to 30 synthetic realisations of the survey predicted by the Besançon Model of the Galaxy. The recently in the RAVE data discovered Stream of Aquarius [8] is marked with a blue circle in the lower left corner.

![Figure 2](image2.png)

**Figure 2.** Example of a comparison between model (Besançon Model of the Galaxy) and survey (RAVE).

Shown here is the measured mean radial velocity per pixel compared to 30 realisations of a synthetic survey predicted by the model for giants. The plot has been smoothed once for better visibility. Contour lines show the differences of the mean values in steps of 0.6 standard deviations. Each pixel is 5 degrees wide in l and b, except of the lowest Galactic latitudes, where the swaths get increasingly wider in l due to decreasing numbers of stars. The blue circle in the bottom left marks the position of the recently in the RAVE data discovered Stream of Aquarius.

*Gaia* provides coordinates and velocities in the ICRS coordinate system. However, for Galactic research it is often desirable to use galactic coordinates instead of ICRS. Also most models of the Galaxy only give Galactic coordinates so in order to compare one data set to the other a coordinate transformation is required. In section 2 we will review the research that has been based on the *Gaia* DR2. In section 3 we describe in detail the transformation procedure for the *Gaia* coordinates and proper motions to Galactic coordinates. An example application of the transformed *Gaia* proper motions, namely the comparison of the measured parameter distributions for the proper motions to the ones predicted by the *Galaxia* Model of the Galaxy is given in section 4. Shortcomings of this preliminary comparison and future work will be described in section 5.
2. Research progress based on Gaia DR2

On the 25th of April 2018, the European Space Agency’s Gaia satellite task force announced that Gaia DR2 contains information on the positions on the sky, parallaxes, proper motions, and brightnesses of more than 1.69 billion stars, three-band photometry of about 1.38 billion stars, and radial velocities of more than 7 million stars.

This unprecedented batch of data not only provides excellent material and opportunity for astronomers studying the Milky Way, but also provides a wealth of information about the Milky Way’s star clusters, nearby dwarf galaxies, as well as our neighbouring galaxies, the Large and the Small Magellanic Clouds (LMC and SMC). Based on the Gaia DR2, Helmi et al. [9] obtained the proper motions of 75 Galactic globular clusters, 9 dwarf galaxies, 1 ultra-faint system, and the Large and Small Magellanic Clouds. They even derived the rotation curves for 5 globular clusters and the LMC, solely based on proper motions that are now competitive with line-of-sight velocity curves. The unprecedented depth, homogeneity, and precision of the Gaia DR2 data in both astrometric and photometric measurements has allowed astronomers to generate the most detailed maps to date.

Using the Gaia DR2 sky survey data, Zhang et al. [10] studied the characteristics and structure of the stellar population of the nearby open cluster (OC) Blanco 1. Using unsupervised machine learning they identified a total of 644 members of the cluster. For the first time, the existence of a leading as well as a trailing tidal tail was found in the cluster. The cluster has no obvious mass stratification and is in the early stage of dynamic evolution.

Based on Gaia DR2, Yang et al. [11] selected a sample of 24 young (< 3 Myr) pulsars with precise parallax measurements and measured the velocity of their local standard of rest (LSR) and the velocity dispersion among their respective local stellar groups, showing that Gaia DR2 is feasible and important to study the velocity of individual systems and the velocity distribution of neutron stars.

Ali & Alharbi [12] investigated the physical and kinematical characteristics of planetary nebulae accompanying PG 1159 stars based on the Gaia DR2 parallaxes and proper motions. They found that most of the studied nebulae arise from progenitor stars in the mass range 0.9 – 1.75\(M_\odot\) and tend to live within the Galactic thick disk, moving with an average peculiar velocity of 61.7 ± 19.2kms\(^{-1}\) at a mean vertical height of 469 ± 79 pc. The locations of the PG 1159 stars on the Hertzsprung–Russel diagram indicate that they have an average final stellar mass and evolutionary age of 0.58 ± 0.08\(M_\odot\) and 25.5 ± 5.3 \times 10^3 yr, respectively.

Xu et al. [13] used young massive stars in Gaia DR2 to trace the spiral arm structure. Although Gaia DR2 can only trace the Perseus, Local, and Sagittarius arms, its data in the fourth quadrant of the Milky Way complement the VLBA’s northern results. Lallement et al. [14] built a 3D map of the dust in the Local arm and the surrounding regions using a new hierarchical inversion algorithm. Chen et al. [15] used extinction to calculate the distance and size of the molecular clouds near the Local arm, showing that the molecular gas is extending to the Perseus arm at both ends. They also propose the existence of a spur in the fourth quadrant connecting the Local arm and the Sagittarius arm.

The search for OCs - the fundamental building blocks of the Galaxy - has made new progress with the release of data from the Gaia satellite. Stars at different distances can be distinguished by their parallaxes, and the accuracy of the proper motion of stars has also significantly improved. Currently, about 1,500 known clusters have been re-certified in the data. Cluster search and authentication throughout the galactic plane can be performed by many different methods. The most important ones are:

- **Visual examination**
  Using astrometry (proper motion, position, parallax) and photometric measurements from Gaia DR2, Sim et al. [16] visually detected 655 (207 of them new) star clusters within 1 kpc of the Sun, taking advantage of the fact that stars in clusters have similar proper motions and are spatially clustered.

- **UPMASK method based on K-means clustering**
  UPMASK is short for Unsupervised Photometric Membership Assignment in Stellar clusters and was developed by Krone-Martins and Moitinho [17]. This method is data driven - apart from the spatial uniformity of field members, UPMASK does not assume a priori parametrisations, such as isochrones or King profiles, of any probability distributions involved. It assumes that cluster members will be clustered in most spaces, including positional space. Field stars, even if clustered
in some spaces, are not expected to cluster in positional space. The core principles of this method can be generalized to other types of data. The UPMASK method does not rely on strong physical assumptions about cluster properties, but only requires that cluster members must have common properties and be more closely distributed in the sky than the random distribution between field stars.

• **SHip method based on FoF clustering**

SHip is short for Star Cluster Hunting Pipeline. In this method developed by Liu & Pang [18], the well established Friend of Friend (FoF) algorithm is first used to cluster stars in a five-dimensional parameter space. The clustering results are then classified and judged by isochrone fitting.

• **Methods based on DBSCAN clustering**

The density-based clustering algorithm (DBSCAN) is another unsupervised clustering algorithm [19]. This density-based algorithm makes use of the notion of distance between two sources in the data to define a set of nearby points as a cluster; it has the advantage over other methods of being able to find arbitrarily shaped clusters. Castro-Ginard et al. [20] used DBSCAN in combination with a supervised learning method such as an artificial neural network to automatically distinguish between real OCs and statistical clusters.

Already known OCs are easier to re-discover because known clusters usually have more members than new OCs, and the relatively low central star density and fewer member stars in new clusters prevented them from being previously detected. However, not all OCs can be re-certified, and only about half of the known clusters have been identified in Gaia DR2 [21]. Combinations of many factors can cause a cluster to be difficult to detect, such as the source density of the background, interstellar extinction, how populated a cluster is, its age, or how its proper motions differ from the field stars.

3. **The transformation procedure**

For the transformation procedure we followed the Gaia Data Release Documentation on the ESA website. The following equations were taken from there: Equation (3.58, here Eq. (1)) to convert Galactic Longitude $l$ and Galactic Latitude $b$ to $[X_{\text{Gal}}, Y_{\text{Gal}}, Z_{\text{Gal}}]$. Equations (3.65, here Eq. (2)) and (3.67, here Eq. (3)) to convert the proper motions to the Galactic Coordinate System, and Equations (3.72) to (3.80) for the propagation of the uncertainties. The program was written in python 3, making extensive use of the numpy library.

\[
\hat{r}_{\text{Gal}} = \begin{bmatrix} X_{\text{Gal}} \\ Y_{\text{Gal}} \\ Z_{\text{Gal}} \end{bmatrix} = \begin{bmatrix} \cos l \cos b \\ \sin l \cos b \\ \sin b \end{bmatrix}.
\] (1)

\[
\hat{p}_{\text{Gal}} = \begin{bmatrix} -\sin l \\ \cos l \\ 0 \end{bmatrix}, \quad \hat{q}_{\text{Gal}} = \begin{bmatrix} -\cos l \sin b \\ -\sin l \sin b \\ \cos b \end{bmatrix}.
\] (2)

\[
\vec{\mu}_{\text{Gal}} = \hat{p}_{\text{Gal}} \mu_l \cos b + \hat{q}_{\text{Gal}} \mu_b.
\] (3)

### 3.1 Comparison of the proper motions in Gaia and Galaxia: Running Galaxia

In order to keep the file sizes manageable we ran the Galaxia Model of the Galaxy a total of 650 times for overlapping cones of 157 square degrees, with a $V$ apparent magnitude limit of 21.5 mag. An example parameter file for a run centered on Galactic longitude $l = 175^\circ$ and Galactic latitude $b = 5^\circ$ is shown in Table 1. Using tools provided by the Galaxia model and described in http://galaxia.sourceforge.net/Galaxia3pub.html, the absolute magnitudes (in the Johnson–Cousins $UBV$ photometric system) of the synthetic stars were then artificially reddened and converted to apparent magnitudes. The 650 data files were then split up into 40,713 evenly sized pixels in the Hammer projection where each pixel projects the same area on the sky. By doing so we could avoid any projection effects which would have irritated the reader.

1https://gea.esac.esa.int/archive/documentation/GDR2/Data_processing/chap_cu3ast/sec_cu3ast_intro/sssec_cu3ast_intro_transforms.html
Example parameter file for a run of the Galaxia model of the Galaxy, centered on Galactic longitude of $175^\circ$ and Galactic latitude of $5^\circ$. For the meaning of each parameter please refer to http://galaxia.sourceforge.net/Galaxia3pub.html.

**Table 1.** Example parameter file for running *Galaxia*

| Parameter          | Value                           |
|--------------------|---------------------------------|
| outputFile         | galaxia_175.5                   |
| outputDir          | /data/galaxia/ubv_Vlt21.5_1.0/175_5 |
| photoSys           | UBV                             |
| magcolorNames      | V,B-V                           |
| appMagLimits[0]    | -1000                           |
| appMagLimits[1]    | 21.5                            |
| absMagLimits[0]    | -1000                           |
| absMagLimits[1]    | 1000                            |
| colorLimits[0]     | -1000                           |
| colorLimits[1]     | 1000                            |
| geometryOption     | 1                               |
| longitude          | 175                             |
| latitude           | 5                               |
| surveyArea         | 157.080000                      |
| fSample            | 1.0                             |
| pop1DID            | -1                              |
| warpFlareOn        | 1                               |
| seed               | 17                              |
| r_max              | 1000                            |
| starType           | 0                               |
| photoError         | 0                               |

3.2 **Comparison procedure**

For each pixel in the Hammer projection, we first transformed the *Galaxia* magnitudes from the UBV photometric system to *Gaia* $G_{BP}$ using the transformation described in Ritter & Huang [22]. We then selected all *Gaia* stars in the same pixel with $G_{BP} \leq \max(G_{BP,Galaxia})$ as the simulated data are not as deep as the *Gaia* data.

We then calculated the mean of the proper motions in $X$, $Y$, and $Z$ for the corresponding pixels in the Hammer projection in both data sets and plotted the absolute values as well as the differences as colour-coded maps (Fig.s 6 – 14).

4. **Results**

The results are visualized in Fig.s 3 to 14. Comparing the absolute numbers of stars in *Galaxia* and in *Gaia* (Fig.s 3 to 5) it is apparent that towards the Galactic plane there are many more stars predicted than measured. This can be explained with the fact that in the Galactic plane stars are too close to each other to be resolved by *Gaia*. In Fig. 5 the Magellanic clouds are easily visible as blue areas in the lower left.
Figure 3. Number of stars predicted by the Galaxia Model of the Galaxy. Note that the missing values in the center are due to a computer glitch.

Figure 4. Number of stars measured by Gaia. The Large and Small Magellanic Clouds can be seen in the bottom left.

Figure 5. Difference in the number in percent of stars predicted by the Galaxia Model of the Galaxy and measured by Gaia.

Comparing the mean values per pixel of the proper motions in $X$, $Y$, and $Z$ we find large systematic discrepancies, which is actually very surprising. Looking at the proper motions in $X$, both Fig. 6 and 7 show dipoles, however while in the model the positive values are towards the center of the Galaxy and the negative values are towards the edges of the Galaxy, the measured values by Gaia are positive towards the East and negative towards the West. This can also be seen in Fig. 8 where the difference in the proper motions between Galaxia and Gaia is shown.

Figure 6. Mean values of the proper motions in X predicted by the Galaxia Model of the Galaxy. Values are in km/s.
Figure 7. Mean values of the proper motions in X measured by Gaia. Values are in km/s.

Figure 8. Difference in the proper motions in X predicted by the Galaxia Model of the Galaxy and measured by Gaia. Values are in percent.

Figure 9. Mean values of the proper motions in Y predicted by the Galaxia Model of the Galaxy. Values are in km/s.

Figure 10. Mean values of the proper motions in Y measured by Gaia. Values are in km/s.

Figure 11. Difference in the proper motions in Y predicted by the Galaxia Model of the Galaxy and measured by Gaia. Values are in percent.
For the proper motions in Y (Figs. 9 to 11) there is a clear dipole in the predicted proper motions with positive values towards the East and negative values with similar amplitude to the West, while the measured proper motions show large negative values everywhere except for the far East and West where the values are slightly positive.

In the Z direction the proper motions predicted by Galaxia (Fig. 13) show a clear dipole with positive values towards the North and negative values towards the South. This is in stark contrast to the values measured by Gaia (Fig. 12) which show a quadrupole with positive values in the North East and South West and negative values towards the North West and South East. This is also reflected in the difference image shown in Fig. 14.

Figure 12. Mean values of the proper motions in Z predicted by the Galaxia Model of the Galaxy. Values are in km/s.

Figure 13. Mean values of the proper motions in Z measured by Gaia. Values are in km/s.

Figure 14. Difference in the proper motions in Z predicted by the Galaxia Model of the Galaxy and measured by Gaia. Values are in percent.

5. Shortcomings and future work

Our preliminary results still only look at the total numbers from all stars predicted in the Galaxia model. Given that the uncertainties in the measurements are increasing with the distance to the stars, we should really look at the proper motions as a function of the distances. Another shortcoming of the presented analysis is that we have not convolved the parameters predicted by the model with the uncertainties of the Gaia measurements. These shortcomings will be addressed in the subsequent further analysis of the data.
6. References
[1] Gaia Collaboration et al.: 2018, Astronomy and Astrophysics 616, 1
[2] Sharma S., Bland-Hawthorne J., Johnston K. V. and Binney J.: 2011, Astrophysical Journal 730, 3
[3] Robin, A. C., Reyle, C., Derriere, S., and Picaud, S.: 2003, Astronomy and Astrophysics 409, 523
[4] Ibata R. A., Gilmore G., Irwin M. J.: 1995, Monthly Notices of the Royal Astronomical Society 277, 781
[5] Chou M., Cunha K., Majewski S. R., Smith V. V., Patterson R. J., Martinez-Delgado, D., & Geisler, D.: 2010, Astrophysical Journal 708, 1290
[6] Ritter A.: 2012, PhD Thesis: Towards a better model of the Galaxy - Comparison of the RAVE spectroscopic survey to the Besancon Model of the Galaxy
[7] Steinmetz, M.: 2003, ASP Conference Series 298, 381
[8] Williams M. E. K., et al.: 2011, Astrophysical Journal 728, 102
[9] Helmi A., van Leeuwen F., McMillan P. J., Massari D., et al.: 2018, Astronomy and Astrophysics 616, A12
[10] Zhang Y., Tang S.-Y., Chen W. P., Pang X., Liu J. Z.: 2020, Astrophysical Journal 889, 99
[11] Yang, M., Dai S. Li D., Tsai C.-W., et al.: 2021, Research in Astronomy and Astrophysics 21, 141
[12] Ali A. and Alharbi W. R.: 2021, Research in Astronomy and Astrophysics 21, 151
[13] Xu Y., Bian S. B., Reid M. J., et al.: 2018, Astronomy and Astrophysics 616, L15
[14] Lallement R., Babusiaux C., Vergely J., Katz D., et al.: 2019, Astronomy and Astrophysics 625, A135
[15] Chen B.-Q., Li G.-X., Yuan H.-B., Huang Y., et al.: 2020, Monthly Notices of the Royal Astronomical Society 493, 351
[16] Sim G., Lee S. H., Ann H. B., and Kim S.: 2019, Journal of the Korean Astronomical Society 52, 145
[17] Krone-Martins A. and Moitinho A.: 2014, Astronomy and Astrophysics 561, A57
[18] Liu L. and Pang X.: 2019, ApJS 245, 32
[19] Ester M., Kriegel H.-P., Sander J., Xu X.: 1996, KDD'96: Proceedings of the Second
[20] Castro-Ginard A., Jordi C., Luri X., Julbe F., et al.: 2018 Astronomy and Astrophysics 618, A59
[21] Cantat-Gaudin T., Jordi C., Vallenari A., Bragaglia A., et al.: 2018, Astronomy and Astrophysics 618, A93
[22] Ritter A. and Huang C.: Transformations From Standard Photometric Systems To The Gaia Passbands, 2020 Journal of Physics: Conference Series 1593, 012039

Acknowledgements
This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.