Spatial Distribution of Variable Stars in the Milky Way Galaxy

Tim Higginbotham, Yazmin Suastegui-Perulero, Luke Miles

April 30, 2025

Abstract

This project explores the spatial distribution of variable stars within the Milky Way galaxy using datasets from AAVSO, Simbad, OGLE, ZTF, ASAS-SN, and Gaia. We pursued three main hypotheses: (1) variable stars show meaningful clustering by constellation and a correlation between their period and luminosity, (2) the types of variable stars differ across different constellations, and (3) Mira variables follow a coherent spatial structure that can be revealed by applying the Period-Luminosity Relation (PLR) and converting their positions into a Galactocentric frame. Together, these approaches shed light on the organization and evolution of different types of variable stars.

1 Introduction

Variable stars—stars whose brightness changes over time—are important astrophysical tools. Their variability can arise from pulsations, eclipses by companion stars, or intrinsic changes in their structure. Beyond their dynamic nature, these stars offer insight into stellar evolution and Galactic structure. This project began as a collaborative effort to better understand the large-scale distribution of these stars using publicly available datasets. Initial efforts focused on the American Association of Variable Star Observers (AAVSO) database, where we began examining positional trends among variable stars in different constellations. Over time, the project evolved to include more refined hypotheses, culminating in an in-depth spatial analysis of Mira variables using multi-survey crossmatching and distance estimation techniques.

Understanding the distribution of Mira variables across the Milky Way provides insight into stellar evolution and Galactic structure. Because Miras are older, intermediate-mass stars, they are excellent tracers of the Galactic disk, bulge, and possibly halo.

Hypothesis 1: Variable Stars and Their Constellations

This project began with an investigation into whether variable stars are evenly distributed across the night sky. Using the AAVSO database, one team member proposed studying the sky distribution of these stars by constellation and seeing if there was a trend with their period and luminosity. Without access to distance data in AAVSO alone, this analysis was constrained to 2D positional maps, supported by supplemental data from the Simbad database to gather parallax information for a subset of stars.

While this hypothesis was more exploratory, it helped establish foundational tools for catalog access, plotting, and general data exploration. Clustering effects were noted qualitatively, though further refinement was needed to make meaningful distance-based inferences. It is noted in figure 1 that a weak correlation between period and magnitude for all variable stars was found, but the scope was generally too broad to make a significant conclusion.

Correlation Matrix

	Period	MagDelta
Period	1.000000	0.227278
MagDelta	0.227278	1.000000

Figure 1: Correlation Matrix of Period and Change in Luminosity (MagDelta) for all variable stars.

Hypothesis 2: Variation of Star Types Across Constellations

The second hypothesis explored the distribution of variable star types within and across constellations. Datasets were parsed using pandas, and new columns were added to calculate counts and proportions of each star type. Bar graphs were created to visualize both the global distribution across 14 constellations and individual breakdowns for each region.

The most common types were SR—M (semi-regular or Mira-type) and E (eclipsing) variables. Monoceros and Crux, for example, were dominated by eclipsing variables, while Cassiopeia and Sagittarius showed a higher prevalence of SR and SR—M types. In nearly all cases, less common variable types composed less than 5% of the sample. Limitations included the lack of consistent positional precision and distance information in the AAVSO dataset, making 3D spatial analysis impossible at this stage. Figure 2 shows the variable star distribution for Sagitarrius constellation, visualizing the breakdown.

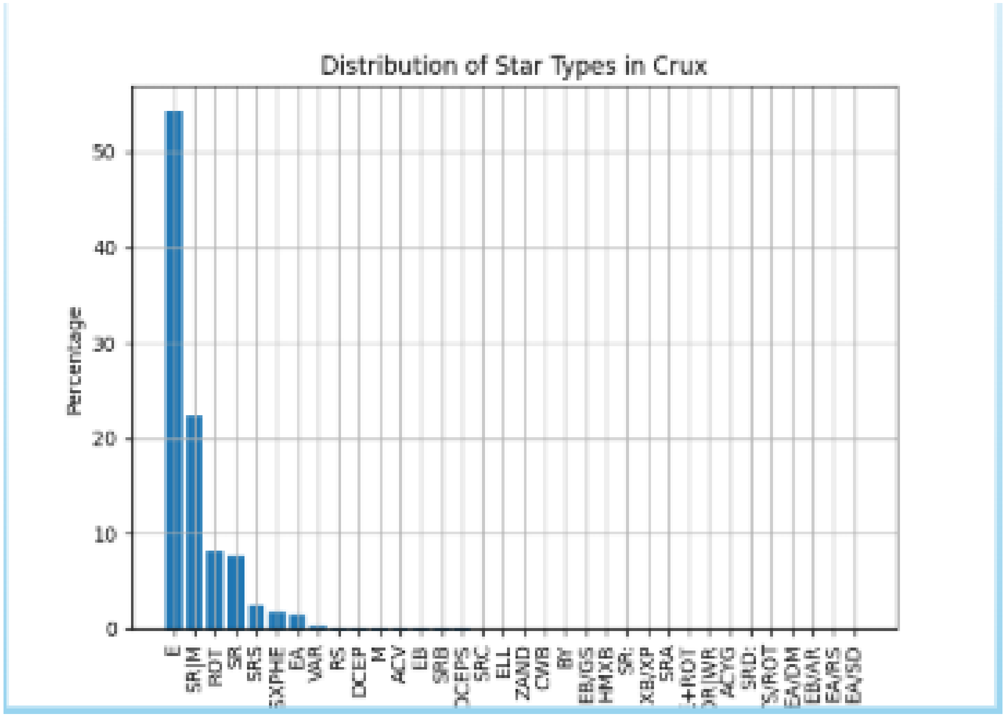

Figure 2: Histogram of variable star types in the Sagittarius Constellation.

Hypothesis 3: Spatial Structure of Mira Variables in the Galaxy

The third and most detailed hypothesis involved a spatial mapping of Mira variables using precise distance measurements. Unlike the broader AAVSO dataset, this portion of the study used data from OGLE, ASAS-SN, and ZTF, cross-matched with Gaia and AllWISE catalogs. K-band magnitudes were acquired primarily from AllWISE, and parallax/radial velocity/proper motion data from Gaia DR3.

Absolute magnitudes for each Mira were calculated using a fitted Period-Luminosity Relation (PLR), filtered and aligned with literature fits by Whitelock (2008) and Feast (1989). Using these fits, distances were computed and stars were transformed into Galactocentric coordinates using the `astropy` package.

Multiple 2D projections (X vs Y, X vs Z, and Y vs Z) and a 3D visualization were created to inspect structure. The Galactic bulge, disk, and potential halo features were visually identifiable, though projection crowding made distinguishing individual arms difficult. Overlays with Milky Way structure diagrams and density plots helped reveal large-scale structure.

2 Data and Methodology

We compiled a sample of Mira variable stars by cross-matching catalogs from OGLE-IV, ASAS-SN, and ZTF surveys. These surveys provide photometric data, including K-band magnitudes and pulsation periods, necessary for applying the period-luminosity relation.

Using the PLR in the form:

$$M_K = a \log(P) + b \tag{1}$$

where M_K is the absolute K-band magnitude and P is the pulsation period, we calculated distances for each star, and the relation derived is shown in figure 3:

$$d = 10^{(K - M_K + 5)/5} \text{ pc} \tag{2}$$

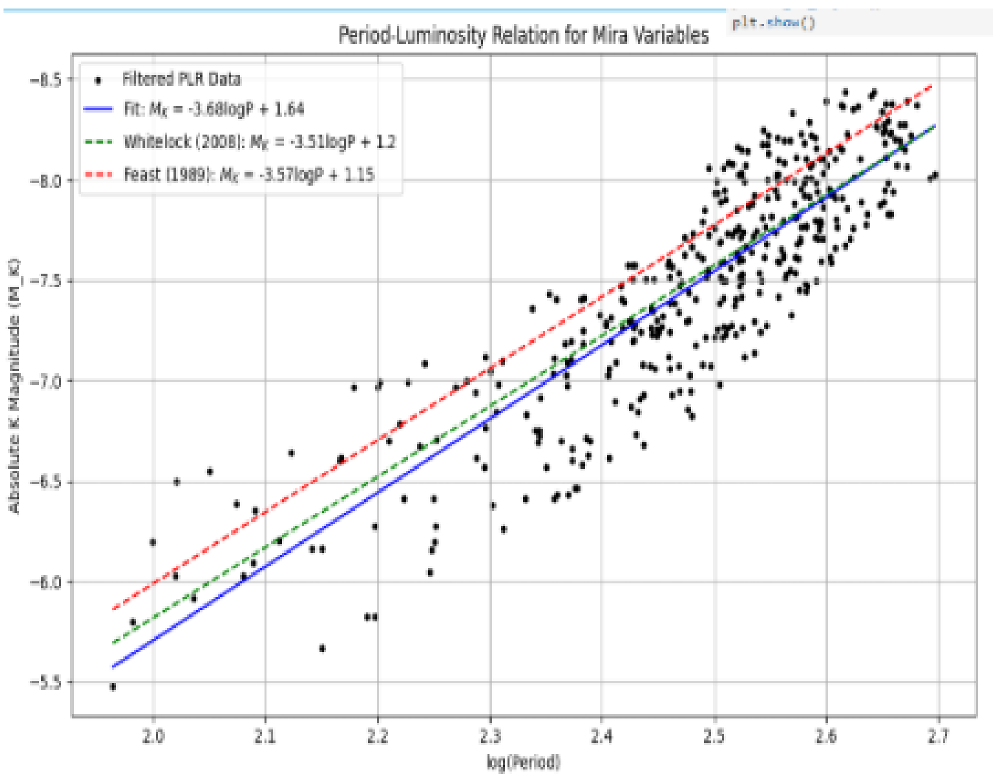

Figure 3: Derived PLR with fit and compared to sources in literature.

Right ascension and declination were converted into Galactocentric Cartesian coordinates (X, Y, Z) using the Astropy library, assuming a solar position at $(-8.122, 0, 0)$ kpc in the Galactocentric frame.

3 Results

The spatial distribution of Mira variables was visualized in two-dimensional and three-dimensional plots. Key views included the X-Y plane (top-down view of the Galactic disk) and the X-Z plane (vertical slice through the disk).

The X-Y projection in figure 4 reveals a concentration of Miras toward the Galactic center, consistent with expectations for evolved disk and bulge populations. In the X-Z projection, figure 4, the vertical spread of stars reflects the disk thickness and the presence of some stars at higher Galactic latitudes, possibly tracing older populations or structures such as the halo or thick disk.

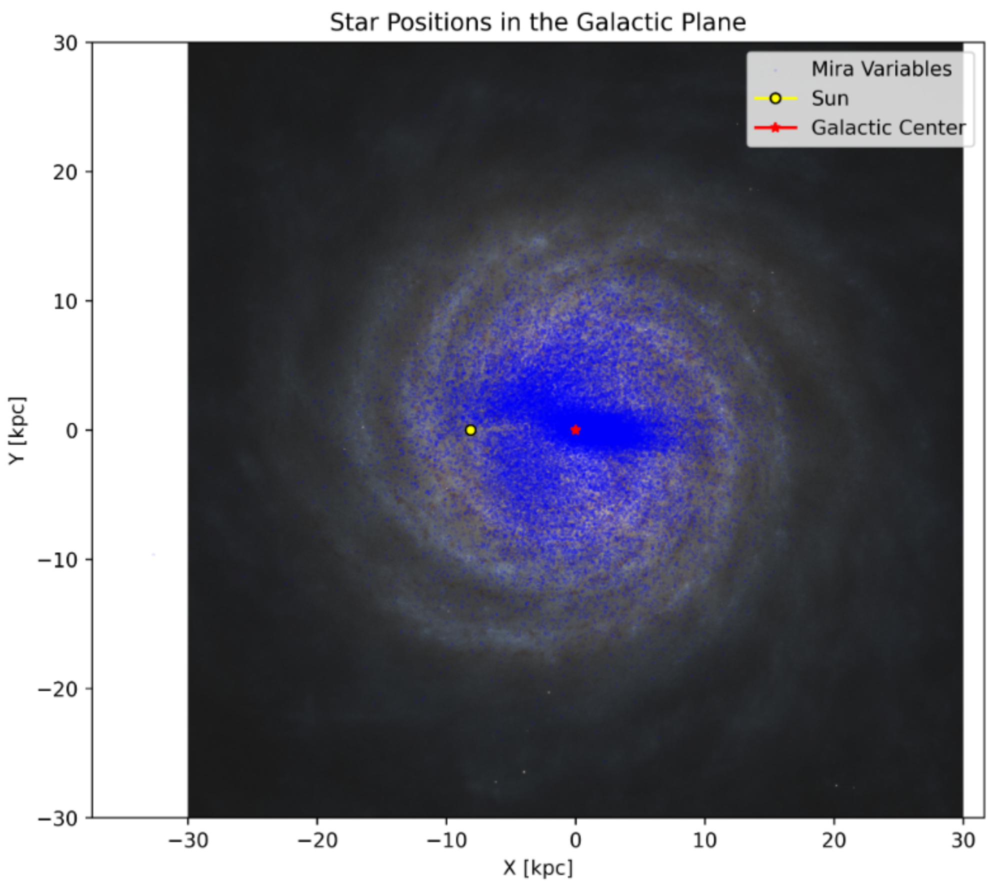

Figure 4: XY view of Miras in the Milky Way with ESO model of Milky Way behind.

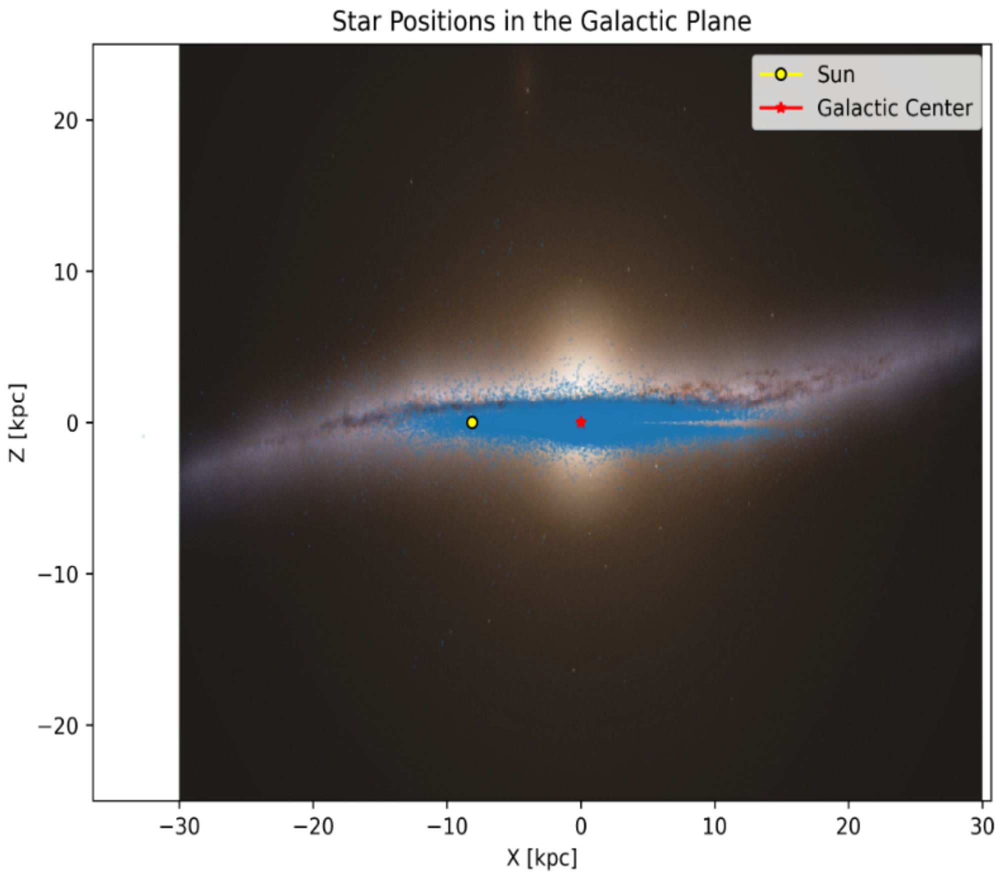

Figure 5: XZ view of Miras in the Milky Way with ESO model of Milky Way behind.

4 Discussion

Each of the three hypotheses addressed the spatial nature of variable stars from increasing levels of depth. The first provided basic familiarity with variable star sky distributions. The second added insight into astrophysical populations by type, and the third achieved a full 3D Galactic mapping using distance estimation techniques.

Limitations included poor positional accuracy in early datasets and sparse Gaia radial velocity coverage, though the Mira subset offered enough data to attempt velocity dispersion and kinematic analysis (not included here). Future work could involve exploring the connection between variable star kinematics and halo structure, and refining Mira selection using additional quality filters.

The high density of stars in the inner regions led to some challenges in visualization, which were addressed through alpha blending and density shading. Distance uncertainties, potential extinction effects, and catalog completeness also play roles in shaping the observed distribution.

5 Conclusion

This study demonstrates how variable stars, particularly Miras, can serve as valuable tracers of Galactic structure. By progressing from broad spatial trends to precise 3D mapping, we highlight the diverse utility of variable star populations. Mira variables, due to their well-defined Period-Luminosity relation and infrared brightness, are especially suited for probing distant Galactic regions. Integrating datasets across multiple sky surveys enables deeper insight than any one catalog alone.