

Unveiling VVV/WISE Mira variables on the far side of the Galactic disk

Distances, kinematics, and a new extinction law

R. Albarracín^{1,2,*}, M. Zoccali^{1,2}, J. Olivares Carvajal^{1,2}, Á. Rojas-Arriagada^{3,2,4,5},
J. H. Minniti⁶, M. Catelan^{1,2}, M. De Leo^{1,2,7}, F. Gran⁸, R. Contreras Ramos^{1,2},
Á. Valenzuela Navarro^{1,2}, and C. Salvo-Guajardo^{1,2}

¹ Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile

² Millennium Institute of Astrophysics, Av. Vicuña Mackenna 4860, 82-0436 Macul, Santiago, Chile

³ Departamento de Física, Universidad de Santiago de Chile, Av. Victor Jara 3659, Santiago, Chile

⁴ Núcleo Milenio ERIS, Santiago, Chile

⁵ Center for Interdisciplinary Research in Astrophysics and Space Exploration (CIRAS), Universidad de Santiago de Chile, Santiago, Chile

⁶ Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

⁷ Dipartimento di Fisica e Astronomia “Augusto Righi”, Alma Mater Studiorum, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy

⁸ Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, 06304 Nice, France

Received 29 August 2024 / Accepted 5 November 2024

ABSTRACT

Context. The structure and kinematics of the Milky Way disk are largely inferred from the solar vicinity. To gain a comprehensive understanding, it is essential to find reliable tracers in less explored regions such as the bulge and the far side of the disk. Mira variables, which are well studied and bright standard candles, offer an excellent opportunity to trace intermediate and old populations in these complex regions.

Aims. We aim to isolate a clean sample of Miras in the VISTA Variables in the Vía Láctea survey using Gaussian process algorithms. This sample will be used to study intermediate and old age populations in the Galactic bulge and far disk.

Methods. Near- and mid-infrared time-series photometry were processed using Gaussian Process algorithms to identify Mira variables and model their light curves. We calibrated selection criteria with a visually inspected sample to create a high-purity sample of Miras, integrating multiband photometry and kinematic data from proper motions.

Results. We present a catalog of 3602 Mira variables. By analyzing photometry, we classify them by O-rich or C-rich surface chemistry and derive selective-to-total extinction ratios of $A_{K_s}/E(J - K_s) = 0.471 \pm 0.01$ and $A_{K_s}/E(H - K_s) = 1.320 \pm 0.020$. Using the Mira period-age relation, we find evidence supporting the inside-out formation of the Milky Way disk. The distribution of proper motions and distances aligns with the Galactic rotation curve and disk kinematics. We extend the rotation curve up to $R_{GC} \sim 17$ kpc and find no strong evidence of the nuclear stellar disk in our Mira sample. This study constitutes the largest catalog of variable stars on the far side of the Galactic disk to date.

Key words. stars: AGB and post-AGB – stars: variables: general – Galaxy: bulge – Galaxy: disk – Galaxy: kinematics and dynamics – Galaxy: structure

1. Introduction

Although we are able to study the Milky Way (MW) galaxy in great detail, de-projecting the position of the stars in the sky into a three-dimensional map is still challenging due to the difficulty of deriving precise distances for individual stars. This task is especially hard for stars in the Galactic plane, as dust extinction, heavily affecting distances and often degenerated with it, is significantly larger. Given that reliable geometric distances are currently available only within a few kiloparsecs of the Sun (Bailer-Jones 2023), we rely on specific stellar tracers, for which we are able to derive distances because we have independent information about their intrinsic luminosity.

Among tracers, pulsating variable stars are the most common ones, as most of them follow tight period-luminosity (P-L) relations (Catelan & Smith 2015). RR Lyrae stars have been used to trace the structure of the old component of the Galactic bulge (e.g., Dékány et al. 2013; Pietrukowicz et al. 2015; Prudil et al. 2019; Du et al. 2020; Olivares Carvajal et al. 2024), and classical Cepheids have proven to trace the structure of the near and far Galactic disk (e.g., Skowron et al. 2019; Minniti et al. 2021). The nature of these classes of variable stars restricts the analysis to strictly old or young stellar populations, respectively. Mira variables, in turn, offer a unique possibility to trace the range of intermediate-age populations in our Galaxy.

Mira variables are highly evolved stars from the thermally pulsating asymptotic giant branch (AGB). AGB stars exhibit pulsation in different modes, with complex interactions between

* Corresponding author; rialbarracin@mpia.de

the pulsation, the convective envelope, and the stellar atmosphere (Freytag et al. 2017; Höfner & Olofsson 2018; Chiavassa et al. 2024). The modes and regularity of the pulsation between cycles define different subclasses of long-period variables (LPVs) occupying different P-L sequences (Wood 2015). Indeed, observations of the Large Magellanic Cloud (LMC) have shown that LPVs lie in distinct sequences depending on their subclass (Soszyński et al. 2009). Mira variables, in particular, pulsate in the fundamental mode, exhibiting amplitudes of $\Delta V > 2.5$ mag with periods ranging from 80 days to over 1000 days, following a single P-L sequence (Glass & Feast 1982). Mira variables have been proven to follow tight P-L relations in the near-infrared (near-IR) and infrared bands (e.g., Feast et al. 1989; Whitelock et al. 2008; Yuan et al. 2018; Iwanek et al. 2021b; Sanders 2023). They are well suited as standard candles for cosmological applications (Huang et al. 2018, 2024), local group studies (Menzies et al. 2015), and the MW (Matsunaga et al. 2009; Grady et al. 2020; Urigo et al. 2020; Iwanek et al. 2023). A recent study by Sanders et al. (2022, hereafter S22) identified a large number of Miras in the MW nuclear stellar disk (NSD) from VISTA Variables in the Vía Láctea (VVV) data. The existence of Miras in this poorly known component is especially important as their age, constrained through their period, allows for the formation of the main Galactic bar to be dated (Baba & Kawata 2020; de Sá-Freitas et al. 2023). In addition, thanks to their precise distances and proper motion (PM), they allow for models of the NSD to be computed (Sanders et al. 2024).

At a given mass, any thermally pulsating AGB star begins to show fundamental-mode pulsations at a very narrow range of radii (Trabucchi et al. 2019). Thus, a dependence of the pulsation period on stellar mass, and therefore age, is expected. Recent theoretical studies of this relation can be found in the literature (see Eggen 1998; Trabucchi & Mowlavi 2022). However, most period-age (P-A) relations have been calibrated empirically. Stars near the Sun with hotter kinematics are, on average, older (De Simone et al. 2004), and indeed, this relation has been long observed for Miras (Merrill 1923; Feast 1963). Using the exquisite astrometric precision of *Gaia* (Gaia Collaboration 2016), Zhang & Sanders (2023) have obtained a P-A relation, using the velocity dispersion of Mira variables in the solar neighborhood. As is discussed in both Nikzat et al. (2022) and Zhang & Sanders (2023), the exact shape of this relation is still highly debated, and some key anchors, such as LMC clusters, can have significant contamination. Although the correlation between period and age is not under discussion, a spread in the relation is expected. Trabucchi & Mowlavi (2022) used a set of theoretical pulsation models to show that a Mira with a period of 350 days can have ages that differ by up to 3 Gyr.

Low and intermediate-mass stars in the AGB also undergo significant mass loss (Höfner & Olofsson 2018). Strong convection and shockwaves due to pulsation generate winds that lead to cool envelopes, where dust and molecule formation are expected (Freytag & Höfner 2023). Changes in the surface ratio of C/O, due to the third dredge-up, affect the dominant molecular species formed in the envelope, from oxygen-based molecules such as TiO and SiO to carbon-based molecules like CN and C₂ (Cristallo et al. 2011). This transition in chemical abundance greatly impacts the spectral energy distribution (Iwanek et al. 2023) and, in turn, the mean magnitudes and the corresponding P-L relation (Whitelock et al. 2003; Soszyński et al. 2005; Yuan et al. 2017; Iwanek et al. 2021b). Therefore, an assessment of the Mira chemical class is crucial for adequate distance determination.

A challenge in the characterization of light curves from Miras is their long-term trends and period variations. Indeed, it has been observed that Mira variables show long-term modulations independent of the main pulsation period (e.g., Zijlstra et al. 2002; Templeton et al. 2005; Uttenthaler et al. 2011; Ou & Ngeow 2022). Multiple mechanisms responsible for these trends have been proposed: stellar winds triggered by thermal pulses (Vassiliadis & Wood 1993), coupling of the pulsation with convection (Freytag et al. 2017), or the presence of dust shells (Zijlstra et al. 1996; Ou & Ngeow 2022). A work frame that accommodates these variations in the modeling is necessary for adequate identification of Miras.

In this paper, we aim to identify Mira variables in the VVV survey by means of a Gaussian process (GP) algorithm to study the 3D distribution of these stars along with the kinematics via PMs. In Section 2, we describe the data considered from different surveys, and Section 3 describes the algorithm used for the identification of Miras and modeling of the light curves. Section 4 explains the separation of C-rich and O-rich stars in our sample and the determination of an extinction law. Section 5 presents the 3D spatial distribution of the stars. Section 6 presents the kinematics of the sample. Finally, in Section 7, we present the discussion and conclusions of our study.

2. Observations and photometry

2.1. VVV data

The catalog of Mira variables presented in this work is based on point spread function (PSF) fitting photometry performed on multi-epoch near-IR data from the VVV ESO public survey (Minniti et al. 2010). Since 2010, the VVV survey has observed nearly 1700 deg² over the MW bulge and southern disk. Further data is available as a part of VVV eXtension (VVVx Minniti et al. 2018), extending the baseline to roughly 13 years. In this study, we use the full footprint of the bulge region, from $-10^\circ \leq \ell \leq +10^\circ$ and $-10^\circ \leq b \leq +5^\circ$, together with tiles from the disk, from $-22^\circ \leq \ell \leq -10^\circ$ and $-1 \leq b \leq 1$, encompassing tiles b201 to b396, d069 to d076, and d107 to d114. The footprint of the catalog in the sky can be seen in Fig. 1. The PSF-based photometry was carried out across all available epochs, including ~ 100 in the K_s band and fewer than 12 in each of the J and H bands. The average seeing is ~ 1 arcsec, as images with seeing larger than 1.7 arcsec were discarded from the analysis.

Multi-epoch photometry was derived using a photometric pipeline based on the DAOPHOT II/ALLSTAR package (Stetson 1987). An initial, high-threshold photometry was performed to detect bright stars and coordinate transformation across all available epochs in each VIRCAM chip. The images were stacked accordingly, and deep, low-threshold photometry was then performed, creating the most complete stellar master list. Using this master list, PSF photometry was performed on each epoch, allowing the centroids for each star to be refined. This resulted in a list of stars with consistent identification across all epochs and a position that can change in time. Instrumental magnitudes were normalized to the photometric system of a reference epoch by means of the DAOMATCH/DAOMASTER codes (Stetson 1987). Only this reference epoch was then photometrically calibrated to the 2MASS reference system, as described below. Proper motions, on the other hand, were derived as described in Contreras Ramos et al. (2017), following the prescriptions from Anderson et al. (2006) and Bellini et al. (2014).

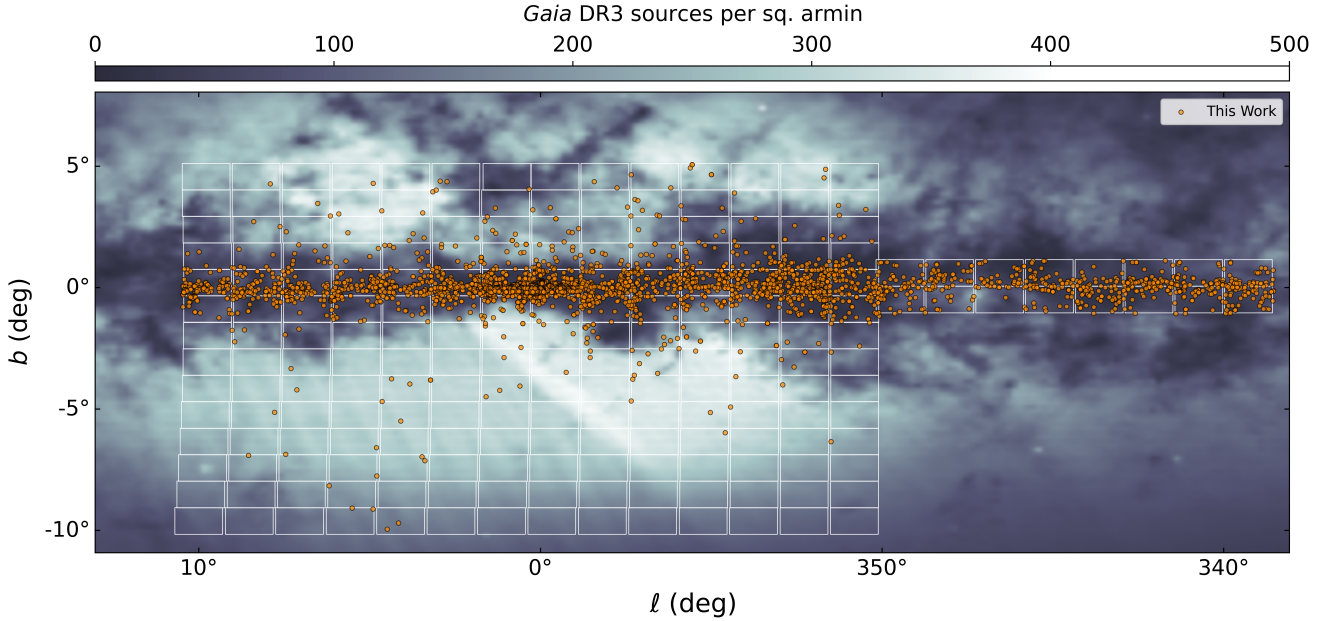


Fig. 1. *Gaia* DR3 source density around the region covered by the VVV survey. White boxes represent VVV tiles considered in this study. Orange dots represent Mira variables identified by our GP algorithm. Color indicate the amount of sources in the *Gaia* DR3 catalog. Density map data were taken from the *Gaia* archive.

The photometry has been calibrated to the VVV photometric system as explained in detail in Zoccali et al. (2024). Briefly: bright, isolated stars in common with the 2MASS catalog (Skrutskie et al. 2006) were identified within each VIRCAM detector. The 2MASS magnitudes were transformed to the VVV photometric system using the color and extinction terms reported in González-Fernández et al. (2018). A zero point was then derived in each band as the median difference between the VVV instrumental magnitudes and the transformed 2MASS magnitudes and applied to the former.

By construction, the PM of each individual star was derived with respect to the average motion of a local sample of bulge red giant branch stars. Therefore, they are relative PMs. Nonetheless, they were later calibrated to the *Gaia* data release 3 (DR3) astrometric system by means of >80 common stars, selected within a quarter of a VVV detector. This is the unit area for which we derive a common astrometric calibration to the *Gaia* system (see Zoccali et al. 2024, for more details about this step).

2.2. WISE data

In addition to the VVV photometry, we complemented the variability information with data from the Wide-field Infrared Survey Explorer (WISE Wright et al. 2010). WISE is a 40 cm space telescope with the mission of mapping the entire sky in four mid-IR bands; W1 ($\lambda_{\text{eff}} = 3.4\mu\text{m}$), W2 ($\lambda_{\text{eff}} = 4.6\mu\text{m}$), W3 ($\lambda_{\text{eff}} = 12\mu\text{m}$) and W4 ($\lambda_{\text{eff}} = 22\mu\text{m}$). The initial WISE mission was conducted from 2010 to 2011, after which the telescope was reactivated as part of the Near-Earth Object WISE Reactivation Mission (NEOWISE-R; Mainzer et al. 2011), providing additional data in the W1 and W2 bands from 2013 to the present. We cross-matched all sources from the VVV catalog with the WISE Multiepoch Photometry Table (WISE Team 2020)¹ and the NEOWISE-R Single Exposure (L1b) Source Table (Neowise Team 2020) using the NASA/IPAC Infrared Science Archive².

¹ <https://wise2.ipac.caltech.edu/docs/release/allwise/>

² <https://irsa.ipac.caltech.edu/applications/Gator/>

Matches were performed with a tolerance of 1 arcsecond, and magnitudes were corrected for saturation effects following the prescriptions in Cutri et al. (2015). Out of our final sample of 3 602 stars, 2782 have measurements in the W1 and W2 bands, 2242 in W3, and 2027 in W4. The median and maximum values for individual data points in each of the WISE bands were 23, 22, 11, 11, and 35, 32, 14, 14 for W1, W2, W3, and W4, respectively. We note that even though the magnitudes are corrected for saturation, several matched sources fall within the saturation range for the WISE bands, specifically 593 in W1, 752 in W2, 382 in W3, and 24 in W4.

Furthermore, the large PSF of WISE is likely to cause blending in highly crowded regions near the Galactic plane. Since Mira variables are very bright in the mid-infrared, we consider the effects of contamination to be minor for the mid-infrared color excess used in Sect. 4.1. However, for precise distance determinations, the contribution of blended sources in the WISE magnitudes affects both the measured median magnitude and the extinction estimates. We expand on this issue in Sect. 5.1 and Appendix C.

2.3. Variable Star Zoo Mira candidates

In 2018, our group started a citizen science project named Variable Star Zoo, which was hosted at the Zooniverse platform³. The users were asked to classify variable stars by comparing their phased light curves with some predefined templates. An input catalog contained roughly 55,000 candidate variable stars from the VVV survey, located within $-10^\circ \lesssim \ell \lesssim +10^\circ$ and $-1.6^\circ \lesssim b \lesssim 1.8^\circ$. The available templates included two eclipsing binaries, an RR-Lyrae and a Cepheid, a Mira, a microlensing event, as well as an unusual variable source and a pure-noise light curve. A candidate variable was considered classified, hence removed from the input catalog as soon as it reached 15 classifications. A more extensive discussion of the Variable Star Zoo Project is presented in Ordenes-Huanca et al. (2022).

³ <https://www.zooniverse.org>

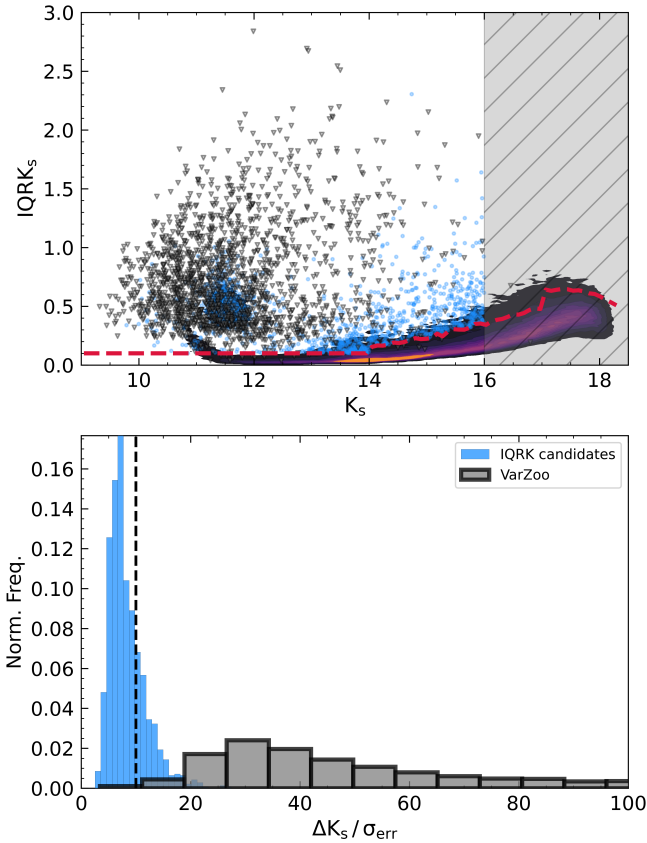


Fig. 2. Amplitude-based cuts in Inter Quartile Range in K_s band and Amplitude over error for an example VVVx tile. *Top:* blue points show candidates that meet IQRKs and amplitude over noise criteria. The dashed red line shows the IQRKs threshold as a function of K_s magnitude; black triangles are the control VarZoo sample. Points inside the gray-shaded region are excluded from further analysis. *Bottom:* histogram of the amplitude of variability over the source-averaged photometric error. Blue histograms are sources that passed the IQRKs cuts, and the gray histogram is the VarZoo sample. The dotted black line defines the threshold value to further consider candidates ($\Delta K_s/\sigma \geq 10$).

For this work, we initially selected the ~ 2000 stars that were classified as Mira by 11 or more volunteers. Subsequent visual inspection carried out independently by four team members reduced the sample to 1883 Mira candidates, which hereafter will be referred to as the VarZoo sample. By using the VarZoo sample as a calibrator for the algorithms explained in Sect. 3, we constructed our final Mira sample, shown in the sky in Fig. 1.

3. Mira variable identification

In order to select candidate Miras, we followed the prescription by S22. We constructed cuts based on an interquartile range of magnitudes in the K_s band (IQRKs; Fig. 2), which we optimized until all the stars in the parameter space occupied by the VarZoo sample were selected, but not many more. We applied different cuts for brighter and fainter stars in order to separate intrinsic variability from statistical fluctuations due to low signal-to-noise ratio (SNR). For stars brighter than $K_s=14$ we considered as candidate Mira any star with $IQRK_s > 0.1$. For sources in the range $14 \leq K_s \leq 16$ we required $IQRK_s \geq 1\sigma$ above the median value. Sources that satisfy these amplitude cuts are shown as blue dots in the top panel of Fig. 2. We neglected sources with $K_s \geq 16$

as their light curves would be dominated by the noise (shaded gray area in Fig. 2, top). We are aware that this first cut does not remove spurious variability due to saturation. However, removing sources in the saturation range would remove many real Mira variables. The top panel of Fig. 2 shows these cuts applied to a VVVx tile, along with the VarZoo sample. We see that the majority of VarZoo stars lie above the imposed threshold. To further clean the candidates from artificial variability due to noise, we impose a cut in the ratio of light curve amplitude to the average photometric noise of $\Delta K_s/\sigma \geq 10$. From the bottom panel of Fig. 2, we see that almost the entire VarZoo sample, represented as a black histogram, lies above the threshold (dotted black line in Fig. 2, bottom). We use these cuts to feed our algorithm with an initial list of candidates showing a strong intrinsic variability based on amplitude only. For further consideration as candidates, a strong periodic signal consistent with Mira variables needs to be detected.

Mira variables are characterized by a prominent main oscillation due to pulsation, with periods over 80 days and a K_s amplitude $\Delta K_s \gtrsim 0.4$ mag (Kerschbaum et al. 2006; Iwanek et al. 2021a). Together with the main pulsation, Mira and other red giants often show additional variability (see He et al. 2016; Matsunaga et al. 2017). The timescale for this additional variability can be of the same order as the main period or up to several times larger (Nicholls et al. 2009). For these reasons, flexible modeling that separates Mira variables from other variability classes and noise while accounting for quasi-periodic variability is necessary.

We performed a fast Lomb-Scargle (Lomb 1976; Scargle 1982; Press & Rybicki 1989) periodogram⁴ on the candidate light curves, retaining sources where any of the best 3 periods, with false alarm probabilities of less than 10%, are over 50 days. To identify LPVs, we further inspect the periodicity of the candidates by means of a Generalized Lomb-Scargle (GLS) periodogram (Zechmeister & Kürster 2009). This method has been reported as having the highest accuracy in recovering Mira periods (Graham et al. 2013b). We cleaned the periodogram from period aliases related to the survey sampling rate by using the window method described in Kramer et al. (2023). In some cases, the periodogram powers for the 3 best GLS periods were approximately the same. In such cases we selected the period with minimal conditional entropy as defined by Graham et al. (2013a).

After rejecting individual data points with errors 3σ above the median error, we fed the K_s band light curves and the selected period to a grid of iterative GP models. Similarly to S22, we used a combination of quasi-periodic and non-periodic kernels in order to capture the behavior of Mira variable light curves, detailed in Eq. (A.1). For each iteration, we selected the model that minimized the Aikake information criterion (AIC). We fitted a simple polynomial to the GP model in order to identify the median trend and subtract it. We phase-folded the de-trended light curve using the period re-adjusted by the GP algorithm. We fitted a Fourier series of three harmonic terms and calculated the distance between each point and the model adequately normalized, as prescribed in Baeza-Villagra et al. (in prep.). Points with a distance over 3σ above the median were rejected and not considered for further analysis. The remaining points were fed to the next iteration of the GP, where a new model fitted. For the model of the next iteration to be accepted, the reported likelihood had to exceed that of the previous iteration (i.e., $\log_{10} \mathcal{L}_i \leq \log_{10} \mathcal{L}_{i+1} + 10$), with a maximum of 5 iterations.

⁴ See docs.astropy.org/en/stable/timeseries/

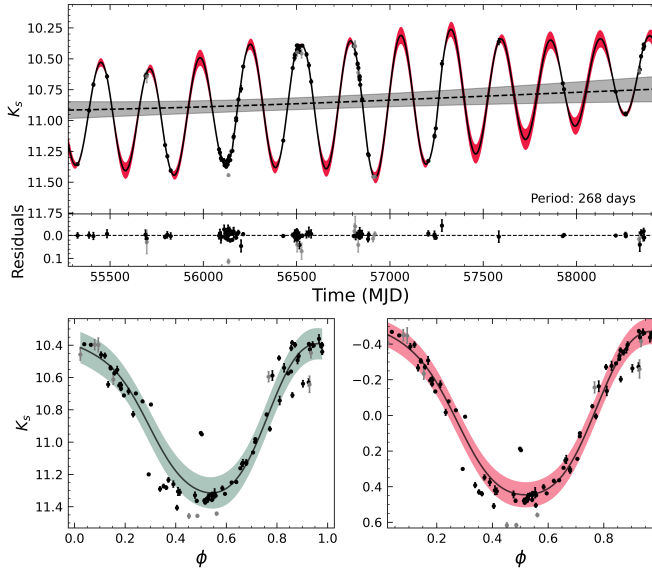


Fig. 3. Example of the GP modeling for a single source. *Top*: time series photometry for this star; red shades denote the confidence intervals for the GP model, shaded black line denotes the median trend with corresponding confidence interval in gray. Gray and black points are masked and accepted points, respectively. *Bottom panels*: phased light curves for the star without (left) and with (right) the median trend correction applied. Colored regions are the confidence intervals for the Fourier fit.

This procedure allows for flexible modeling of the light curve while accounting for stochastic behavior and saturation effects that otherwise would hamper the performance of the \mathcal{L} parameter as a quality indicator for Mira candidates. An example of the described modeling for a VarZoo Mira candidate is shown in Fig. 3.

3.1. Multiband modeling

The method described above only applies to the K_s light curves. In the other bands, the number of epochs is only a few at most. Therefore, in order to accurately derive the median magnitudes for the other photometric bands, we constructed a light curve template based on the K_s data. Specifically, we applied a Fourier transform to the model for the K_s data obtained in the previous step. To apply this model to other bands, we multiplied the Fourier coefficients of the K_s band by the amplitude ratios derived in S22, which describe how the amplitude in a given band relates to the measured amplitude in the K_s band. After applying an inverse Fourier transform to the coefficients, a simple fit on the available data in the given band was performed. In order to derive a median magnitude and its uncertainty, we obtained model fit sampling over the confidence intervals of amplitude ratios and the photometric errors. For each combination of sample points and amplitude ratios, we obtained a best-fit model and a corresponding polynomial fit and repeated this procedure 1000 times. The average of the fitted polynomials is a distribution from which a final median magnitude and uncertainty were obtained. We emphasize that the number of observations per band and their phase have a large impact on the uncertainty of the median magnitude. Although phase shifts are expected in Mira light curves between different bands (Iwanek et al. 2021a), we chose not to include these additional parameters due to the small number of data points in bands different from K_s . This phase shift is close to 0 in the near-IR bands

and $\Delta\phi \sim 0.1$ for WISE filters (Iwanek et al. 2021a). Figure 4 shows an example of the multiband fit for a source in the VarZoo sample.

3.2. Candidate selection

Mira candidates were selected based on the reported likelihood of the fit and the amplitude of the Fourier series. Fig. 5 shows the distribution of the VarZoo sample (cyan) and GP candidates (gray) in the Fourier amplitude, likelihood, and period space. For the VarZoo sample, members of the collaboration visually inspected each light curve and assigned a quality score ranging from 10 (best) to -10 (bad). For a given source, the total score is simply the sum of the individual scores. We bin the distribution of VarZoo stars and show the average total score per bin in the left panel of Fig. 5. Motivated by the distribution of the VarZoo sample and the respective scores, we chose a standard cut in the amplitude of $\Delta K_s \geq 0.4$ and a cut in the likelihood of $\log \mathcal{L} \geq 20$. We recognize that these cuts, especially the likelihood cut, remove variables that might be Miras, but have a low-quality light curve in our photometry. We have chosen to exclude sources with lower likelihood values as we aim to compile a high-purity catalog for variables for which we can derive reliable distances and extinctions.

Mira variables show a clear dependence between the period of the star and the amplitude of the light curve (Soszyński et al. 2011; Huang et al. 2018, S22). Therefore, we adopted a selection box designed to broadly match the distribution of the VarZoo Miras also in the period amplitude plane (Fig. 5; right panel). These cuts resulted in 1420 stars from the VarZoo sample that meet the criteria of light curve morphology and amplitude expected from a Mira variable; these stars will be referred to as VarZoo Miras. We analyzed the distribution of GP candidates (gray points in Fig. 5) and see that the vast majority of processed sources lie outside the borders of the selection box. By applying the same cuts to the output of the GP algorithm, we obtained 12,346 candidates. These candidates include the 1420 stars from the VarZoo sample that meet the aforementioned criteria.

From this list of candidates, we performed a final visual inspection and also removed duplicated sources in the overlapping region between tiles. We find that 3602 sources are clearly Miras, and we can derive reliable distances and extinctions for them. These are the sources contained in the catalog released with the present paper.

3.3. Comparison with other catalogs

The present work is based on the same images from the VVV survey as in S22, analyzed with the same PSF fitting technique. The pipeline to identify Mira variables and measure their mean magnitudes is also very similar. Therefore, a direct comparison of the results is in order. A first important difference is that the work of S22 is restricted to a $3^\circ \times 3^\circ$ region around the Galactic center, while the present work includes the whole VVV area, as shown in Fig. 1. The catalog by S22 includes a total of 1782 Miras, while there are only 939 Miras in our catalog in the common sky region. Specifically, there are 1013 variables in S22 that were not identified as reliable Miras in the present work, whereas 170 variables that we label as Miras are not present in the S22 catalog. As for the first sample, present in S22 but not in the present work, we report a few light curve examples in Fig. 6, all of them folded with the period reported in S22. In a few cases, like the one at the top right corner, it is clear that we are looking at a variable star that is most likely a Mira, too. Nonetheless, this

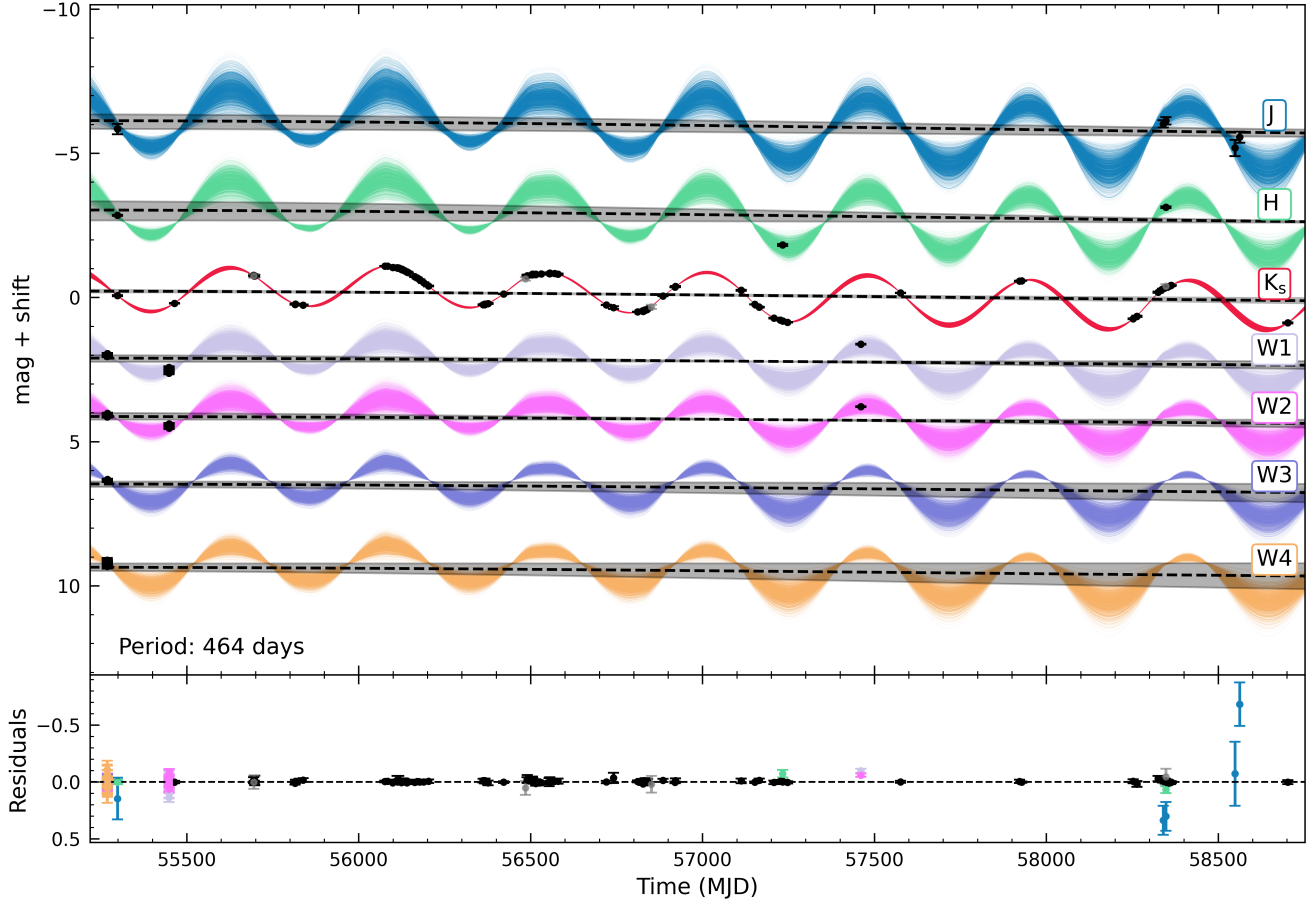


Fig. 4. Example multiband light curve for a source in our catalog. *Top:* fitted models for a candidate in the different bands, ordered in increasing wavelength from top to bottom: J, H, K_s , W1, W2, W3, W4. Black points are photometric measurements in each band. Solid lines indicate each of the sampled models. Dashed gray lines are the modeled median trends with the respective confidence intervals in black-shaded regions. *Bottom:* residuals of the band modeling. Points are color-coded according to the observing band, as in the top panel.

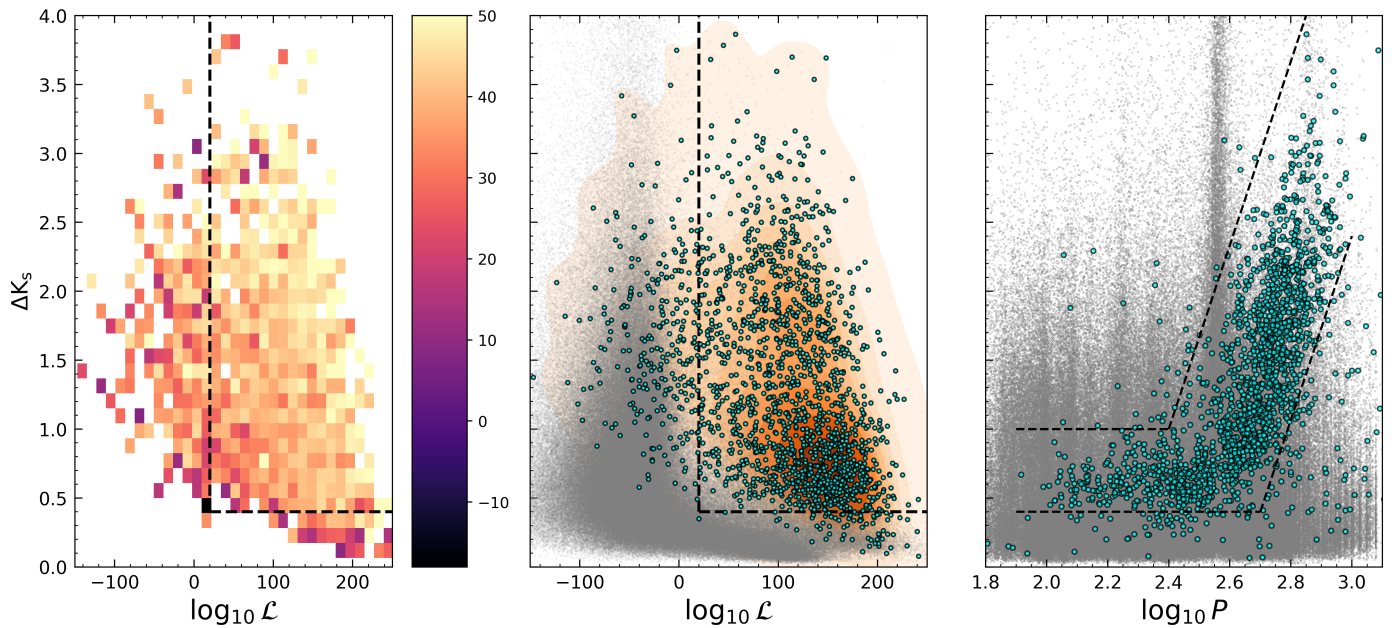


Fig. 5. Distribution of VarZoo and GP candidates in likelihood, amplitude, and period space. *Left:* likelihood and amplitude distribution of the VarZoo sample; the color of the bins denote the average vote score. *Middle:* distribution of likelihood and amplitude of the VarZoo sample and GP candidates; gray points indicate GP candidates, cyan points are the Varzoo sample, with the Orange contours indicating the density of VarZoo sample in the presented space. *Right:* $\log P$ and amplitude distribution; dotted black lines indicate the selection borders for Mira variables.

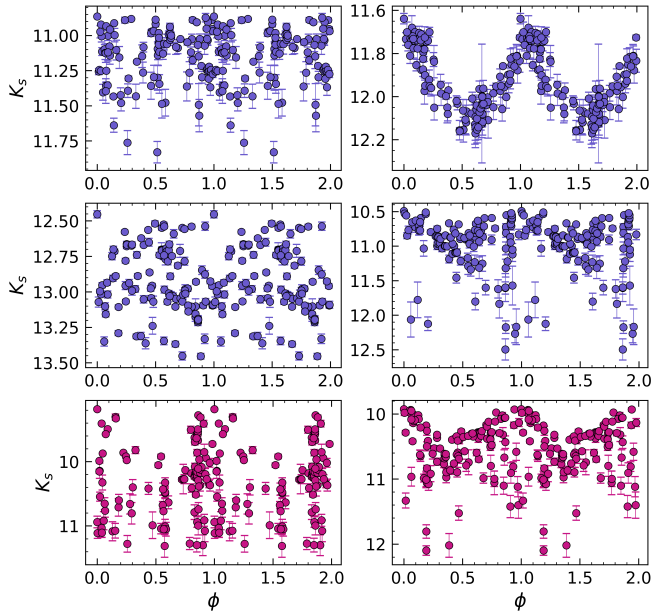


Fig. 6. Example light curves for stars in other catalogs. *Top and Middle panels:* light curves of stars identified as Miras in S22 but not included in the present catalog. *Bottom panels:* same as above, but for sources in the catalog by Matsunaga et al. (2009), not included in the present sample. All the light curves were folded using the period reported by the corresponding authors.

one was not fed to our algorithm because we only considered light curves whose amplitude was larger than 10 times the mean error on the data points. The vast majority of the stars in S22 not included in this work have light curves like those in the middle and top left panels of Fig. 6. In these cases, we might say that the variability is real, but we were unable to assign a reliable period and mean magnitude to them.

Two considerations must be kept in mind while performing this comparison. First, for the Miras in common between the two catalogs, the derived periods are virtually identical for all except ~ 20 stars. We phase the stars with both periods, and found that only in two cases, the period selected by our algorithm produces a folding worse than that reported by S22. These stars were removed from further consideration. Second, stars with $K_s \sim 11$ mag, or brighter, are saturated in VVV. The PSF photometry performed by ALLSTAR does a good job of recovering a reasonable magnitude (see discussion in Appendix B, and Fig. 2 in S22). Nonetheless, we believe that a different prescription for dealing with saturated stars is most likely the reason why stars like those shown in the middle and top left panels might have light curves compatible with Miras in the S22 photometry but not in the present one. We would also like to emphasize the importance of visual inspection of the final sample. For example, the star whose light curve is shown in the middle right panel was included in the initial catalog, and the GP derived reasonable parameters for it. It was later discarded by visual inspection. This last step is often neglected in a world where machine learning algorithms are faster and more effective by the day. While we acknowledge that it is impossible to visually classify hundreds of thousands of light curves, we wish to stress that, although subjective, visual classification is still the most reliable method we have. When the numbers allow, this step should not be skipped.

The bottom panels of Fig. 6 show a couple of light curves for stars in the catalog by Matsunaga et al. (2009) that have not been classified as Mira in the present work. There are 85 such

variables. The large majority of these stars are very bright, and the effect of saturation is stronger. It is, therefore, not surprising that we miss most of them.

Both S22 and Matsunaga et al. (2009) argue that most of the Miras in their catalogs are located very close to the Galactic center, specifically in the NSD. Unfortunately, we cannot perform a one-to-one comparison of the distances derived in the present work with those of S22, as the latter are not reported in their catalog. Still, we can compare the distance distribution derived in the two works. In the present work, we tested the P-L relation derived by Sanders (2023) from a sample of nearby Miras with *Gaia* DR3 parallaxes, converted to the VVV system as reported in Table 1 of S22. We also tested the P-L relation, from the same table, derived by Sanders (2023) for Miras in the LMC, assuming a distance modulus to the LMC of $\mu_{\text{LMC}} = 18.477 \pm 0.026$ (Pietrzyński et al. 2019). Although the two relations are very similar, the extinction law derived further below is quite sensitive to the value of the P-L coefficients, and that, of course, has a non-negligible impact on the final distance distribution. The main reason for our preference for the LMC-based P-L is the derived rotation curve, as is discussed in Sect. 6 below. In addition, previous works have shown that the derivation of precise distances for nearby Miras is very challenging, even with *Gaia* DR3 parallaxes (see discussion in Sect. 5.1 of Huang et al. 2024).

The orange and green histograms in Fig. 7 show the distance distribution of the whole sample of Miras in the present work, derived with the MW-based P-L by Sanders (2023, orange), and with LMC-based P-L by the same authors, adopted here (green). The comparison shows that, had we used the MW-based P-L, we would not have found a significantly different distance distribution, especially close to the Galactic center. If anything, we would have assigned larger distances to each of our variables. Having said that, the purple histogram shows the distance distribution of the Miras in common with S22. While the distribution shows a peak at the distance of the Galactic center, we cannot confirm that most of these variables are located in the NSD due to its relatively small size and the typical distance errors. Of course, we cannot say anything about the location of the 1013 stars not included in our catalog. The very few stars that we have in common with Matsunaga et al. (2009) are also shown in Fig. 7, and they are mostly close to the Galactic center, but they are too few to allow us to draw any conclusion.

We further contrasted our catalog with Mira catalogs based on data from the Optical Gravitational Lensing Experiment (OGLE; Udalski et al. 2015). We consider the catalog from Iwanek et al. (2022) of OGLE Miras in the Galactic bulge and disk and the VVV/OGLE catalog from Nikzat et al. (2022), finding 59 and 56 stars in common, respectively. As is noted by Nikzat et al. (2022), Miras in the OGLE catalog are likely to be heavily saturated in VVV, impacting the quality of the light curve and thus hampering the ability of the GP algorithm to identify the sources.

Another important point of comparison is in the distribution of periods of the samples. We analyze the Mira period distribution of our catalog and several catalogs from the literature in the bottom panel of Fig. 7. We observe an overall uniform distribution with two small peaks at $P \sim 200$ days and $P \sim 320$ days for the whole sample. As was recently noted by Suresh et al. (2024), the Mira period distribution from catalogs in the literature exhibits two peaks at ~ 200 days and ~ 350 , shown as dashed grey lines in Fig. 7. Even though our sample does not show a strong bimodality, the two small peaks in our period distribution are broadly consistent with the expected location of the peaks from other

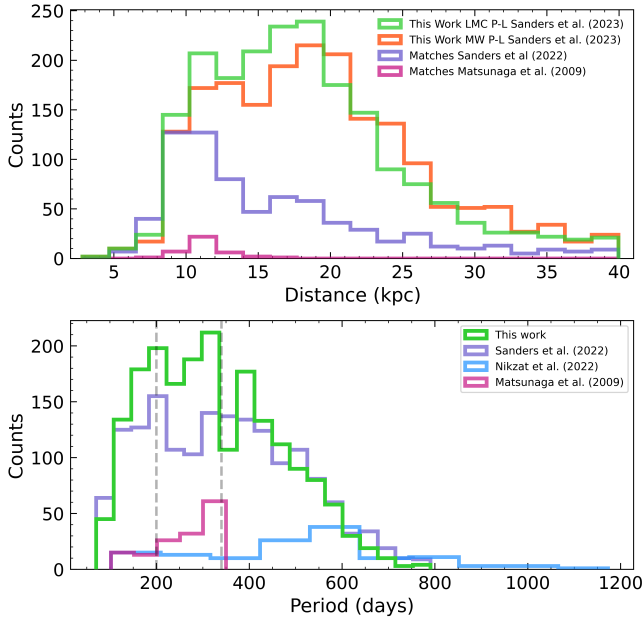


Fig. 7. Comparison of different P-L relations and catalogs. *Top*: distance distribution of our sample using the P-L relations from Sanders (2023) using MW Miras (orange) and LMC Miras (green). We also show the distance distribution of matching Miras from Matsunaga et al. (2009, pink) and S22 (purple), also using the P-L relation using LMC Miras from Sanders (2023) adopted in this study. *Bottom*: period distribution of different catalogs. We show our final sample (green), the S22 catalog (purple), Matsunaga et al. (2009) catalog (pink), and the OGLE/VVV Mira catalog from Nikzat et al. (2022). We added two tentative peaks for the period distribution as vertical gray lines.

studies, such as Matsunaga et al. (2009), Nikzat et al. (2022) and Suresh et al. (2024). However, we also note that the 350-day peak is greatly affected by the adopted alias removal procedure, resulting in a gap at $P \sim 365$ days, which can explain the difference between the expected and observed peaks.

4. Sample properties

4.1. Surface chemistry

As was previously stated, AGB stars are typically categorized as C-rich or O-rich depending on the abundance ratio [C/O]. C-rich stars with $[C/O] > 1$ are known as C stars, while O-rich stars with $[C/O] < 1$ are named M stars, and intermediate stars with $[C/O] \sim 1$ are labeled S stars (Höfner & Olofsson 2018). The third dredge-up process, together with the metallicity and mass of the star, determines the resulting chemical composition (Gail et al. 2009). Typically, C-rich stars are associated with metal-poor or younger systems such as the LMC (Soszyński et al. 2009). O-rich stars, instead, are linked to metal-rich older populations like the MW bulge, although noticeably Matsunaga et al. (2017) found 5 C-rich Miras in this region. Therefore, the stellar type based on the [C/O] ratio conveys important information about the stellar content and is crucial for determining precise distances.

Many methods have been employed in the literature to distinguish between M and C stars using photometric data only. M and C type stars have been observed to exhibit different color indexes in the near-IR (Feast et al. 1989; Soszyński et al. 2009; Lebzelter et al. 2018) owing to their different compositions of circumstellar dust (Ita & Matsunaga 2011; Höfner et al. 2016; Matsunaga et al. 2017; Höfner et al. 2022, S22) and spectral

features at several wavelengths due to C and O molecular bands (Sanders & Matsunaga 2023). For Miras in the LMC, the distinction between C-rich and O-rich chemistry has been achieved using optical and near-IR Wesenheit indexes (Soszyński et al. 2005, 2009; Cioni 2010; Lebzelter et al. 2018), and color index combinations (Matsunaga et al. 2017; Suh & Hong 2017; Suh 2018). Furthermore, Iwanek et al. (2021b) demonstrated that this distinction could also be made using only the period and a mid-IR color index. This surface chemistry distinction is not exclusive to color indexes: Cioni et al. (2003) found amplitudes of C stars to be, on average, higher than those of M stars. By using the same LMC sample of M and C type stars as Yuan et al. (2017), Huang et al. (2018) showed that these populations can be distinct on the basis of their periods and amplitudes in the H near-IR band. Despite relatively vast literature discussing how to untangle these AGB stars, the implementation is not straightforward to apply to our sample, as large star-by-star distance spread directly affects the colors due to extinction, blending the two populations in simple color-period diagrams.

Distinguishing between O-rich and C-rich surface composition is crucial for adequate source-by-source reddening determination, and distances. To achieve this goal using the available photometric filters, we first explored the differences in their light curve properties. We built a preliminary sample of O-rich and C-rich Miras based on the Period versus K_s band amplitude measured by the Fourier fit. We restricted this analysis to VarZoo Miras only, as they have the best light curves in the sample. Motivated by the distribution of LMC Miras identified and classified as C/O rich by Soszyński et al. (2009) and then studied in near-IR by Yuan et al. (2017), we label as O-rich the stars with $\log_{10}(P) \leq 2.4$ and amplitudes $0.4 \leq \Delta K_s \leq 0.8$, and C-rich the stars with $\log_{10}(P) \geq 2.4$ and amplitudes $\Delta K_s \geq 1.5$ (Fig. 8, bottom). Stars in the middle region can be of both types and therefore, we did not classify them into either type at this stage. By comparing to the Yuan et al. (2017) sample, shown in the top panel of Fig. 8, we estimate the contamination of the O-rich sample to be $< 10\%$, and the contamination of the C-rich sample to be negligible.

In order to see how these samples separate in the two-color planes, in Fig. 9, we plot the distribution of the preliminary subsamples of C-rich and O-rich Miras using the median magnitudes derived from the models in Sect. 3.1. We obtain the absolute magnitudes per band using the LMC O-rich P-L relation from Sanders (2023), for the VVV filters, and the P-L relation from Iwanek et al. (2021b), for the WISE bands. Comparing the distribution of our sample with the extinction law from Wang & Chen (2019, black arrows), we see that although our Miras show a non-negligible spread along the reddening vector, the separation between the C-rich and O-rich samples is still visible. In order to understand this separation, we perform linear fits in the same color excess planes to the LMC Mira sample from Soszyński et al. (2009), together with the near-IR photometry from Yuan et al. (2017) and WISE/Spitzer data by Iwanek et al. (2021a). As is noted by Yuan et al. (2018), C-rich Miras from the LMC occupy a distinct sequence with a slope (magenta arrows) different from the Galactic reddening vector, hence allowing us to disentangle the reddening for these sources. To separate the samples, we use a linear cut with the slope of the Wang & Chen (2019) extinction law and a zero point that maximizes the **F1** score, defined as:

$$\mathbf{F1} = \frac{2 \cdot \mathbf{TP}}{2 \cdot \mathbf{TP} + \mathbf{FP} \cdot \mathbf{FN}}, \quad (1)$$

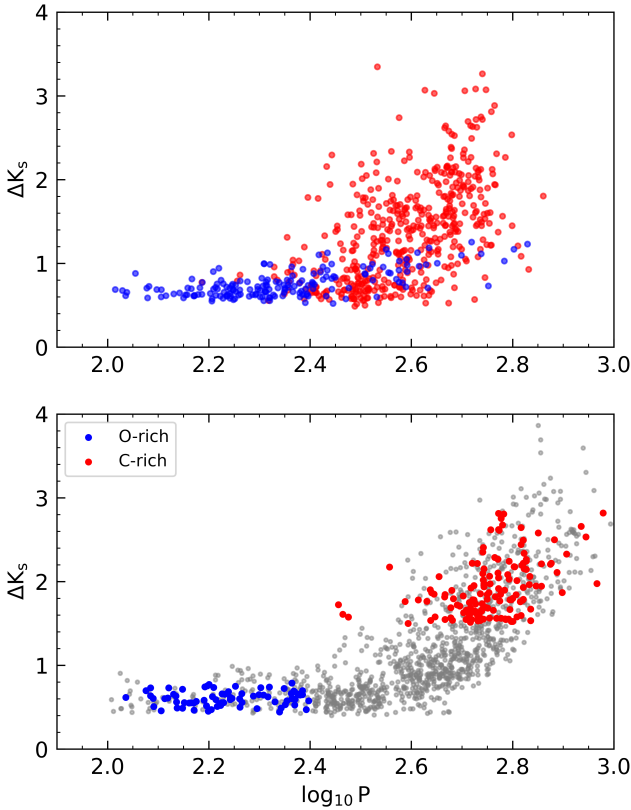


Fig. 8. Period-amplitude distribution of VarZoo and LMC Miras. Red points indicate C-rich Miras, and blue points indicate O-rich Miras for both samples. *Top:* LMC Miras from Yuan et al. (2017) classified using Weseinheit indexes. *Bottom:* Miras from the VarZoo sample with the assigned classifications based on period and amplitude ranges.

TP are the True Positives (O-rich Miras classified as such), **FP** are the False Positives (C-rich Miras labeled as O-rich), and **FN** are the False Negatives (O-rich Miras labeled as C-rich). The **F1** score adopts values from 0 to 1, indicating a complete misclassification and a perfect classification, respectively. This cut, using WISE filters and the extinction law from Wang & Chen (2019), allowed us to isolate stars with clear excess in color that separate from the O-rich region without a previous assumption of an extinction law in the VVV bands. As shown in Fig. 9, the separation is maximal in the top left panel. Nonetheless, we applied the same process also to a combination of only VVV filters, because not all the stars in the sample were identified in the WISE photometry. We repeated the procedure using the extinction law derived in Sect. 4.2; the results for the zero points obtained for the different color planes are detailed in Table 1. O-rich stars tend to follow narrower sequences in the color planes, aligned with the interstellar extinction vector due to their optically thin envelopes, whereas C-rich stars extend to redder colors with a larger scatter. This effect is more notorious for WISE filters, as flux contributions from the circumstellar shell become more important in the mid-IR for C-rich Miras, whose spectral energy distribution can be described by a double Planck law (Iwanek et al. 2021a). To classify Mira candidates by surface chemistry in our catalog, we used the above color planes, favoring the use of WISE data where available. We prioritized color planes by the angular difference between the circumstellar extinction vector (magenta arrows in Fig. 9) and the interstellar extinction vector (black arrow in Fig. 9). This order is also showcased in Table 1. It is important to note that our classification

Table 1. Separating lines for O-rich and C-rich Miras in the Color Excess planes considered.

CE X	CE Y	Slope	Zero-point	F1
$E(J - W2)$	$E(K_s - W1)$	0.179	0.770	0.953
$E(H - W1)$	$E(K_s - W2)$	0.565	0.770	0.952
$E(J - H)$	$E(H - K_s)$	0.554	0.122	0.830
$E(H - K_s)$	$E(J - K_s)$	0.356	0.078	0.830

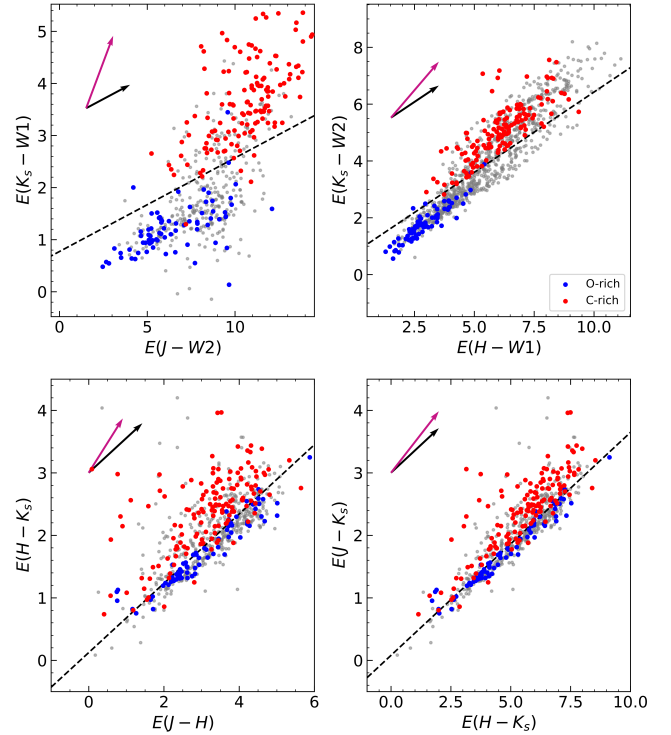


Fig. 9. Distribution of C-rich and O-rich VarZoo Miras in excess color planes detailed in Table 1. Red points show C-rich Miras, Blue points indicate O-rich Miras, and gray points show the rest of the VarZoo sample. Black arrows indicate the extinction law from Wang & Chen (2019) in the top panels and this study in the bottom panels. Magenta arrows show the fit for C-rich Miras from Yuan et al. (2017) in the same color planes. Dotted black lines show the derived separations between C-rich and O-rich Miras.

method requires measurements in at least three or four bands, depending on the color plane used. However, due to the high extinction affecting the J and H bands and the blending issues with WISE photometry near the Galactic plane, 512 sources in our catalog lack the necessary filter combinations. As a result, we could not provide a classification for these sources.

We show the distribution of C-rich and O-rich for stars of our Mira sample in a color-magnitude diagram (CMD) in Fig. 10. We clearly see that C-rich Miras have, on average, redder colors, owing to their optically thick dust envelopes.

4.2. Extinction law

In order to derive distances and reddening for each source, again, we obtain the absolute magnitudes in the VVV filters using the near-IR P-L relation by Sanders (2023) using LMC Miras in the VVV photometric system, detailed in Sect. 5. For distance purposes, we focus on the K_s band as it has the highest

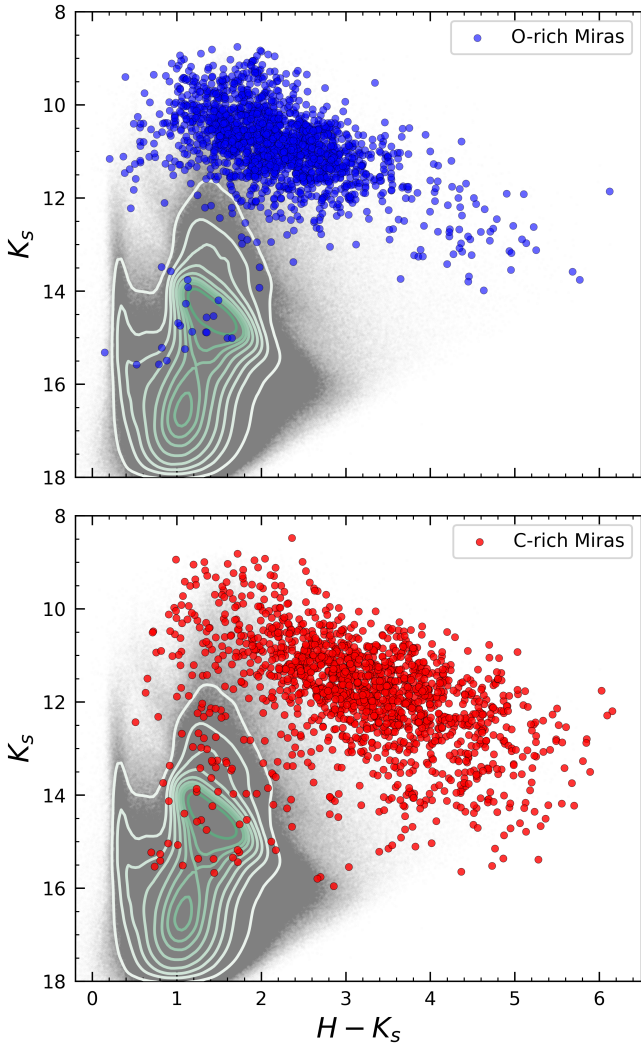


Fig. 10. CMD using H and K_s of Miras in our sample. *Top:* CMD of O-rich Miras. *Bottom:* CMD of C-rich Miras. Both panels include a sample of stars from VVV as gray points, with the corresponding density contours in green. In this CMD, the extension of C-rich Miras to redder colors is clear.

sampling and a lower typical error. To obtain the total extinction in the K_s band (A_{K_s}) from the observed color excess, we need an accurate value of the total-to-selective extinction ratios in a filter combination. As is discussed in Matsunaga et al. (2016) in the case of classical Cepheids, and in Nikzat et al. (2022) in the case of Miras, differences in the adopted extinction law directly impact the determined distances, especially in the highly obscured regions close to the Galactic plane. The values of the selective-to-total extinction ratios $A_{K_s}/E(H - K_s)$ in the literature span a wide range of values, from 1.10 (Alonso-García et al. 2017) to 1.61 (Nishiyama et al. 2009). In the case of $A_{K_s}/E(J - K_s)$, the range is smaller, from 0.428 to 0.528, drawn from the same two references. Using classical Cepheids at the disk far side, Minniti et al. (2020) determined a value of 1.308 and 0.465, for $A_{K_s}/E(H - K_s)$ and $A_{K_s}/E(J - K_s)$, respectively. Due to this non-negligible spread, it is important to determine an extinction law suitable for our sources.

To determine the extinction law for our sample of Miras, we used O-rich Miras from the VarZoo sample, cleaned from C-rich contaminants, using the WISE color excess planes and the Wang & Chen (2019) extinction law (Fig. 8 top), as detailed

in Sect. 4.1. We also imposed additional cuts in likelihood $\log_{10} \mathcal{L} \geq 50$ to only consider the highest quality stars. As mentioned, Miras can show trends in their average magnitudes over several cycles (Uttenhaler et al. 2011). From the modeling described in Sect. 3, we defined the parameter MinMax K_s to describe the absolute difference in magnitude of the polynomial fitting the median trend of the light curve. For this sub-sample, we impose $\text{MinMax}K_s \leq 0.05$ mag to include only light curves that have a stable mean brightness over the time baseline of the survey. We performed a linear fit to the distribution of O-rich Miras in the color excess planes, imposing a fixed intercept of 0 and a 3σ clipping to exclude outliers. From this, we obtained the following mean color excess ratios:

$$\frac{E(J - K_s)}{E(H - K_s)} = 2.801 \pm 0.016, \quad (2)$$

$$\frac{E(J - H)}{E(J - K_s)} = 0.6438 \pm 0.0021. \quad (3)$$

Assuming that the extinction law in the near-IR bands follows a power law $A_\lambda \propto \lambda^{-\alpha}$, we can get the power law index from the mean color excess ratios. For the VISTA bands, the effective wavelengths (λ_{eff}) of each filter are 1.254 μm (J), 1.646 μm (H), and 2.149 μm (K_s). From Eqs. (2) and (3) we obtained $\alpha = 2.11 \pm 0.02$. The derived selective-to-total extinction ratios are:

$$\frac{A_{K_s}}{E(J - K_s)} = 0.471 \pm 0.010, \quad (4)$$

$$\frac{A_{K_s}}{E(H - K_s)} = 1.320 \pm 0.020. \quad (5)$$

From these values and the measured color excess, we can calculate A_{K_s} for our Mira sample.

5. Three-dimensional spatial distribution

5.1. Distance determination

To obtain distances to individual stars, we used the average K_s magnitude derived in Sect. 3 and Sect. 3.1, together with the extinction law derived in Sect. 4.2. The distances presented here are based on the H and K_s band magnitudes and selective-to-total extinction ratios. We chose this combination of filters because these bands have lower typical errors in average magnitude, and the H and K_s based extinction ratios since $\sim 38\%$ of the sample do not have J band measurements. We calculate the distance modulus using the expression

$$d = 10^{1+0.2(K_s - A_{K_s} - M_{K_s})}, \quad (6)$$

where K_s is the intensity-based mean magnitude in the K_s band derived from the modeling described in Sect. 3, A_{K_s} is the extinction in the K_s band, and M_{K_s} is the absolute magnitude obtained from the P-L relation. As was mentioned in Sect. 3.3, we use the P-L relation from Sanders (2023) using LMC Miras. The adopted P-L relations were transformed to the VVV photometric system using the coefficients from González-Fernández et al. (2018), obtaining

$$\begin{aligned} M_J &= -6.30 - 3.32 (\log_{10} P - 2.3) - 7.29 \mathcal{H}_{2.6} (\log_{10} P - 2.3)^2, \\ M_H &= -6.67 - 3.33 (\log_{10} P - 2.3) - 6.85 \mathcal{H}_{2.6} (\log_{10} P - 2.3)^2, \\ M_{K_s} &= -7.01 - 3.73 (\log_{10} P - 2.3) - 6.99 \mathcal{H}_{2.6} (\log_{10} P - 2.3)^2, \end{aligned} \quad (7)$$

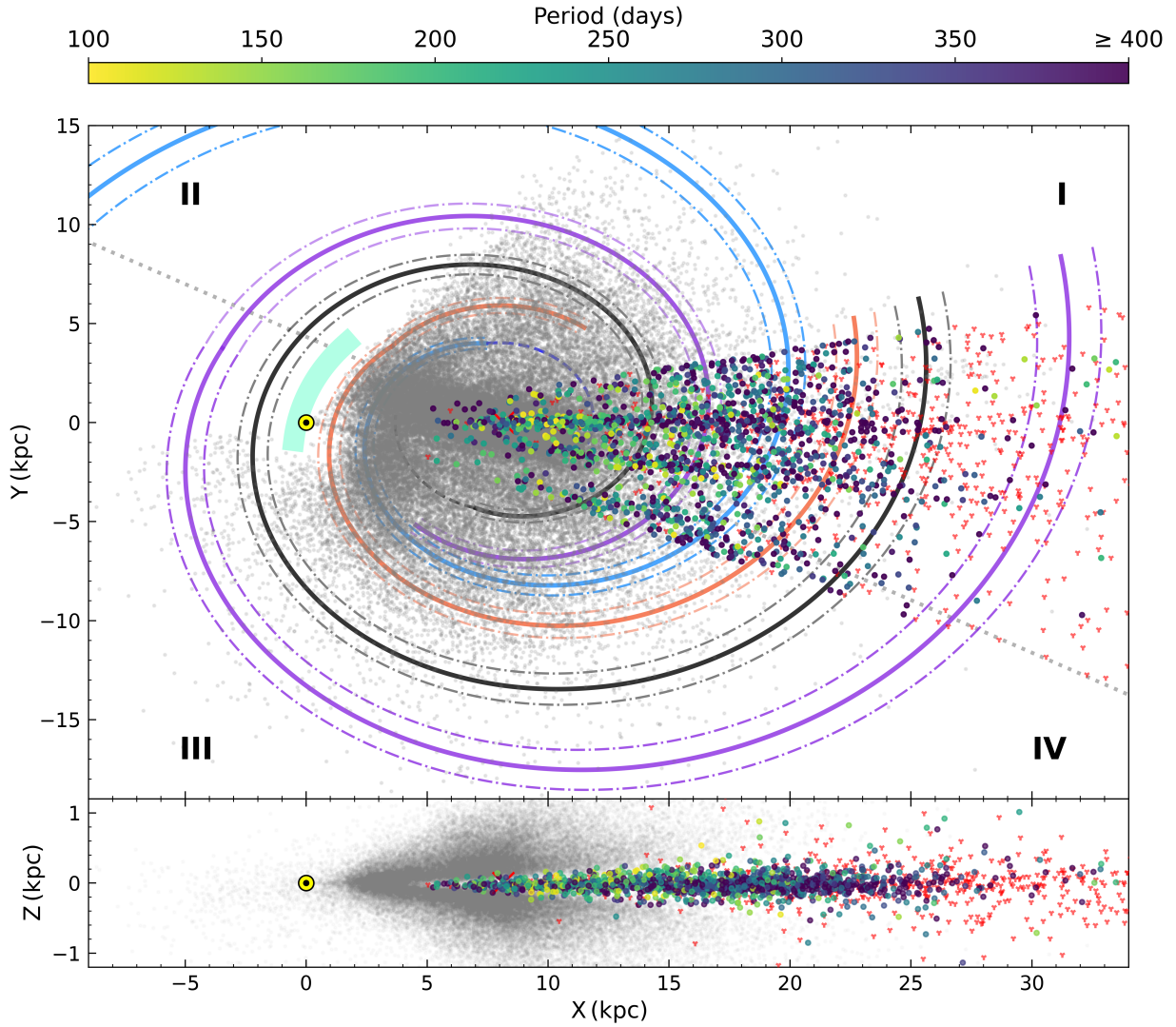


Fig. 11. Three-dimensional distributions of Miras in the MW. The position of O-rich stars in our sample is shown in circles color-coded by period. C-rich stars are included as red ticks. We over-plot the Miras from OGLE project presented in Iwanek et al. (2022) with the distances reported in Iwanek et al. (2023) as gray dots. The Sun is located at $(X,Y)=(0,0)$, and marked with a yellow dot, and the Galactic center is marked with a red cross at $(X,Y)=(8.34,0)$. Log-periodic spiral arm segment fit of the local arm based on VLBI trigonometric parallaxes of high-mass star-forming regions by Reid et al. (2019) is represented with a turquoise-colored section. The thick lines represent qualitative logarithmic spiral arm fits from Minniti et al. (2021), with the thin dashed lines marking the 1σ arm widths reported in Reid et al. (2019) (Sct-Cen arm, blue; Sgr-Car arm, orange; Perseus arm, black; Outer arm, purple). The line of nodes of the Galactic warp model by Skowron et al. (2019) is also shown for reference (dotted gray line). The bottom panel shows the XZ view of the Mira distribution in the MW.

here, $\mathcal{H}_{2.6}$ indicates a simple Heaviside function, which takes a value of 1 for $\log P \geq 2.6$ and 0 otherwise. The statistical uncertainty of the distance can be derived by applying the propagation of the errors to Eq. (6):

$$(\Delta d)^2 = (0.46d)^2(\delta K_s^2 + \delta A_{K_s}^2 + \delta M_{K_s}^2). \quad (8)$$

Here, δK_s is the mean magnitude error in the K_s -band, δA_{K_s} is the extinction error, and δM_{K_s} is the absolute magnitude error. The value of δK_s was obtained from the sampling of the median model derived in Sect. 3.1, and δM_{K_s} was derived from the error propagation of the P-L relation and the reported coefficients from Sanders (2023). We note that the term δA_{K_s} is the largest source of errors due to the large color excess in the sample and the larger uncertainties of the extinction law. We note that, as P-L relations, extinction law, and modeled median magnitudes in the WISE bands have been used in this study, it is also possible

to use WISE filters for distance determination. In Appendix C, we determined distances using only WISE filters by following the same procedure detailed in the current section and compared these distances with our adopted values. We found that the WISE-based distances are in good agreement with the adopted distances, but exhibit significantly larger errors. We attribute this increased uncertainty to the larger PSF of WISE, causing blending effects near the Galactic plane, as well as uncertainties in the extinction law. Therefore, we favor the use of K_s -band-based distances in the following sections.

In Appendix D, we present the header of the final catalog, available at the CDS.

5.2. 3D distribution

The distribution of our Miras, projected to the plane of the Galaxy, is shown in Fig. 11. O-rich Miras are shown as dots,

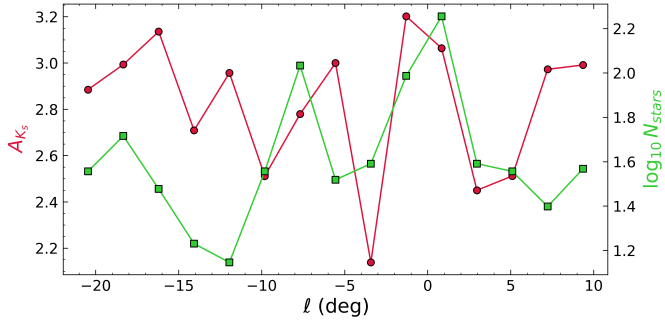


Fig. 12. Correlation between stellar counts and extinction. We make equally spaced bins using the Galactic longitude and show the average extinction in the K_s band (red) and the logarithm of the number of Miras per bin (light green).

color-coded by their respective period. We note that Miras with periods longer than ~ 400 days are often neglected from distance analysis due to the onset of Hot Bottom Burning (HBB) at the corresponding mass range, as well as the increasing mass loss rates expected for more massive stars (Uttenhaler 2015; Uttenhaler et al. 2019). These processes increase the scatter from the P-L relation at higher periods, thus making distances less reliable. Here, we report distances to these sources, but they are not included in the analysis. We choose to also include C-rich Miras in the distribution (red ticks in Fig. 11), although their distances are likely over-estimated due to additional reddening from the circumstellar shell.

A clear feature in Fig. 11 is the presence of sequences along lines of sight with higher stellar density. To explain the origin of these features, we plot in Fig. 12 the average extinction in the K_s band (red) and the number of sources (green) in longitude bins. We see a clear correlation between the two curves, showing that the number of detected Miras is larger for lines of sight-piercing regions with higher average extinction. We interpret this as due to low-extinction Miras being saturated in VVV.

Even though saturation effects are expected for bright sources at $K_s \leq 12$ (Contreras Ramos et al. 2017), Sanders et al. (2022) showed that no significant bias in the saturation regime was observed for Miras in the VIRAC2 catalog (Smith et al. 2018). Given that both the VIRAC2 catalog and the catalog from Contreras Ramos et al. (2017) are based on the same VVV images using PSF-based methods, we consider that no significant biases in the magnitudes are present for our selected sources, and thus, the provided distances are reliable. We expand on this point in the Appendix B.

In order to explore the correlations between the position of the Miras in the Galaxy and their age, we use the P-A relation of Zhang & Sanders (2023) and a Sun-Galactic center distance of $R=8.34$ kpc (Reid et al. 2014). To get precise distances, we limit the analysis to O-rich stars with errors in distance of less than 10% and periods of less than 400 days, resulting in 972 total sources. We bin the data in four age bins, as is shown in Fig. 13. To understand how the distance distribution changes with age, we run a grid of Gaussian mixture modeling with a maximum of 3 components, choosing the number of components that minimizes the AIC. For the two oldest bins, we observe three components, where the two furthest components increase both in weight and distance as age decreases. For the two youngest bins, the distribution is best represented by a single Gaussian. We find that the position of the component with the highest weight is located at $R_{GC} = 7.40$ kpc, 9.25 kpc, 10.19 kpc, 10.88 kpc as

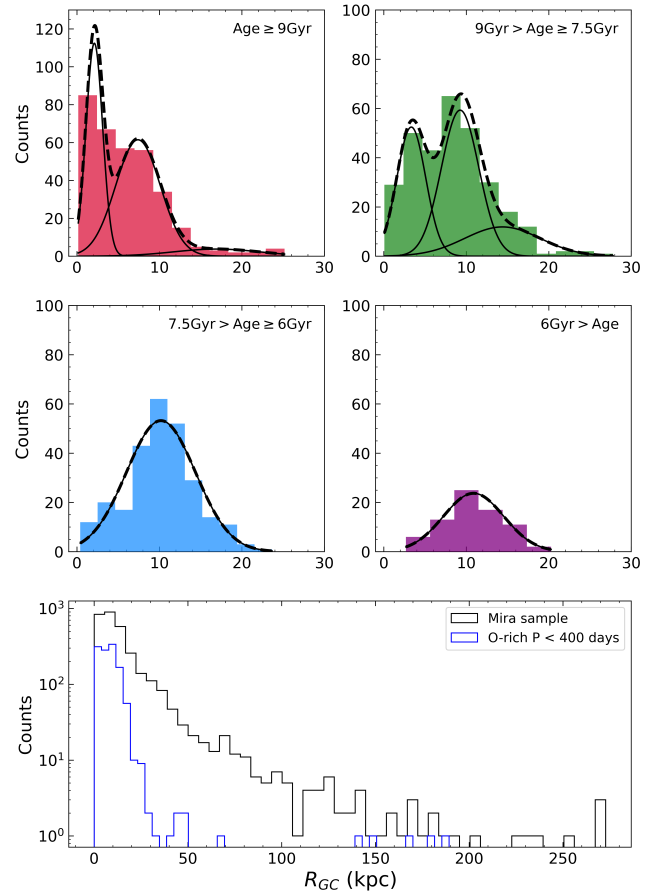


Fig. 13. Histograms of galactocentric distance (R_{GC}) binned by age. *Top and middle panels:* radius of stars in the Galaxy binned by age. The age range decreases from top to bottom and left to right. Dashed black lines represent the Gaussian mixture model fitted to the distribution. Individual components are presented as bold black lines. *Bottom:* full histogram of galactocentric distances for the full sample (black) and the O-rich Miras with periods under 400 days (blue).

age decreases, pointing to a reverse correlation between the age of the stars and their radius in the Galaxy. Although saturation effects decrease the observed number of Miras at low R_{GC} , this effect should mainly impact the component closest to the Galactic center and have a slight effect at higher R_{GC} . This is indeed seen in the decreasing weight of the innermost component as age decreases, with it completely disappearing for the youngest bins. We note that the absolute count per bin is not directly related to the underlying distribution of stars as a function of R_{GC} , as the innermost bins include stars closer than the GC and eastward or westward of it, while at larger R_{GC} we are only sampling a narrow cone. Despite the observational effects influencing the distribution of stars, the scenario described by the correlation between the position and age of stars in the sample is consistent with the inside-out formation of the Galactic disk, where star formation started in the inner regions of the MW and gradually moved outward (Chiappini et al. 1997; Schönrich & McMillan 2017). The presence of a centrally confined component traced only by old Miras (≥ 8 Gyr) is also consistent with a dominantly old component in the Galactic bulge (see Joyce et al. 2023, and references therein).

In the bottom panel of Fig. 13, we see the distribution of distances for the full sample (black) and the O-rich Miras with periods shorter than 400 days (blue). We see that the

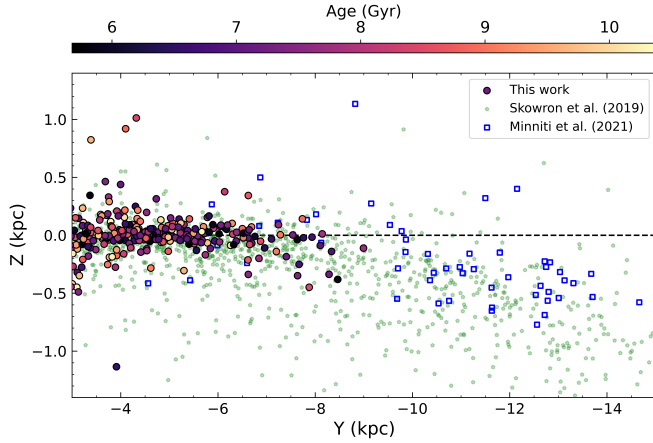


Fig. 14. Projection of our Mira sample in Galactic Y versus Z coordinates. Circles represent Miras color-coded by their estimated age; We include the classical Cepheids samples from Skowron et al. (2019, green polygons) and Minniti et al. (2021, blue squares). We include only sources in the respective samples inside the region $-15 \text{ kpc} \lesssim Y \lesssim -5 \text{ kpc}$ and $-1.5 \text{ kpc} \lesssim Z \lesssim 1.5 \text{ kpc}$.

vast majority of the sample is located well within the Galactic disk, with the noticeable exception of ~ 10 distance indicator Miras. Miras at these distances were also found by Nikzat et al. (2022) using VVV/OGLE data and are likely members of substructure within the Galactic halo, such as unknown globular clusters, dwarf spheroidal galaxies, or even the Sagittarius dwarf spheroidal galaxy (Ibata et al. 1994).

As is shown in Fig. 11, a significant number of Miras lies in the region affected by the warp of the Galactic disk, from $-15 \text{ kpc} \lesssim Y \lesssim -5 \text{ kpc}$ and $-1.5 \text{ kpc} \lesssim Z \lesssim 0 \text{ kpc}$. In Fig. 14 we explore the position of Miras in our catalog in that region. We find that Miras do follow the warping of the Galactic disk, aligned with the distribution of classical Cepheids located in the region. We do not find any correlation between the Mira position in the Galaxy and star-forming regions such as the spiral arms, likely due to their ages ($\gtrsim 5 \text{ Gyr}$) and migration effects inside the Galaxy.

6. Kinematics

In this section, we analyze the kinematic footprint of our Mira sample using the VVV PM data. The distribution of longitudinal proper motions ($\mu_\ell^* \cos b$), calibrated to the *Gaia* reference system and projected to a line of sight through the Galactic center, is shown in Fig. 15, as a function of the heliocentric distance. Stars in the Bulge region have a wider spread in proper motions consistent with older bulge populations (see Fig. 5 from Olivares Carvajal et al. 2024). As the distance increases, a transition to disk-like kinematics is observed, following the model by Mróz et al. (2019) based on the near disk and projected to a line of sight through the Galactic center.

As was mentioned in Sect. 3.3, we now investigate the signal of the NSD seen by our sample of Miras and determined distances. In Fig. 16 we present the histograms of PMs of stars in our sample with distances consistent with the projected position of the NSD in the Galaxy. Following Fig. 7 from Sanders et al. (2024), we apply cuts in Galactic coordinates $|\ell| \leq 0.4^\circ$ and $|b| \leq 0.4^\circ$ and study two period ranges from $100 \text{ days} \leq P \leq 250 \text{ days}$ and $450 \text{ days} \geq P$. Additionally, we impose a cut in $R_{GC} \leq 1 \text{ kpc}$ to isolate the signal of the NSD. It is important to notice that this

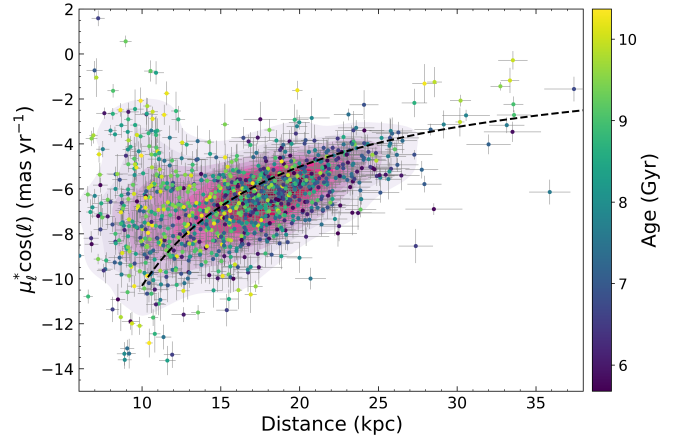


Fig. 15. Rotation curve in proper motion space. O-rich Miras in our sample are shown as circles color-coded by age. Pink contours show the density percentiles. The dashed line represents a model of the rotation curve from Mróz et al. (2019) as seen for a line of sight through the Galactic center.

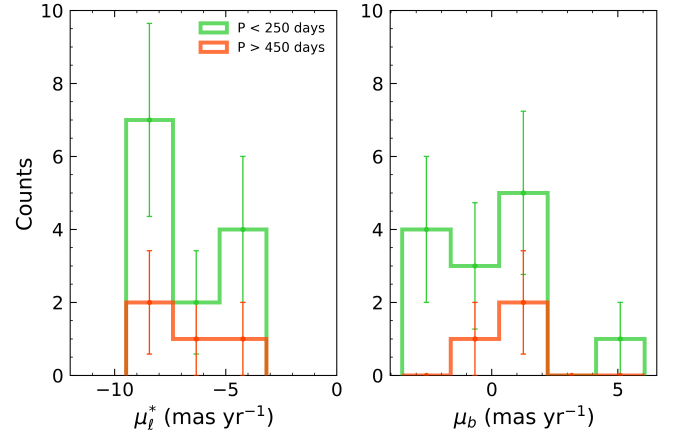


Fig. 16. Histogram of PMs for stars consistent with the NSD position. *Left*: histogram of longitudinal proper motions. Green lines correspond to Miras with periods in the range $100 \text{ days} \leq P \leq 250 \text{ days}$ and orange to $P \geq 450 \text{ days}$. The Galactic center has $\mu_\ell = -6.41 \text{ mas yr}^{-1}$. *Right*: same as left panel for μ_b .

cut far exceeds the usual radius of the NSD from Sormani et al. (2022) of $R_{NSD} \lesssim 300 \text{ pc}$, yet we found it necessary to include a minimum amount of stars for a reasonable comparison, and this value is still consistent with the model from Debattista et al. (2018). From Fig. 16, we observe a slight double-peaked distribution for the older period range and no peaks for the youngest Miras in the period ranges proposed in Sanders et al. (2024). This slight bimodality for Miras of periods shorter than 250 days cannot be explained by a rotating structure such as the NSD, as the corresponding distribution in μ_b is not consistent with disk-like kinematics. We conclude that in the Mira sample compiled and analyzed in this work, no strong signal of the NSD is observed.

To further inspect the kinematics of Miras in the bulge and far disk, we calculated the transverse velocities detailed in Du et al. (2020):

$$v_{\ell,HC}^* = 4.74 \frac{\mu_\ell^*}{\text{mas yr}} \frac{d}{\text{kpc}} \text{ km s}^{-1}, \quad (9)$$

where $v_{\ell,HC}$ is the transverse velocity relative to the Sun. To transform these velocities to a coordinate system in the Galactic

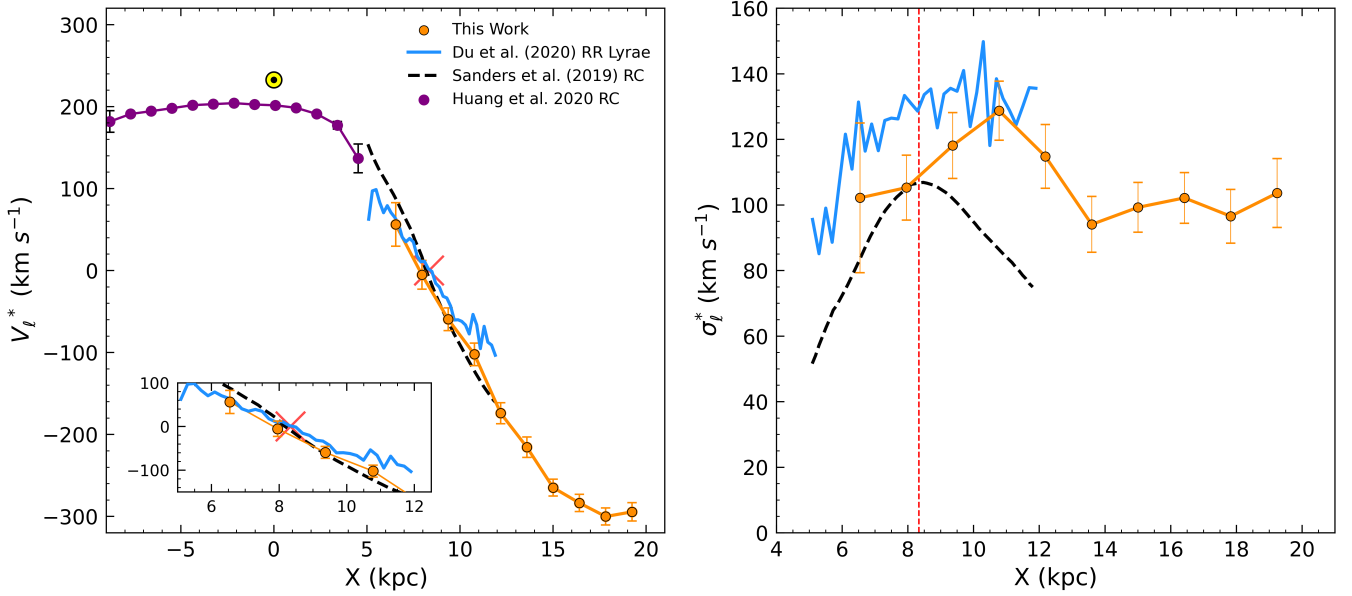


Fig. 17. Rotation curve of the Galaxy seen by our Mira sample. *Left:* Galactocentric Transverse velocity *Right:* velocity dispersion of the Mira sample. We add the corresponding curves from Sanders et al. (2019) (black) and Du et al. (2020) (blue). We include the rotation curve derived from RC stars from Huang et al. (2020) (purple). The velocity and position of the Sun are showcased in yellow. The red cross and the dashed line indicate the position of the Galactic center at a value of $V_l^* = 0$.

center $v_{l,GC}$, we use the relation,

$$v_{l,GC}^* = v_{l,HC}^* - U_\odot \sin \ell \cos b + (V_\odot + V_{LSR}) \cos \ell \cos b, \quad (10)$$

where $(U_\odot, V_\odot, W_\odot)$ is the solar peculiar velocity relative to the Local Standard of Rest (V_{LSR}). Here, we adopted the values (11.1, 12.24, 7.25) km s⁻¹ reported by Schönrich et al. (2010) for the solar peculiar velocity, and $V_{LSR} = 232.8$ km s⁻¹ (McMillan 2017). We binned the sample by distance, including only stars with $P < 400$ days, errors in distance $\leq 10\%$, and errors in $\mu_l^* \leq 1$ mas/yr. These cuts result in 776 stars. Figure 17 shows their average transverse velocity (left), along with the velocity dispersion within each bin (right). We also include rotation curves based on RR Lyrae (Du et al. 2020; average of metal-rich and metal-poor components) and Red Clump (RC) stars (Sanders et al. 2019). To compare with the rotation of the near side of the Galactic disk, we calculate the rotation velocities for the sample of RC stars presented in Huang et al. (2020). We only focus on stars that have ages comparable to our sample (i.e., age ≥ 5 Gyr) as the adopted cut in a period of 400 days translates into an age of ~ 5.6 Gyr. We see a nice continuity between the rotation curves of nearby RC stars, bulge RR Lyrae, bulge RC stars, and Miras.

Specifically, we find great agreement with the RR Lyrae-based rotation curve from Du et al. (2020) for the Mira closer to the Galactic center. For further distances, at $X \sim 11$ kpc, we see that the shape of the rotation curve resembles the distribution of RC stars by Sanders et al. (2019). At distances larger than ~ 15 kpc, we find that the rotation curve flattens. Even though this effect is expected (Mróz et al. 2019), this can also be attributed to the lower sensitivity of proper motions at large distances.

We showed in Fig. 7 that the distance distribution obtained with the MW-based P-L by Sanders (2023) would be similar to the present one. However, we note that for the rotation curve in Fig. 17 with the MW-based P-L, the derived distances would be larger, particularly for the Miras in the outer disk. This directly impacts the V_l , as is shown in Eq. (10). This, combined with

the given PMs, results in larger rotation velocities (>300 km/s) at large R_{GC} . In that case, the rotation curve in Fig. 17 would not be symmetric and would not show the expected flattening at $V_l^* \sim -250$ km/s. This is the main reason why our data favor the LMC-based P-L relation, over the one based on *Gaia* parallaxes for nearby Miras.

7. Discussion and conclusions

We have developed an algorithm to identify and characterize Mira candidates in the VVV survey, filtering candidates from all available Mira-like variables in order to match a validated sample of bona fide Miras. We used the photometric information from both the VVV survey and WISE to obtain average magnitudes in multiple filters. After a final validation through visual inspection, we found 3602 Miras in the VVV survey that we consider as a pure sample.

We have compiled and released, in this work, the largest catalog of variable stars to date on the far side of the Galactic disk, providing mean magnitudes, distances, and PMs.

The designed GP algorithm is tailored to find Mira candidates based on the K_s band time series of the VVV survey. The modeled K_s band is then used to determine average magnitudes in other VVV filters and WISE data. We also note that our choice of modeling for the long median trends in Miras was a simple polynomial to favor the detection of Miras with reliable average magnitudes. As a consequence, Miras with complex median magnitude trends might have been neglected from the analysis.

We used the validated VarZoo sample, imposing a higher likelihood threshold to determine a Mira-based extinction law in the inner region of the Galaxy. We found great agreement with the extinction laws found by other studies such as Minniti et al. (2020, classical Cepheids) and Majaess et al. (2016, multiple variable stars).

Concerning the distinction between O-rich and C-rich Miras, we analyzed the separation in color excess planes and found clear

differences between bona fide C-rich and O-rich Miras, as is shown in Fig. 9. After disentangling the populations, we derived a new extinction law and computed distances to the O-rich Miras, finding most of the sample in the bulge region past the Galactic center, on the far side of the Galactic disk, as is seen in Fig. 11. The kinematic information via PMs is consistent with the rotation of the Galaxy, determined using several probes on this side of the disk (see Fig. 15).

To exploit the correlation between the period and the age of the sample, we made bins of different ages and studied the distance distribution. As shown in Fig. 13, we found a strong correlation between the distances and age of the sample, with younger stars lying at larger Galactocentric distances. This is consistent with the inside-out picture of the formation of the Galactic disk. We also investigated the warping of the Galactic disk traced by the Mira sample in Fig. 14, finding that both the magnitude and the location of the warp are consistent with other studies based on younger tracers (Skowron et al. 2019; Dékány et al. 2019; Minniti et al. 2021).

We studied the Galactic rotation curve by means of the transverse velocities, as is shown in Fig. 17. We found great consistency in the bulge region with other tracers, especially with RR Lyrae (Du et al. 2020). This consistency can be understood as the shorter-period (older) Miras in our sample are more dominant closer to the Galactic center, tracing older populations in the same age bracket as RR Lyrae. At $X \sim 12$ kpc, the average age of Miras, per bin, resembles that of RC stars ($\lesssim 9$ Gyr), and we see a nice agreement with the rotation curve from Sanders et al. (2019). For the furthest Mira ($X \sim 14$ kpc), we see a flattening of the rotation curve at $V_{\ell}^* \sim -260$ km/s, consistent with the flattening of the rotation curve observed with classical Cepheids (Mróz et al. 2019). However, this could also be attributed to the sensitivity of PM-based velocities at these large distances. Still, the shape of the rotation curve is consistent with RC stars from Huang et al. (2020) in the near side of the Galactic disk, showing a symmetric shape with respect to the Galactic center. We also observe that the velocity dispersion increases close to the Galactic center, then decreases as we approach the Galactic disk. The dispersion remains flat at higher distances due to the errors in PMs introducing a higher dispersion.

The catalog presented in this study represents the largest sample of variable stars on the far side of the Galactic disk and opens an opportunity to study intermediate-age and old stellar populations in this under-explored region of the MW.

Data availability

The full version of Table D.1 is available in electronic form at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/693/A28>.

Acknowledgements. We would like to thank the anonymous referee for suggestions that helped to improve this manuscript. We gratefully acknowledge the help from Massimo Griggio for the supplied data. R.A. acknowledges Manuel Cavieres, Francisco Jara-Ferreira, Nicolás Cristi and Benjamín Silva for the insightful discussions. R.A. acknowledges support for this work by the National Agency for Research and Development (ANID) BASAL Center for Astrophysics and Associated Technologies (CATA) through grants AFB170002, ACE210002, and FB210003, and by the ANID Millennium Science Initiative, ICN12_009 and AIM23-0001, awarded to the Millennium Institute of Astrophysics (MAS). M.Z. and A.R.A. acknowledges support from DICYT through grant 062319RA and from ANID through FONDECYT Regular grant No. 1230731. J.O.C. acknowledges support from the ANID Doctorado Nacional grant 2021-21210865 and by ESO grant SSDF21/24. F.G. gratefully acknowledges support from the French

National Research Agency (ANR)- funded projects "MWDisc" (ANR-20-CE31-0004) and "Pristine" (ANR-18-CE31-0017). A.V.N. acknowledges support from ANID Scholarship Program Doctorado Nacional 2020–21201226. Based on observations collected at the European Southern Observatory under ESO programs 0103.D0386(A), 105.20MY.001, 179.B2002, and 198.B2004. We gratefully acknowledge the use of data from the VVV ESO Public Survey program ID 179.B-2002 taken with the VISTA telescope and data products from the Cambridge Astronomical Survey Unit (CASU). The VVV Survey data are made public at the ESO Archive. Based on observations taken within the ESO VISTA Public Survey VVV, Program ID 179.B-2002. This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, and NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology. WISE and NEOWISE are funded by the National Aeronautics and Space Administration. The Geryon cluster at the Centro de Astro-Ingeniería UC was extensively used for the calculations performed in this paper. BASAL CATA PFB-06, the Anillo ACT-86, FONDECUIP AIC-57, and QUIMAL 130008 provided funding for several improvements to the Geryon cluster. This publication makes use of data products from the Two Micron All Sky Survey (2MASS), which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. This work has also 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. It also made use of NASA's Astrophysics Data System and of the Vizier catalog access tool, CDS, Strasbourg, France (Wenger et al. 2000). The original description of the Vizier service was published in (Ochsenbein et al. 2000). Finally, we acknowledge the use of the following publicly available software: Celerite (Foreman-Mackey et al. 2017), TOPCAT (Taylor 2005), pandas (Wes McKinney 2010), IPython (Pérez & Granger 2007), numpy (van der Walt & Varoquaux 2011), matplotlib (Hunter 2007), Astropy, a community developed core Python package for Astronomy (Astropy Collaboration 2013, 2018), and Aladin sky atlas (Bonnarel et al. 2000; Boch & Fernique 2014).

References

- Alonso-García, J., Minniti, D., Catelan, M., et al. 2017, *ApJ*, 849, L13
 Anderson, J., Bedin, L. R., Piotto, G., Yadav, R. S., & Bellini, A. 2006, *A&A*, 454, 1029
 Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33
 Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123
 Baba, J., & Kawata, D. 2020, *MNRAS*, 492, 4500
 Bailer-Jones, C. A. L. 2023, *AJ*, 166, 269
 Bellini, A., Anderson, J., van der Marel, R. P., et al. 2014, *ApJ*, 797, 115
 Boch, T., & Fernique, P. 2014, in *Astronomical Society of the Pacific Conference Series*, 485, Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset, & P. Forshay, 277
 Bonnarel, F., et al. 2000, *A&AS*, 143, 33
 Brewer, B. J., & Stello, D. 2009, *MNRAS*, 395, 2226
 Catelan, M., & Smith, H. A. 2015, *Pulsating Stars* (Weinheim: Wiley-VCH)
 Chiappini, C., Matteucci, F., & Gratton, R. 1997, *ApJ*, 477, 765
 Chiavassa, A., Kravchenko, K., & Goldberg, J. A. 2024, *Liv. Rev. Comput. Astrophys.*, 10, 2
 Cioni, M. R. L. 2010, in *EAS Publications Series*, 40, eds. L. Spinoglio, & N. Epchtein, 137
 Cioni, M. R. L., Blommaert, J. A. D. L., Groenewegen, M. A. T., et al. 2003, *A&A*, 406, 51
 Contreras Ramos, R., Zoccali, M., Rojas, F., et al. 2017, *A&A*, 608, A140
 Cristallo, S., Piersanti, L., Straniero, O., et al. 2011, *ApJS*, 197, 17
 Cutri, R. M., Mainzer, A., Conrow, T., et al. 2015, *Explanatory Supplement to the NEOWISE Data Release Products*
 de Sá-Freitas, C., Fragkoudi, F., Gadotti, D. A., et al. 2023, *A&A*, 671, A8
 De Simone, R., Wu, X., & Tremaine, S. 2004, *MNRAS*, 350, 627
 Debattista, V. P., Earp, S. W. F., Ness, M., & Gonzalez, O. A. 2018, *MNRAS*, 473, 5275
 Dékány, I., Minniti, D., Catelan, M., et al. 2013, *ApJ*, 776, L19
 Dékány, I., Hajdu, G., Grebel, E. K., & Catelan, M. 2019, *ApJ*, 883, 58
 Du, H., Mao, S., Athanassoula, E., Shen, J., & Pietrukowicz, P. 2020, *MNRAS*, 498, 5629
 Eggen, O. J. 1998, *AJ*, 115, 2435
 Feast, M. W. 1963, *MNRAS*, 125, 367
 Feast, M. W., Glass, I. S., Whitelock, P. A., & Catchpole, R. M. 1989, *MNRAS*, 241, 375

- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *AJ*, **154**, 220
- Freytag, B., & Höfner, S. 2023, *A&A*, **669**, A155
- Freytag, B., Liljegren, S., & Höfner, S. 2017, *A&A*, **600**, A137
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gail, H. P., Zhukovska, S. V., Hoppe, P., & Trieloff, M. 2009, *ApJ*, **698**, 1136
- Glass, I. S., & Feast, M. W. 1982, *MNRAS*, **199**, 245
- González-Fernández, C., Hodgkin, S. T., Irwin, M. J., et al. 2018, *MNRAS*, **474**, 5459
- Grady, J., Belokurov, V., & Evans, N. W. 2020, *MNRAS*, **492**, 3128
- Graham, M. J., Drake, A. J., Djorgovski, S. G., Mahabal, A. A., & Donalek, C. 2013a, *MNRAS*, **434**, 2629
- Graham, M. J., Drake, A. J., Djorgovski, S. G., et al. 2013b, *MNRAS*, **434**, 3423
- He, S., Yuan, W., Huang, J. Z., Long, S., & Macri, L. M. 2016, *AJ*, **152**, 164
- Höfner, S., & Olofsson, H. 2018, *A&A Rev.*, **26**, 1
- Höfner, S., Bladh, S., Aringer, B., & Ahuja, R. 2016, *A&A*, **594**, A108
- Höfner, S., Bladh, S., Aringer, B., & Eriksson, K. 2022, *A&A*, **657**, A109
- Huang, C. D., Riess, A. G., Hoffmann, S. L., et al. 2018, *ApJ*, **857**, 67
- Huang, Y., Schönrich, R., Zhang, H., et al. 2020, *ApJS*, **249**, 29
- Huang, C. D., Yuan, W., Riess, A. G., et al. 2024, *ApJ*, **963**, 83
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, *Nature*, **370**, 194
- Ita, Y., & Matsunaga, N. 2011, *MNRAS*, **412**, 2345
- Iwanek, P., Kozłowski, S., Gromadzki, M., et al. 2021a, *ApJS*, **257**, 23
- Iwanek, P., Soszyński, I., & Kozłowski, S. 2021b, *ApJ*, **919**, 99
- Iwanek, P., Soszyński, I., Kozłowski, S., et al. 2022, *ApJS*, **260**, 46
- Iwanek, P., Poleski, R., Kozłowski, S., et al. 2023, *ApJS*, **264**, 20
- Joyce, M., Johnson, C. I., Marchetti, T., et al. 2023, *ApJ*, **946**, 28
- Kerschbaum, F., Groenewegen, M. A. T., & Lazaro, C. 2006, *A&A*, **460**, 539
- Kramer, D., Gowanlock, M., Trilling, D., McNeill, A., & Erasmus, N. 2023, *Astron. Comput.*, **44**, 100711
- Lebzelter, T., Mowlavi, N., Marigo, P., et al. 2018, *A&A*, **616**, L13
- Lomb, N. R. 1976, *Ap&SS*, **39**, 447
- Mainzer, A., Bauer, J., Grav, T., et al. 2011, *ApJ*, **731**, 53
- Majaess, D., Turner, D., Dékány, I., Minniti, D., & Gieren, W. 2016, *A&A*, **593**, A124
- Matsunaga, N., Kawadu, T., Nishiyama, S., et al. 2009, *MNRAS*, **399**, 1709
- Matsunaga, N., Feast, M. W., Bono, G., et al. 2016, *MNRAS*, **462**, 414
- Matsunaga, N., Menzies, J. W., Feast, M. W., et al. 2017, *MNRAS*, **469**, 4949
- McMillan, P. J. 2017, *MNRAS*, **465**, 76
- Menzies, J. W., Whitelock, P. A., & Feast, M. W. 2015, *MNRAS*, **452**, 910
- Merrill, P. W. 1923, *ApJ*, **58**, 215
- Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, *New A*, **15**, 433
- Minniti, D., Saito, R. K., Gonzalez, O. A., et al. 2018, *A&A*, **616**, A26
- Minniti, J. H., Sbordone, L., Rojas-Arriagada, A., et al. 2020, *A&A*, **640**, A92
- Minniti, J. H., Zoccali, M., Rojas-Arriagada, A., et al. 2021, *A&A*, **654**, A138
- Mróz, P., Udalski, A., Skowron, D. M., et al. 2019, *ApJ*, **870**, L10
- Neowise Team 2020, *NEOWISE-R Single Exposure (L1b) Source Table*, NASA IPAC DataSet, IRSA144
- Nicholls, C. P., Wood, P. R., Cioni, M. R. L., & Soszyński, I. 2009, *MNRAS*, **399**, 2063
- Nikzat, F., Ferreira Lopes, C. E., Catelan, M., et al. 2022, *A&A*, **660**, A35
- Nishiyama, S., Tamura, M., Hatano, H., et al. 2009, *ApJ*, **696**, 1407
- Nogueras-Lara, F., Schödel, R., Gallego-Calvente, A. T., et al. 2019, *A&A*, **631**, A20
- Ochsenbein, F., et al. 2000, *A&AS*, **143**, 23
- Olivares Carvajal, J., Zoccali, M., De Leo, M., et al. 2024, *A&A*, **687**, A312
- Ordenes-Huanca, C., Zoccali, M., Bayo, A., et al. 2022, *MNRAS*, **517**, 6191
- Ou, J.-Y., & Ngeow, C.-C. 2022, *AJ*, **163**, 192
- Pérez, F., & Granger, B. E. 2007, *Comput. Sci. Eng.*, **9**, 21
- Pietrukowicz, P., Kozłowski, S., Skowron, J., et al. 2015, *ApJ*, **811**, 113
- Pietrzyński, G., Graczyk, D., Gellenne, A., et al. 2019, *Nature*, **567**, 200
- Press, W. H., & Rybicki, G. B. 1989, *ApJ*, **338**, 277
- Prudil, Z., Dékány, I., Catelan, M., et al. 2019, *MNRAS*, **484**, 4833
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, *ApJ*, **783**, 130
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, *ApJ*, **885**, 131
- Sanders, J. L. 2023, *MNRAS*, **523**, 2369
- Sanders, J. L., & Matsunaga, N. 2023, *MNRAS*, **521**, 2745
- Sanders, J. L., Smith, L., Evans, N. W., & Lucas, P. 2019, *MNRAS*, **487**, 5188
- Sanders, J. L., Matsunaga, N., Kawata, D., et al. 2022, *MNRAS*, **517**, 257
- Sanders, J. L., Kawata, D., Matsunaga, N., et al. 2024, *MNRAS*, **530**, 2972
- Scargle, J. D. 1982, *ApJ*, **263**, 835
- Schönrich, R., & McMillan, P. J. 2017, *MNRAS*, **467**, 1154
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, **403**, 1829
- Skowron, D. M., Skowron, J., Mróz, P., et al. 2019, *Science*, **365**, 478
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Smith, L. C., Lucas, P. W., Kurtev, R., et al. 2018, *MNRAS*, **474**, 1826
- Sormani, M. C., Sanders, J. L., Fritz, T. K., et al. 2022, *MNRAS*, **512**, 1857
- Soszyński, I., Udalski, A., Kubiak, M., et al. 2005, *Acta Astron.*, **55**, 331
- Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009, *Acta Astron.*, **59**, 239
- Soszyński, I., Dziembowski, W. A., Udalski, A., et al. 2011, *Acta Astron.*, **61**, 1
- Stetson, P. B. 1987, *PASP*, **99**, 191
- Suh, K.-W. 2018, *J. Korean Astron. Soc.*, **51**, 155
- Suh, K.-W., & Hong, J. 2017, *J. Korean Astron. Soc.*, **50**, 131
- Suresh, A., Karambelkar, V., Kasliwal, M. M., et al. 2024, *PASP*, **136**, 084203
- Taylor, M. B. 2005, in *Astronomical Society of the Pacific Conference Series*, **347**, Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, 29
- Templeton, M. R., Mattei, J. A., & Willson, L. A. 2005, *AJ*, **130**, 776
- Trabucchi, M., & Mowlavi, N. 2022, *A&A*, **658**, L1
- Trabucchi, M., Wood, P. R., Montalbán, J., et al. 2019, *MNRAS*, **482**, 929
- Udalski, A., Szymański, M. K., & Szymański, G. 2015, *Acta Astron.*, **65**, 1
- Urago, R., Omodaka, T., Nagayama, T., et al. 2020, *ApJ*, **891**, 50
- Utenthaler, S. 2015, in *Astronomical Society of the Pacific Conference Series*, **497**, Why Galaxies Care about AGB Stars III: A Closer Look in Space and Time, eds. F. Kerschbaum, R. F. Wing, & J. Hron, 49
- Utenthaler, S., van Stiphout, K., Voet, K., et al. 2011, *A&A*, **531**, A88
- Utenthaler, S., McDonald, I., Bernhard, K., Cristallo, S., & Gobrecht, D. 2019, *A&A*, **622**, A120
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Comput. Sci. Eng.*, **13**, 22
- Vassiliadis, E., & Wood, P. R. 1993, *ApJ*, **413**, 641
- Wang, S., & Chen, X. 2019, *ApJ*, **877**, 116
- Wenger, M., et al. 2000, *A&AS*, **143**, 9
- Wes McKinney 2010, in *Proceedings of the 9th Python in Science Conference*, eds. S. van der Walt, & J. Millman, 56
- Whitelock, P. A., Feast, M. W., van Loon, J. T., & Zijlstra, A. A. 2003, *MNRAS*, **342**, 86
- Whitelock, P. A., Feast, M. W., & Van Leeuwen, F. 2008, *MNRAS*, **386**, 313
- WISE Team 2020, *AllWISE Multiepoch Photometry Table*, NASA IPAC DataSet, IRSA134
- Wood, P. R. 2015, *MNRAS*, **448**, 3829
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Yuan, W., Macri, L. M., He, S., et al. 2017, *AJ*, **154**, 149
- Yuan, W., Macri, L. M., Javadi, A., Lin, Z., & Huang, J. Z. 2018, *AJ*, **156**, 112
- Zechmeister, M., & Kürster, M. 2009, *A&A*, **496**, 577
- Zhang, H., & Sanders, J. L. 2023, *MNRAS*, **521**, 1462
- Zijlstra, A. A., Loup, C., Waters, L. B. F. M., et al. 1996, *MNRAS*, **279**, 32
- Zijlstra, A. A., Bedding, T. R., & Mattei, J. A. 2002, *MNRAS*, **334**, 498
- Zoccali, M., Rojas-Arriagada, A., Valenti, E., et al. 2024, *A&A*, **684**, A214

Appendix A: Gaussian Process Modeling.

Previous studies have found great success using GPs to model light curves of several types of stellar variability (Brewer & Stello 2009), including Mira variables (He et al. 2016; Yuan et al. 2017, S22). A general GP can be described as a stochastic model consisting of a mean function ($\mu_\theta(x)$) and a covariance generally named Kernel ($\mathcal{K}(\alpha, \theta)$). The set of parameters from the mean function and the Kernel functions (x, θ, α) can be optimized, maximizing the likelihood function ($\ln \mathcal{L}(\theta, \alpha)$).

GP algorithms are generally avoided for large dataset applications due to their computational complexity. However, limiting the problem to one-dimensional data and using combinations of stationary exponential kernels greatly improves the performance (Foreman-Mackey et al. 2017). This approach is implemented in the python package CELERITE (Foreman-Mackey et al. 2017) chosen for this work.

Each light curve that meets the variability criteria detailed in Sect. 3 is then fed to a grid of GPs with kernels described by Eq. (A.1).

$$\mathcal{K}_{ij} = \mathcal{K}_{SHO} \sum_{i=0}^1 \mathcal{K}_{DRW}^i + \sum_{j=0}^1 \mathcal{K}_{DRW}^j + \mathcal{K}_\sigma \quad (\text{A.1})$$

Where \mathcal{K}_{SHO} is a damped harmonic oscillator kernel, \mathcal{K}_{DRW} is a damped random walk kernel, and K_σ is a white noise kernel. both i and j have allowed values of 0 and 1, meaning 4 possible kernel combinations were considered in the analysis. We select a kernel combination based on the AIC, thus rewarding the use of simpler models for the light curves.

For each iteration of the model, the fitted period was recovered from the maximum value of the power spectral density (PSD) of the fitted kernel. Since all considered kernels only include 1 term corresponding to a periodic kernel, the peak of the PSD corresponds to the dominant frequency according to the model, from which the period is recovered.

We emphasize that we chose to only include one periodic kernel (\mathcal{K}_{SHO}), so models will only fit for one strongly periodic signal, while accommodating for stochastic effects and noise. However, complex median trends than cannot be captured by the chosen kernel combination will report a low \mathcal{L} , and thus, were automatically neglected. As is discussed in Sect. 3, this is made by design to isolate distance tracers, at the expense of possible real Miras not being considered in the final catalog.

Appendix B: Saturated and blended sources.

As was previously noted, effects from saturation in the VVV survey begin to manifest for sources brighter than $K_s \leq 12$ (Contreras Ramos et al. 2017). To investigate this effect, we perform a crossmatch between our VVV catalog and GALACTICNUCLEUS (Nogueras-Lara et al. 2019). We show the result of the crossmatch for non-variable sources in gray and for stars in our sample in Fig. B.1. Only VVV magnitudes have been corrected by the pulsation of the star, and thus, this introduces a scatter in the linear relation.

As was mentioned, both VVV and GALACTICNUCLEUS begin to show saturation effects at $K_s \leq 12$ and $K_s \leq 11$, respectively. The shorter exposure times and smaller pixels in the HAWK-I camera from GALACTICNUCLEUS make saturation effects less apparent in this saturated regime.

To further inspect the possible effect of saturation in the VVV photometry, we perform a match between the catalog and

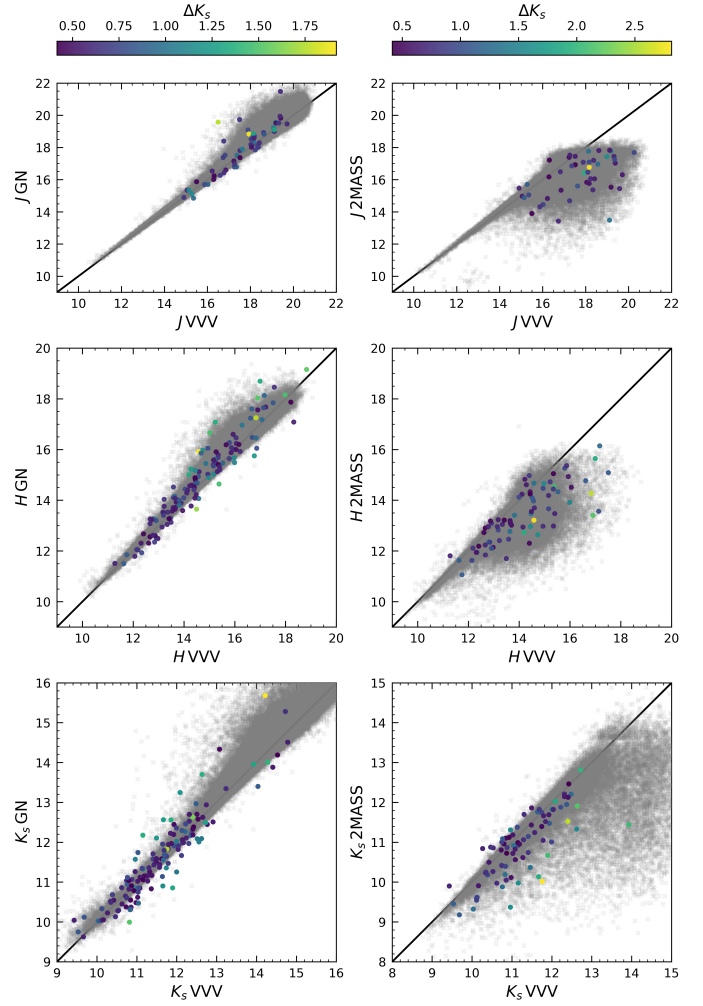


Fig. B.1. Comparison of magnitudes in GALACTICNUCLEUS, 2MASS, and VVV. Panels show the comparison of the same bands in the respective surveys. Gray points show all matched sources, and stars in our catalog present in both surveys are shown color-coded by our modeled amplitude in the K_s band. Solid black lines indicate a 1:1 relation between magnitudes. We see no evidence of biases in the reported magnitudes between the surveys.

the 2MASS catalog (Skrutskie et al. 2006). 2MASS begins to show saturation effects at $K_s \leq 8.5$, effectively being saturation free at the K_s magnitude range important for this study. The comparison between these catalogs is shown in the right hand side panels in Fig. B.1. We observe that the non-variable sources follow a linear relation in both catalogs, however, the confusion effects in the 2MASS catalog adds a significant scatter to the relation by over-estimating the brightness of the sources. Despite this, Miras in common between the catalogs do not show a clear bias or trend for stars fainter than $K_s \leq 10$, and scatter around a 1:1 relation is driven by the lack of correction by pulsation in the 2MASS magnitudes and the confusion of the photometry.

We conclude that even at the saturation regime of VVV, no significant bias in the Mira sources for magnitudes $K_s \geq 10$ is observed. Thus, we deem the VVV magnitudes as reliable even at these saturated magnitudes for the purposes of this study.

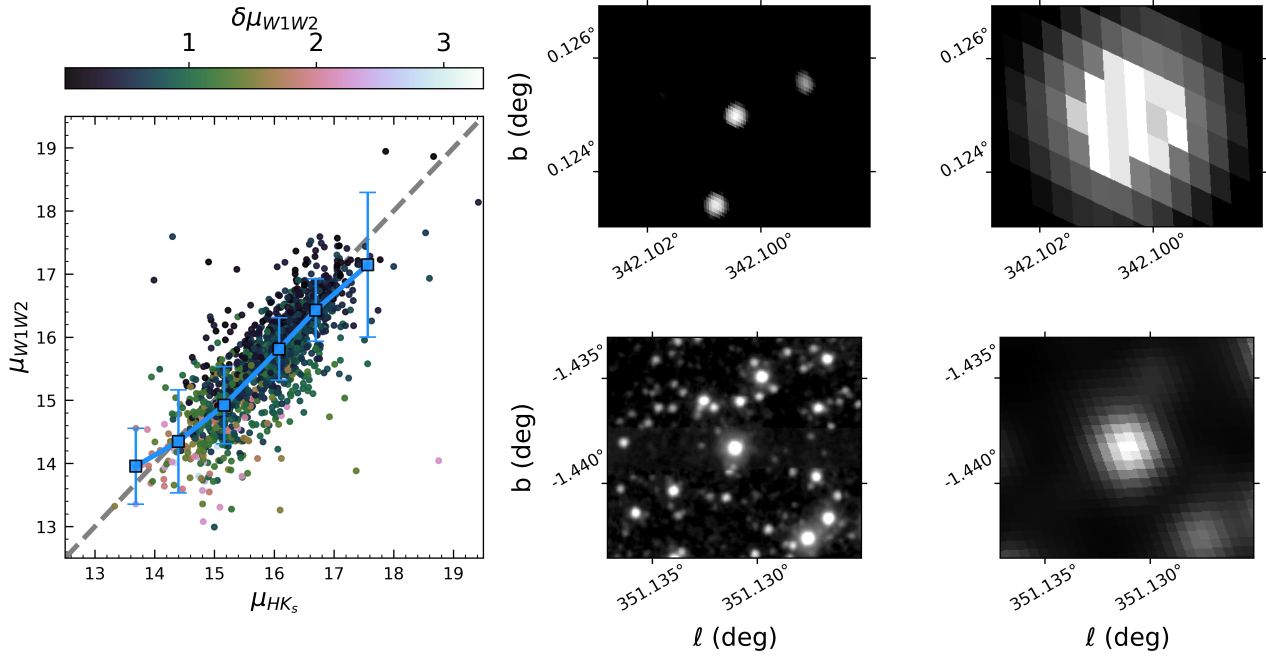


Fig. C.1. Distance modulus comparison between VVV based distances adopted here and WISE-based ones. *Left:* Distance modulus comparison for Miras with WISE photometry available. Miras are shown in dots colored by the error in the WISE based distance modulus. The dashed gray line shows a 1:1 relation between modulus. Blue squares show the median per bin. *Middle panels:* VVV images in the H band of two Miras in our catalog. *Right panels:* WISE images in the W1 band of the same Miras as the middle panels. Here, the blending caused by the larger PSF from WISE is clear.

Appendix C: WISE distances.

In this appendix, we discuss the distance determination for our sample using photometry from the WISE survey. We follow the same procedure described in Sect. 5.1, utilizing the average magnitudes in the W1 and W2 bands detailed in Sect. 3.1, and applying the extinction law from Wang & Chen (2019).

In the left panel of Fig. C.1, we present the comparison between our adopted VVV-based distance modulus and the WISE-based ones. We exclude sources with $W1 < 8.1$ and $W2 \leq 6.7$, as these are heavily saturated in the respective bands (Cutri et al. 2015), leading to unreliable distance determinations. Overall, we observe good agreement between the adopted distances and those derived from WISE photometry; however, the VVV-based distances are systematically slightly larger across the entire range.

To investigate this discrepancy, we examine images from both the WISE survey (top panels in Fig.C.1) and the VVV survey (bottom panels in Fig.C.1). Due to the lower spatial resolution of WISE, nearby sources are blended with the target Mira variable, contributing additional flux to the observed magnitudes in the WISE bands. This blending effect results in brighter observed magnitudes, which leads to smaller distance modulus when compared to those derived from the higher-resolution VVV data. The additional flux from contaminant sources impacts both the observed magnitude and the derived extinction, shifting the distance distribution to lower values.

As was mentioned for the VVV distances in Sect. 5.1, the majority of the error budget from Eq. (8) is dominated by the extinction term. For the WISE-based distances, the comparatively larger uncertainties in the extinction law from Wang & Chen (2019), relative to our study, result in overall larger distance errors. Another interesting feature in Fig. C.1 is the larger errors in WISE distances for closer Miras. We attribute this behavior to

the closer Miras being near the WISE saturation range, resulting in higher errors.

From this, we conclude that although WISE-based distances are generally reliable, blending near the Galactic plane has a considerable effect on the distances. The higher spatial resolution of the VVV survey reduces the impact of blending from nearby sources, resulting in more accurate distances.

Appendix D: Catalog Example.

In this appendix, we include the header of our final catalog, available at the CDS.

ID	ℓ deg	b deg	P days	J mag	δJ mag	H mag	δH mag	K_s mag	δK_s mag	W1 mag	$\delta W1$ mag	W2 mag	$\delta W2$ mag	W3 mag	$\delta W3$ mag	W4 mag	$\delta W4$ mag	d kpc	δd kpc	type	μ_α masyr ⁻¹	μ_δ masyr ⁻¹
b33220745043	-2.291557	0.011355	182.0	17.16	0.04	13.32	0.02	11.39	0.003	-	-	-	-	-	-	-	-	16.9	0.4	O	-6.20	-0.39
b33550389427	2.030869	-0.078912	277.8	16.84	0.11	12.86	0.08	10.69	0.004	9.03	0.12	8.11	0.13	-	-	-	-	15.2	0.9	O	-8.35	-0.25
b32810812649	-8.149293	-0.309705	202.6	18.27	0.03	13.68	0.01	11.28	0.003	9.72	0.01	9.07	0.02	8.42	0.03	6.73	0.18	13.2	0.3	O	-5.23	0.39
b30521329762	-0.139154	-1.590120	254.1	11.96	0.02	11.03	0.02	10.04	0.002	-	-	-	-	-	-	-	-	21.4	0.6	C	-3.02	-1.62
d0734062863	-15.069632	-0.240343	518.1	-	-	18.35	0.09	13.55	0.005	9.67	0.06	7.44	0.06	5.62	0.10	4.35	0.11	22.7	2.7	C	-7.99	0.50
b30350392048	-3.855943	-2.254720	530.5	-	-	13.42	0.07	10.62	0.009	5.90	0.01	5.21	0.01	3.00	0.14	1.34	0.16	20.5	2.4	C	-5.55	-2.32
b33651156387	4.192921	-0.012697	277.0	19.12	0.04	14.23	0.02	11.64	0.002	10.26	0.01	9.33	0.01	8.96	0.25	6.60	0.19	18.2	0.7	O	-5.75	-0.82
d11031023247	-16.263470	0.798953	189.0	14.33	0.06	12.60	0.06	10.87	0.004	9.43	0.11	8.89	0.12	7.65	0.10	6.84	0.13	15.5	0.7	O	-11.35	-2.62
b25411639020	7.417047	-6.940512	468.4	14.18	0.16	13.01	0.13	11.87	0.004	9.22	0.05	6.95	0.05	3.66	0.10	1.46	0.11	84.0	9.6	C	-7.63	0.38
b33910291605	7.610559	0.114559	335.0	18.91	0.04	13.78	0.02	11.12	0.003	9.43	0.02	8.53	0.02	9.88	0.11	-	-	16.1	0.7	O	-6.79	-0.39
b31720969793	-3.381637	-0.574201	445.4	15.85	0.06	12.85	0.06	10.75	0.004	7.91	0.01	6.65	0.01	5.51	0.10	4.65	0.11	26.1	2.1	C	-6.20	-0.87
b30330874013	-3.683423	-2.413739	673.4	-	-	13.16	0.03	10.36	0.006	6.58	0.03	5.12	0.01	2.78	0.11	0.72	0.12	25.6	4.1	C	-11.09	1.90
b33121184475	-3.355265	-0.060248	382.4	18.27	0.07	13.37	0.05	10.82	0.002	9.10	0.00	8.00	0.00	7.44	0.05	6.97	0.08	16.9	0.9	O	-7.17	-1.09
b31630239842	-5.459128	-0.701698	495.1	-	-	11.83	0.01	9.80	0.004	6.91	0.03	5.86	0.03	5.03	0.08	4.30	0.10	20.3	1.9	C	-9.90	-0.69
b31850962773	-1.697197	-0.581808	486.6	18.37	0.02	14.36	0.05	11.28	0.005	7.80	0.05	5.92	0.03	4.27	0.13	2.33	0.15	20.9	2.0	C	-12.30	1.86
d10810440571	-19.915534	0.028824	270.6	18.83	0.08	14.30	0.06	11.95	0.003	-	-	-	-	-	-	-	-	23.9	1.2	O	-6.87	0.12
b33560444936	1.983577	-0.358159	135.0	16.80	0.11	13.63	0.10	11.93	0.005	10.14	0.06	9.76	0.07	-	-	-	-	19.3	1.2	O	-6.29	0.03
b33420338839	0.225009	0.031585	120.1	16.27	0.03	12.54	0.01	10.81	0.002	9.02	0.06	8.40	0.06	-	-	-	-	10.2	0.2	O	-6.84	-0.55
d1125039047	-13.998910	0.470195	145.6	15.26	0.01	12.30	0.01	10.76	0.004	9.75	0.02	9.16	0.02	9.62	0.06	-	-	13.4	0.2	O	-8.56	0.13
b34650374368	-2.427427	1.048136	602.4	-	-	17.27	0.27	13.23	0.008	6.98	0.21	4.79	0.27	4.25	0.70	2.54	0.79	38.9	8.3	C	-4.04	-7.43
b3204024990	0.441987	-0.635423	331.1	-	-	16.02	0.02	11.76	0.003	9.38	0.01	7.87	0.01	6.32	0.02	-	-	8.1	0.4	O	-7.59	1.12
b32840438006	-8.178010	-0.176617	545.0	20.32	0.09	14.36	0.01	11.21	0.003	8.62	0.01	7.20	0.01	5.39	0.10	1.63	0.12	22.5	2.6	O	-8.16	-0.63
b33430475918	0.251130	-0.230975	394.5	17.81	0.21	12.81	0.18	10.20	0.006	7.95	0.01	6.41	0.04	5.13	0.06	4.70	0.08	12.5	1.5	O	0.16	1.18
b38761431710	-2.709962	4.586580	530.5	17.09	0.01	16.27	0.02	15.54	0.003	-	-	-	-	-	-	-	-	694.4	74.2	C	-24.35	5.01
b32950343869	-6.781690	-0.001258	480.8	-	-	15.29	0.06	11.71	0.003	8.74	0.08	7.20	0.09	6.37	0.08	5.83	0.13	18.3	1.8	O	-6.07	-0.88
b33430439937	0.337004	-0.157376	186.6	15.73	0.00	11.96	0.01	10.25	0.004	8.94	0.06	8.12	0.06	-	-	-	-	11.7	0.2	O	-7.31	6.45
d1111216378	-14.851840	0.079387	485.4	19.19	0.11	13.74	0.09	10.89	0.004	8.86	0.05	7.65	0.05	6.83	0.04	5.88	0.06	19.8	2.1	O	-4.57	1.62
b32961614518	-5.771306	-0.290818	662.3	15.95	0.01	14.56	0.12	12.14	0.005	7.87	0.13	5.97	0.14	3.42	0.13	1.53	0.15	71.4	12.2	C	-3.06	-8.41
b32930449905	-6.902989	-0.210528	320.0	17.26	0.04	12.90	0.04	10.58	0.002	8.97	0.00	8.11	0.00	7.77	0.03	6.81	0.09	14.9	0.7	O	-8.83	-0.13
d0752138561	-11.607840	-0.065624	175.0	13.05	0.06	11.14	0.05	9.98	0.003	9.14	0.01	8.77	0.01	8.37	0.02	-	-	13.6	0.5	O	-8.55	-0.43
b33241627731	-1.237215	-0.134515	490.2	17.68	0.11	13.97	0.09	10.97	0.003	10.72	0.05	8.48	0.05	7.69	0.03	3.78	0.02	19.1	2.0	O	-5.54	-0.15
b33621531138	4.417195	0.043633	823.1	-	-	16.02	0.04	11.69	0.008	5.92	0.24	3.70	0.26	0.37	0.43	-1.59	0.49	25.0	5.2	C	-3.13	0.48
b34411683989	-4.361689	0.654729	475.1	-	-	18.01	0.10	12.50	0.004	8.97	0.06	6.74	0.07	4.96	0.03	3.00	0.04	8.0	0.9	O	-11.89	3.06

Table D.1. Median magnitudes and Proper Motions for Miras in our catalog.