Mira variables in the Milky Way's nuclear stellar disc: discovery and classification

Jason L. Sanders,[1,2]★ Noriyuki Matsunaga,[3] Daisuke Kawata,[4] Leigh C. Smith,[1,5] Dante Minniti,[6,7] and Philip W. Lucas[5]

[1]*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*

[2]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

[3]*Department of Astronomy, School of Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[4]*Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK*

[5]*Centre for Astrophysics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*

[6]*Instituto de Astrofísica, Facultad de Ciencias Exactas, Universidad Andrés Bello, Fernández Concha 700, Las Condes, Santiago, Chile*

[7]*Vatican Observatory, Vatican City State, V-00120, Italy*

ABSTRACT

The properties of the Milky Way's nuclear stellar disc give crucial information on the epoch of bar formation. Mira variables are promising bright candidates to study the nuclear stellar disc, and through their period–age relation dissect its star formation history. We report on a sample of 1782 Mira variable candidates across the central $3 \times 3 \, \mathrm{deg}^2$ of the Galaxy using the multi-epoch infrared VISTA Variables in Via Lactea (VVV) survey. We describe the algorithms employed to select candidate variable stars and then model their light curves using periodogram and Gaussian process methods. By combining with WISE, 2MASS and other archival photometry, we model the multi-band light curves to refine the periods and inspect the amplitude variation between different photometric bands. The infrared brightness of the Mira variables means many are too bright and missed by VVV. However, our sample follows a well-defined selection function as expected from artificial star tests. The multi-band photometry is modelled using stellar models with circumstellar dust that characterise the mass loss rates. We demonstrate how $\gtrsim 90$ per cent of our sample is consistent with O-rich chemistry. Comparison to period–luminosity relations demonstrates that the bulk of the short period stars are situated at the Galactic Centre distance. Many of the longer period variables are very dusty, falling significantly under the O-rich Magellanic Cloud and solar neighbourhood period–luminosity relations and exhibit high mass-loss rates of $\sim 2.5 \times 10^{-5} M_\odot \, \mathrm{yr}^{-1}$. The period distribution appears consistent with the nuclear stellar disc forming $\gtrsim 8$ Gyr ago although it is not possible to disentangle the relative contributions of the nuclear stellar disc and the contaminating bulge.

Key words: stars: variables: general – stars: AGB – Galaxy: centre – Galaxy: bulge – Galaxy: stellar content

1 INTRODUCTION

In studies of the Milky Way, we are often interested in piecing together the series of events that resulted in what we observe today. In this way, we can study the Milky Way as a detailed exemplar galaxy in the cosmological context of star-forming galaxies across the Universe (Bland-Hawthorn & Gerhard 2016; Barbuy et al. 2018). One key component of the Milky Way is the bar(-bulge) (Blitz & Spergel 1991; Wegg & Gerhard 2013). The bar is an important dynamical driver in the Milky Way responsible for significant restructuring of both the stars and the gas within the disc. Knowledge of the time over which it has had a dynamical impact on the Milky Way is crucial to understanding the dynamical history of the entire Galaxy.

One structure intimately linked to the formation of the Milky Way's bar-bulge is the nuclear stellar disc (NSD). The NSD is a flattened distribution of stars with radius ~ 250 pc (Launhardt et al. 2002), aspect ratio $\sim 5 : 1$ (Nishiyama et al. 2013; Gallego-Cano et al. 2020) and mass $\sim 10^9 M_\odot$ (Sormani et al. 2022) that rotates at approximately $100 \, \mathrm{km \, s^{-1}}$ as confirmed through both radial velocity (Lindqvist et al. 1992; Schönrich et al. 2015; Matsunaga et al. 2015; Schultheis et al. 2021) and proper motion studies (Shahzamanian et al. 2022). The NSD sits between the larger scale Galactic bar/bulge and the nuclear stellar cluster (see the review of Schödel et al. 2014), and coincides with the central molecular zone (CMZ), a region of significant interstellar dust and gas (Morris & Serabyn 1996).

Beyond the Milky Way, nuclear stellar discs are often observed in barred spiral galaxies (Erwin & Sparke 2002; Pizzella et al. 2002; Gadotti et al. 2018; Gadotti et al. 2020). A consistent picture for their formation has been built up from hydrodynamical simulations (e.g. Athanassoula 1992): once a bar forms in a disc galaxy, gas is readily funnelled along the bar towards the centre of the galaxy where it can settle on central 'x2' orbits forming nuclear gas rings. The gas then begins forming stars which approximately inherit the 'x2' orbital geometry and so the resulting stellar population resembles a disc. This paradigm is supported by observations of external galaxies that show nuclear stellar discs are younger, more metal-rich and of a

★ E-mail: jason.sanders@ucl.ac.uk (JLS)

lower velocity dispersion than the surrounding bar stars (Gadotti et al. 2020; Bittner et al. 2020). In the Milky Way, this picture is supported by observations that the NSD and CMZ overlap spatially and kinematically (Schönrich et al. 2015; Schultheis et al. 2021), gas is visibly being funnelled along the bar (Hatchfield et al. 2021) and the CMZ is star-forming today (Morris & Serabyn 1996). Baba & Kawata (2020) have suggested this connection between the history of the NSD and the bar is a way to pin down the epoch of bar formation in the Milky Way. Their simulations demonstrated that the formation of a bar is rapidly followed by a ~ 1 Gyr long intense period of star formation that forms the NSD. The oldest NSD stars then give the bar's formation time. Note that there are studies of the age of stars within the Galactic bar-bulge (e.g. Bovy et al. 2019; Hasselquist et al. 2020) but crucially the dynamical age of the bar can be quite different from the age of the bar stars.

There have been relatively few studies of the detailed star formation history of the NSD. Figer et al. (2004) argued from Hubble Space Telescope photometry that the star formation history of the NSD was quite continuous over time particularly when compared to the Galactic bulge fields that were more consistent with ancient bursts of star formation. Schultheis et al. (2020, 2021) have similarly demonstrated that the metallicity distribution of NSD stars differs from that of the NSC and the Galactic bulge, giving further evidence of its separate formation channel (and possibly epoch). The more extended GALACTICNUCLEUS photometric survey (Nogueras-Lara et al. 2019) allowed a fuller analysis of the colour–magnitude diagrams across the NSD from which Nogueras-Lara et al. (2020) demonstrated that the low number of stars in the earlier analysis of Figer et al. (2004) did not enable clear discrimination between a bursty and continuous star formation history, and instead the giant-branch luminosity functions across a more extended range of fields were consistent with a star formation history with an early (> 8 Gyr) burst and then lower levels until a recent (< 1 Gyr) burst. A very recent burst is corroborated by observations of classical Cepheids in this region with periods indicating ages of ~ 25 Myr (Matsunaga et al. 2011) and confirmation that at least some fraction of the population in these regions must be very old ($\gtrsim$ 10 Gyr) comes from the detection of RR Lyrae there (Minniti et al. 2016; Molnar et al. 2022). However, for probing the detailed star formation at intermediate ages, the differences in the giant branch luminosity function with age are quite subtle (see figure 8 of Nogueras-Lara et al. 2020). For example, when one goes beyond ages of ~ 2 Gyr the red clump has a weak ~ 0.015 mag Gyr^{-1} gradient with age (Girardi 2016; Chen et al. 2017; Huang et al. 2020a) and one must instead rely on the relative fraction and location of red clump giants to red-giant-branch-bump stars.

Alternative age tracers for the NSD are Mira variables. Mira variables are thermally pulsating asymptotic giant branch stars, and are typically recognised as the final stages of the giant branch life of a low to intermediate mass star (Catelan & Smith 2015). Nearly all asymptotic giant branch stars pulsate to some degree through a mechanism driven by convection (Freytag et al. 2017; Xiong et al. 2018). The range of pulsation modes form an entire family of different long period variables (Wood 2015) of which Mira variables have been identified as those pulsating in the fundamental mode with the highest amplitudes, $\Delta V > 2.5$ mag, and periods in the range 80 to 1000 days. Their light curve shapes are distinguished from the similar, but lower amplitude, semi-regular variables (SRV) and OGLE small amplitude red giants (OSARG) by a more regular, near sinusoidal nature although long-term trends and variations in the period are observed (Zijlstra et al. 2002; He et al. 2016; Ou & Ngeow 2022) possibly related to thermal pulses (Vassiliadis & Wood 1993), the

interactions of pulsation with convective flow (Freytag et al. 2017) or the presence of circumstellar dust (Whitelock et al. 2003; Ou & Ngeow 2022). Additionally, as evidenced clearly in observations of the Large Magellanic Cloud (LMC), the classes of long period variables lie on distinct period–luminosity sequences (Wood et al. 1999; Wood 2000; Soszyński et al. 2009), with the Mira variables lying along a single sequence (Glass & Evans 1981; Feast et al. 1989; Ita et al. 2004; Groenewegen 2004; Fraser et al. 2008; Riebel et al. 2010; Ita & Matsunaga 2011; Yuan et al. 2017a,b; Bhardwaj et al. 2019; Iwanek et al. 2021b). The tight period–luminosity relation and high luminosities of Mira variables have made them ideal standard candles both for cosmological studies (Huang et al. 2018, 2020b) and Local Group and Galactic structure studies (Menzies et al. 2011; Whitelock et al. 2013; Menzies et al. 2015; Catchpole et al. 2016; Deason et al. 2017; Menzies et al. 2019; Grady et al. 2019, 2020). Having well-understood Mira variables across a range of local environments will enable their precise calibration as a cosmological tracer, particularly in the era of the James Webb Space Telescope and the Vera Rubin Observatory.

It has long been observationally known that solar neighbourhood Mira variables with shorter periods have hotter kinematics (Merrill 1923) and more extended profiles perpendicular to the Galactic plane (Feast 1963). This behaviour is indicative of shorter period variables belonging to older populations that have undergone more dynamical heating. Furthermore, older LMC and Milky Way clusters are hosts to shorter period Mira variables (Grady et al. 2019). The period of a Mira variable is largely governed by the mass and radius of the star and Mira-like pulsations only begin once a star has reached a narrow radial range at a given mass (Trabucchi et al. 2019). It is therefore expected that the period is a direct indicator of mass, and hence age of the star. However, there are relatively limited theoretical studies of the Mira variable period–age relation (Wyatt & Cahn 1983; Feast & Whitelock 1987; Eggen 1998; Trabucchi & Mowlavi 2022), and stellar population work has largely been done using period–age relations empirically calibrated from the solar-neighbourhood correlations with kinematics (Feast & Whitelock 1987, 2000b; Feast et al. 2006; Feast & Whitelock 2014; Catchpole et al. 2016; López-Corredoira 2017; Grady et al. 2020; Nikzat et al. 2022). A typical period–age $(P - \tau)$ relation is $\tau \approx 13 \,\mathrm{Gyr} \frac{1}{2}(1 + \tanh((330\,\mathrm{day} - P)/250\,\mathrm{day}))$ where we see Mira variables of $(200, 300, 400)$ day periods have ages of ~ $(9.5, 7.5, 4.5)$ Gyr. However, it is anticipated that there is a significant spread in age at each period with Trabucchi & Mowlavi (2022) reporting a 3 Gyr range for the age of a 350 day period Mira variable using a set of theoretical pulsation models. Nonetheless, due to their potentially excellent resolution at intermediate ages and their high intrinsic brightness, Mira variables offer ideal age tracers for the inner Galaxy.

The first step in using Mira variables to constrain the star formation history of the NSD is to reliably identify them. Early work in this area targeted OH/IR maser stars in the very central regions of the Galaxy (e.g. Blommaert et al. 1998; Wood et al. 1998). These searches are biased towards longer period stars. Both Glass et al. (2001) and Matsunaga et al. (2009, M09) undertook broader searches for variable stars and have presented samples of Mira variables around the Galactic Centre. The VISTA Variables in Via Lactea (VVV) survey is a multi-epoch infrared survey that has taken observations of the Galactic bulge over a ~10-year baseline. This makes it an ideal survey for extending the sample of long-period variables in the NSD. This is the goal of our work. Our paper is structured as follows: we begin by describing the selection and light-curve modelling of Mira variable candidates in Section 2. Additional details on the specifics of the light curve modelling are given in Appendix A. We go on to
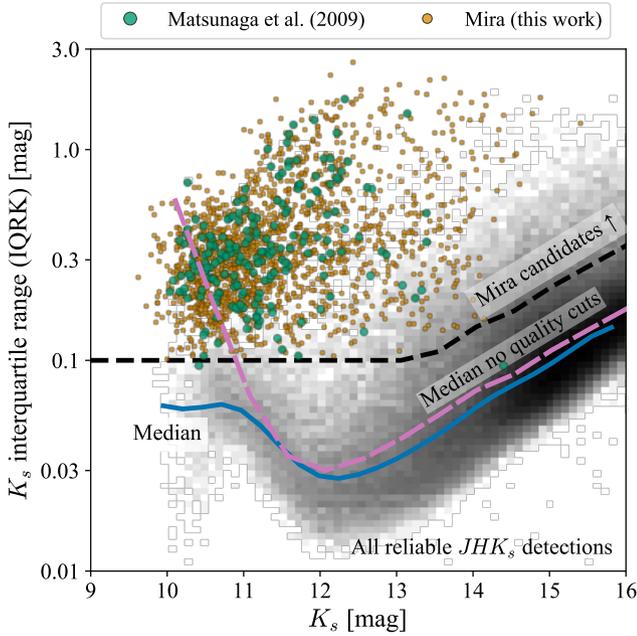
Figure 1. Selection of Mira candidates in K_s magnitude against the interquartile range of K_s (IQRK): background greyscale shows the distribution of all VIRAC2 sources with J, H and K_s detections in a field 10×20 arcmin^2 centred on $(\ell, b) = (0.08, 0)$ deg using only those detections in each light curve that satisfy our astrometric and photometric quality cuts. The blue line gives the median trend and the pink dashed line shows the median trend without applying the astrometric and photometric quality cuts. The black dashed marks the line above which stars are considered candidate Mira variables. The larger green points show the Mira variables from Matsunaga et al. (2009) and smaller orange the final Mira variable sample in this paper.

inspect the properties of our sample in Section 3 focusing on the spatial distribution, the selection effects that impact our sample, the photometric classification and the period–age distribution. We close with our conclusions in Section 4. This is the first of two papers on this sample. Our second paper will focus on the kinematic properties of the sample probed through the proper motion data provided by VIRAC2 (Smith et al. 2018; Sanders et al. 2019, Smith et al., in prep.).

2 DISCOVERY OF NEW NSD MIRA VARIABLES

We describe the sequence of steps taken to extract a sample of NSD Mira variable candidates. First, we begin by describing the primary dataset employed, the VIRAC2 K_s light curve set. We go on to describe the initial sample of likely variable stars and the methods used for modelling their light curves. From this sample, we define a series of cuts to isolate the Mira variable candidates. We further check the quality of our light curve modelling by comparison to overlap Mira variables in the literature. We finally model the multi-band light curves of a set of Mira variable candidates to further refine the periods and perform a visual inspection to weed out any remaining contaminants.

2.1 Primary light curve sample

Our primary source of data is the VISTA Variables in Via Lactea (VVV) survey (Minniti et al. 2010; Saito et al. 2012). The VVV survey is a multi-epoch near-infrared ($ZYJHK_s$) survey of the Galactic bulge and disc conducted using the 16 detector VIRCAM camera (Dalton et al. 2006) mounted on the 4.1m VISTA telescope (Sutherland et al. 2015) at the Cerro Paranal Observatory. The initial 560 deg^2 survey ran from 2010 to 2015 and covered the Galactic bulge ($|\ell| < 10$ deg, -10 deg $< b < 5$ deg) and the southern Galactic disc separated into 1.5×1.1 deg^2 tiles. The primary observations were taken in the K_s band (~ 160 observations except for 8 high cadence tiles with ~ 600 observations) with additional $ZYJH$ observations typically taken at the beginning and end of the survey. In 2016, the VVVX survey commenced, extending the sky coverage of both the bulge and disc components and providing more epochs for the region covered by the initial survey. VVVX extended the JHK_s coverage of the original VVV data resulting in at least ~ 200 K_s observations for each source and an average of ~ 30 J and ~ 20 H observations per source. In order to focus on the NSD region, we only use VVV and VVVX data within Galactic coordinates $|\ell| < 1.5$ deg and $|b| < 1.5$ deg.

The VVV Infrared Astrometric Catalogue (VIRAC, Smith et al. 2018) was generated using VVV epoch aperture photometry (González-Fernández et al. 2018) and calculated relative proper motions (and parallaxes) for ~ 120 million sources. Using the second Gaia data release (Gaia Collaboration et al. 2016, 2018), Sanders et al. (2019) used bright overlapping sources between Gaia and VIRAC to anchor the relative proper motions to Gaia's absolute reference frame. The second version of VIRAC (VIRAC2, Smith et al., in prep.) utilises point spread function (PSF) photometry to deliver more accurate centroids and a deeper source catalogue, and calibrates astrometry to Gaia's reference frame for individual observations (rather than using a post hoc correction). Here we use a preliminary version of the final VIRAC2 dataset. Each VVV image has been processed with the PSF photometry fitting programme DoPhot (Schechter et al. 1993; Alonso-García et al. 2012) and the resulting photometry zero-point calibrated on a chip and time-dependent basis using a pool of 2MASS reference sources. Initial astrometric solutions were computed by grouping nearby detections and then improved by iteratively re-grouping detections based upon fitted astrometry (allowing for detections to be included in multiple groupings). The final set of detections grouped using the derived astrometry is our set of *light curves*.

The deeper photometry provided by PSF fitting comes at the expense of spurious sources detected in the wings of bright objects. Many of these spurious sources have similar magnitudes and variability to our target Mira variables. An initial list of reliable sources was obtained by requiring the sources are non-duplicate (defined as the source having less than 20 per cent of detections shared with other sources), have 10 or more epochs (these sources are fitted with a full five-parameter astrometric solution in VIRAC2) and are detected in more than 20 per cent of observations. Spurious sources are approximately randomly distributed around bright sources and given a large number of observations occasionally there is random alignment and grouping of the spurious sources. Requiring the source is detected in more than 20 per cent of observations is a trade-off between retaining genuine faint sources and removing these spurious sources. The cut is implemented in terms of a fraction of observations such that it is homogeneously applied across the entire VVV survey which can have quite large variations in epoch counts. For the $|\ell| < 1.5$ deg, $|b| < 1.5$ deg region, this results in a primary

catalogue of 27, 714, 023 sources (light curves). We further clean up the light curves of individual sources by removing detections based on the quality of their individual astrometric and photometric fits. Some bright ($11 \lesssim K_s \lesssim 12$) spurious detections can be identified by the high χ^2 of their PSF fits (e.g. see Braga et al. 2019). We reject all individual detections brighter than $K_S = 13.2$ with the DoPhot photometric fit chi-squared > 10. Furthermore, we remove all ambiguous detections (detections shared with another reliable source) and all detections with astrometric chi-squared deviation of the detection relative to the astrometric solution > 11.83 (approximately 3σ outliers for a two degree-of-freedom fit). We also always employ a single 5σ clip for the light curves (estimating σ from the 16th and 84th percentiles) to remove further outliers. Despite these quality cuts, we have found that variable seeing for blended sources can lead to spurious variability and periodicity. In Appendix B we discuss how we calibrate the light curves of suspected blended sources. We refer to this set of resultant light curves as 'cleaned'.

2.1.1 Complementary photometric data

In addition to VVV photometry, we will also utilise some other near infrared and longer wavelength photometry. We cross-match all of our candidates to the DECAPS catalogue (Schlafly et al. 2018, with a 0.3 arcsec radius), the GALACTICNUCLEUS catalogue (Nogueras-Lara et al. 2019, with a 0.4 arcsec radius), the 2MASS catalogue (Skrutskie et al. 2006, with a 1 arcsec radius), the AKARI catalogue (Ishihara et al. 2010, with a 5 arcsec radius), the GLIMPSE catalogue (Churchwell et al. 2009)[1] and the Spitzer-IRAC GALCEN point source catalogue of Ramírez et al. (2008) (with a 0.4 arcsec radius), preferentially keeping the GALCEN data over GLIMPSE, the WISE catalogue (Wright et al. 2010, with a 1 arcsec radius)[2], the 24 μm MIPSGAL catalogue (Gutermuth & Heyer 2015, with a 1.5 arcsec radius), the 7 and 15 μm ISOGAL catalogue (Omont et al. 2003, with a 1.5 arcsec radius) and the Herschel Infrared Galactic plane survey (Hi-GAL, Molinari et al. 2016, with a 3 arcsec radius). Although the point-spread function for some of the surveys is large and so contamination might be expected in the considered crowded regions, we assume the Mira variables are significantly brighter in the mid-infrared than any nearby stars such that contamination is minimal.

2.2 Light curve modelling

From the set of cleaned light curves, we form an initial Mira candidate list by finding highly variable sources in the K_s band. Our search is guided by the previous M09 search for Mira variables in the very inner 20×30 arcmin^2 area around the Galactic centre. These authors first selected stars with photometry three times more variable than the median variability at a given star's magnitude and found 1364 long period variable candidates, of which 549 were assigned periods. No detailed classification was performed such that some level of contamination from young stellar objects (e.g. Guo et al. 2022) is likely (although at the magnitude range this survey probed they would be foreground objects so likely dwarfed in number density by

---

[1] As the GLIMPSE-II catalogue puts a requirement on the sources having similar magnitudes at the two GLIMPSE-II epochs, some variable sources are not present in the combined catalogue and instead we use the results from the Epoch 1 catalogue. This affects ~ 200 stars in our final sample.

[2] We correct the WISE photometry of bright stars using the tables from https://wise2.ipac.caltech.edu/docs/release/neowise/expsup/sec2_1civa.html.
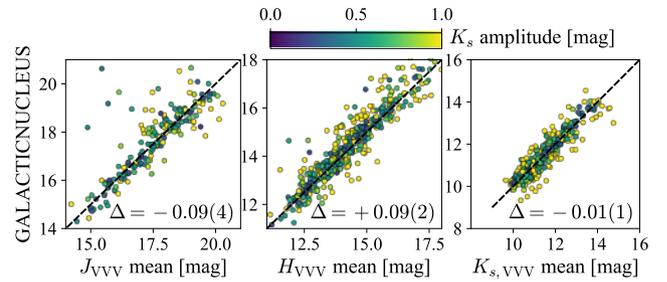

Figure 2. GALACTICNUCLEUS against VVV magnitudes (we plot the mean of our model fits) coloured by amplitude. Δ reports the magnitude offset (GALACTICNUCLEUS – VVV) with the bracketed digit the uncertainty in the last decimal place. In the saturated regime ($K_s \lesssim 11.5$) there is no significant bias. The scatter correlates with Mira variable amplitude.

the background long-period variables) and also that the long-period variables themselves are not guaranteed Mira variables but could contain a mixture of other semi-regular variables. The precise definition of a Mira variable is slightly awkward. Physically they are often linked with high-amplitude fundamental mode pulsation and hence membership of a particular period–luminosity sequence. From an observational perspective, this definition is often approximated as a pure amplitude cut (e.g Soszyński et al. 2013) although Trabucchi et al. (2021) acknowledge that this simple consideration removes lower amplitude stars on the same fundamental period–luminosity relation as the higher amplitude systems. In the M09 analysis, stars are considered as non-Mira variables if the amplitude in any of J, H or K_s is less than 0.4 although this removes only ~ 10 per cent of the stars in their sample. Here we attempt to emulate the fuller selection of M09 keeping in mind that the lower amplitude variables could be semi-regular variable contaminants.

Guided by M09, in each field we construct the median curve of the interquartile range of K_s (IQRK) as a function of K_s for sources detected in J, H and K_s (typically sources detected in all three bands are highly reliable although sources are lost due to no detections in the bluer bands in high extinction regions). No astrometric or photometric quality cuts were applied to this sample although it makes little difference to our selection. The expected IQRK varies significantly with location in the bulge due to both observation quality and crowding. For all cleaned light curves, we compute the median K_s and IQRK. We retain all light curves with more than 20 epochs and with IQRK > 0.1 mag and IQRK greater than 2 times the median line for $K_s > 12$ (see Fig. 1 for the IQRK cut employed). As shown in Fig. 1, this cut encompasses nearly all of the long-period variables with periods presented by M09. Although non-linearities from saturation begin at $K_s = 12$ we still consider sources brighter than this as otherwise we would reject many Mira variables and we have found that period estimation from VVV is still reliable for saturated sources. From Fig. 1, we see that when astrometric and photometric quality cuts are not applied to the parent sample, the IQRK rises significantly at the bright end. Removal of the astrometric and photometric outliers causes the IQRK to plateau at bright K_s at values significantly below the expected IQRK for Mira variables. This gives confidence that spurious variability from saturated sources is not a significant concern for the cleaned light curves. In Fig. 2 we show a comparison of the VVV modelled mean magnitudes compared to the GALACTICNUCLEUS measurements (Nogueras-Lara et al. 2019) for our final Mira variable sample. Although both VVV and GALACTICNUCLEUS suffer from saturation effects for $K_s \lesssim 11$,
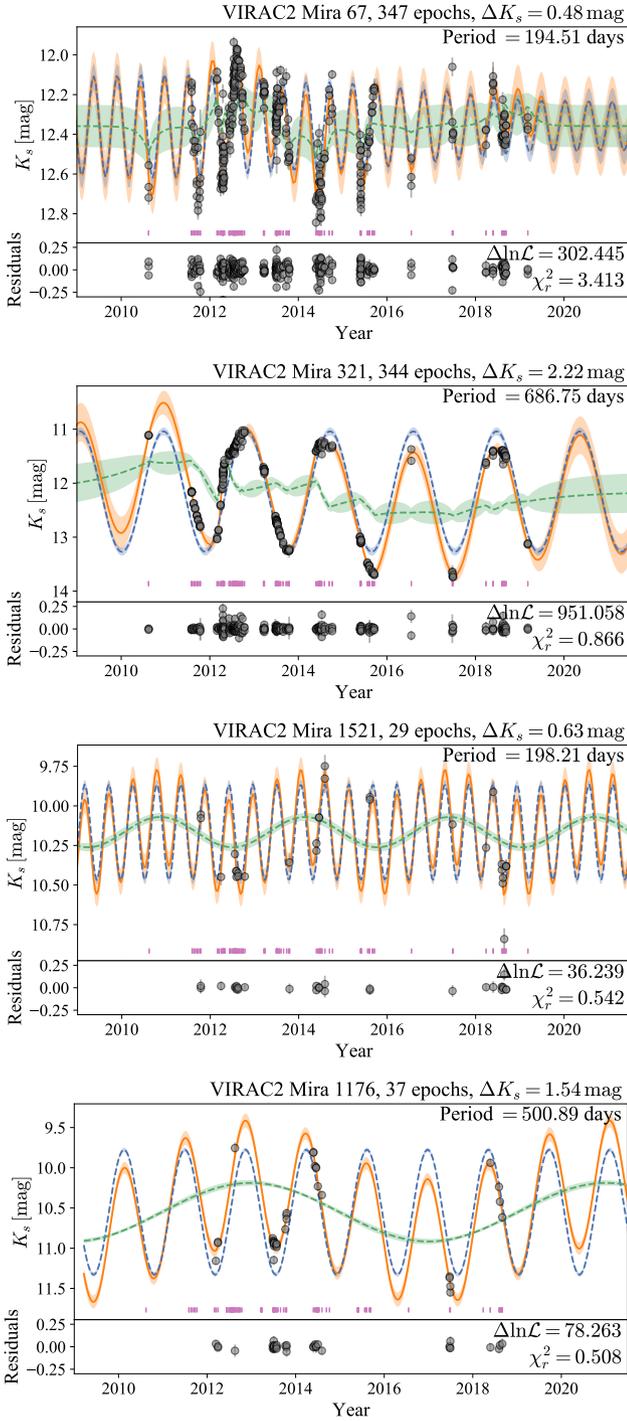
Figure 3. Example light curves for our Mira variable sample. Each set of panels show a different Mira light curve with the model residuals below. The grey points are the cleaned VIRAC2 light curves. The means and standard deviations of the Gaussian process models are shown in orange and the sub-components of the kernels in green and blue. The small pink ticks show the times at which the star could have been detected. $\Delta \ln \mathcal{L}$ gives the difference in log-likelihood with respect to a non-periodic model (with an additional variance) and χ_r^2 the reduced chi-squared of the fit. Note that deviations from pure periodicity are sometimes fitted with the random walk part of the kernel (top two panels) and sometimes a longer periodic signal (other panels). The four examples are chosen as two well-sampled light curves (top – the very top light curve is also in the M09 sample) and two poorly-sampled (bottom). The first and third light curves are low amplitude and short period, whilst the second and fourth are high amplitude and long period.
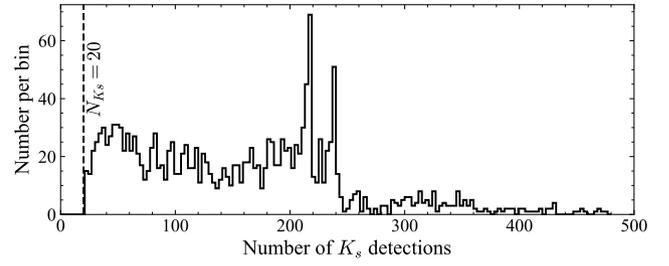

Figure 4. Distribution of the number of K_s detections for our Mira variable sample. The sources have to have at least 20 detections (vertical dashed line).

the effects are weaker in GALACTICNUCLEUS as it has shorter exposure times than VVV and the HAWK-I camera smaller pixels than VIRCAM (Nogueras-Lara et al. 2019). There is no visible bias between the modelled VVV mean magnitudes and GALACTICNU-CLEUS at the bright end giving some confidence in our use of the data in this regime.

From our set of candidate cleaned light curves, we must model the light curve properties to produce a set of Mira variables. The two key properties for identifying Mira variables are their long periods (> 80 days) and high amplitudes $0.25 < \Delta K_s < 3$ (M09, as discussed above, here we adopt a generous lower limit for K_s amplitude to match the long-period variable selection of M09 although as acknowledged by M09 it is expected stars with $\Delta K_s \lesssim 0.4$ are in fact semi-regular variables). We construct Fourier models for each light curve finding the best period. However, Mira light curves tend to not be completely periodic and exhibit a range of other behaviour with both short and long-term amplitude and period variability (Zijlstra et al. 2002; He et al. 2016; Molina et al. 2019). Therefore, as a second step we also use Gaussian process models that can capture quasi-periodic signals. We fully describe the details of these two methods, as well as their application to multi-band photometry, in Appendix A.

We first run a simple Lomb-Scargle periodogram searching periods from 0.5 days to the time-span of the light curve. We continue to consider the star if one of the top three periods (excluding aliases identified through the Lomb-Scargle periodogram on the magnitudes replace by noise) is > 10 days (with a false alarm probability less than 0.001) or if the period found by the 'string-length' method (Lafler & Kinman 1965) is > 10 days. Aliases are defined as the top five peaks in a periodogram of the magnitudes replaced by a constant value (VanderPlas 2018), as well as 1 and 1/2 day periods. On the remaining stars, we run a second Fourier fit with $N_F = 2$ Fourier terms and $N_P = 3$ polynomial terms using 10 days as the minimum period. If the best period is > 50 days, we run a grid of Gaussian process fits selecting the fit that gives the minimum Akaike information criterion. We use the kernel in equation (A4) with $N_O = (1, 2)$, $N_E = (0, 1)$, $\ln c_1 = (-3, -10)$ (the damping of the oscillation) and initialized with the top three periods from the Fourier fit. When two periodic terms are used, we take the period associated with the higher amplitude kernel term as the primary period. If, however, the higher amplitude period is over 1000 days, we use the lower-amplitude period as the primary period (provided it is under 1000 days). In Appendix C we demonstrate the quality of the period recovery using this procedure on a set of literature sources with VIRAC2 data.

In Fig. 3 we show four example light curves for Mira variables in our final sample. We see that the modelling is able to capture the primary periodic signal whilst also having the flexibility to model cycle-to-cycle variations either through a longer periodic component
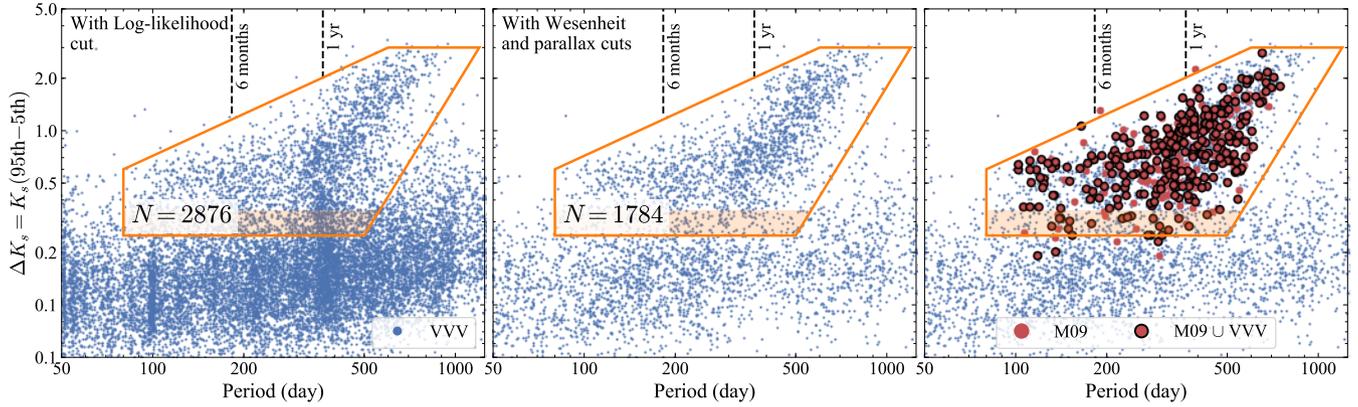
Figure 5. Period–amplitude plot for all processed variables. The left panel shows only sources satisfying our initial cut on log-likelihood whilst the central panel shows the addition of the cuts on Wesenheit magnitudes and parallax. In the right panel we overlay the M09 Mira variables either with black outlines if they have been included in our processing, or no outline if not. The Mira variable period–amplitude sequence is visible and demarcated by the orange selection box, which contains 1784 stars. The shaded region gives the amplitude range over which the variables are likely semi-regular variables. We mark on possible aliases of a year and 6 months which do not appear as overdensities in our parallax-cut sample.

or through a stochastic random-walk component. Fig. 3 also illustrates the range in number of K_s epochs we have for each source. We show the distribution of number of K_s epochs in our final Mira variable sample in Fig. 4. The range in number of detections arises in part due to the VVV and VIRCAM observing strategy (some sources are in the overlaps of the VIRCAM pointings necessary to fill each tile) and in part due to varying observing quality leading to varying levels of blending and saturation and hence unreliable detections not included in our light curve processing.

2.3 Mira variable selection

From our candidate list of long period variables, we remove all sources with a log-likelihood difference between the Gaussian process model and a constant model with an additional variance less than 10^3. This initial cut results in 23496 candidates. We then primarily select Mira variables using the period–amplitude diagram as shown in the left panel of Fig. 5. Here the amplitude, ΔK_s, is the difference between the 95th and 5th percentile for the model computed over one period centred on and averaged over each light curve datum. Guided by the sample of M09, we find that the Mira variables lie on a sequence that begins about $P \approx 80$ day with $\Delta K_s \approx 0.5$, runs horizontally to about $P \approx 300$ day before increasing in amplitude with increasing period up to about $\Delta K_s \approx 3$ at $P \approx 1000$ day. We see that the density of stars changes below $\Delta K_s \approx 0.35$ as this is likely the Mira variable boundary and objects with $\Delta K_s \lesssim 0.35$ are semi-regular variables. We adopt the broad selection box shown in Fig. 5 to match the selection of M09. Within this selection box, we find 2876 stars.

We also employ a number of cuts that remove a further ~ 40 per cent of the sample. Contaminants include young stellar objects (YSOs, see Guo et al. 2022), other fainter giant stars and blended photometry not properly handled in our calibration step. We have found that a small fraction of aliases that appear around 1 year period

also have significant parallaxes ϖ measured in VIRAC2. We therefore remove anything with $|\varpi/\sigma_\varpi| > 5$ (186 stars). Furthermore, we employ several cuts based on Wesenheit magnitude (as the Mira variables follow period–luminosity relations), removing stars with $W_{K_s,H} = K_s - 1.328(H - K_s) > 5 \times 10^{-5}(P - 300 \,\text{days})^2 + 8.7$ (826 stars) or $W_{K_s,J} = K_s - 0.482(J - K_s) > 2 \times 10^{-5}(P - 300 \,\text{days})^2 + 8.7$ (744 stars) or $W_{K_s,[4.5]} = K_s - 1.6(K_s - [4.5]) > 10.5 - P/150$ days (475 stars) where the extinction coefficients are from Fritz et al. (2011). In Appendix D and Fig. D1 we display the impact of these cuts. We use the mean from the light curve fits for K_s, the inverse-variance-weighted mean magnitudes for J and H and the [4.5] GLIMPSE/GALCEN measurement (for GLIMPSE-II this is the average of two epochs separated by six months). These cuts remove potential YSO contaminants unless they are very nearby and bright. We are left with a sample of 1784 stars as shown in the central panel of Fig. 5. We assign the stars a running index and name them 'VIRAC Mira #'. We append to this list the 40 M09 sources with periods that are in VIRAC2 but do not enter our final sample (for reasons discussed in Section 3.2).

2.4 Multi-band light curves

Although the multi-epoch coverage from the VVV survey is primarily in the K_s band, there are also multi-epoch observations available in $ZYJH$. Where available, we further complement the JHK_s epochs with the data from M09 (including for those sources without measured periods in M09) and the multi-epoch data from 2MASS (from the IRSA tables fp_psc and pt_src_rej). The three photometric systems (VISTA, SIRIUS and 2MASS) are slightly different. However, this is only a minor concern as the light curve modelling method can accommodate small magnitude shifts. The WISE satellite initially surveyed every region of the sky over two epochs separated by approximately half a year before exhausting its coolant, and from 2013 was re-purposed for the NEOWISE survey which takes two groups of observations per year in $W1$ and $W2$. We take the observations from the IRSA allwise_p3as_mep and neowiser_p1bs_psd tables selecting high quality detections with moon_mask = 0, saa_sep ≥ 5 and qi_fact ≥ 0.5. The photometry is corrected as per footnote[2].

From VVV, 2MASS, the M09 dataset and WISE, we can construct

---

[3] Using the Akaike or Bayesian information criteria instead results in minor differences in the final sample. Choosing a cut of ΔAIC< -10 results in 1783 stars in the final sample and ΔBIC< -10 results in 1776 stars, compared to our default log-likelihood cut resulting in 1784 stars.
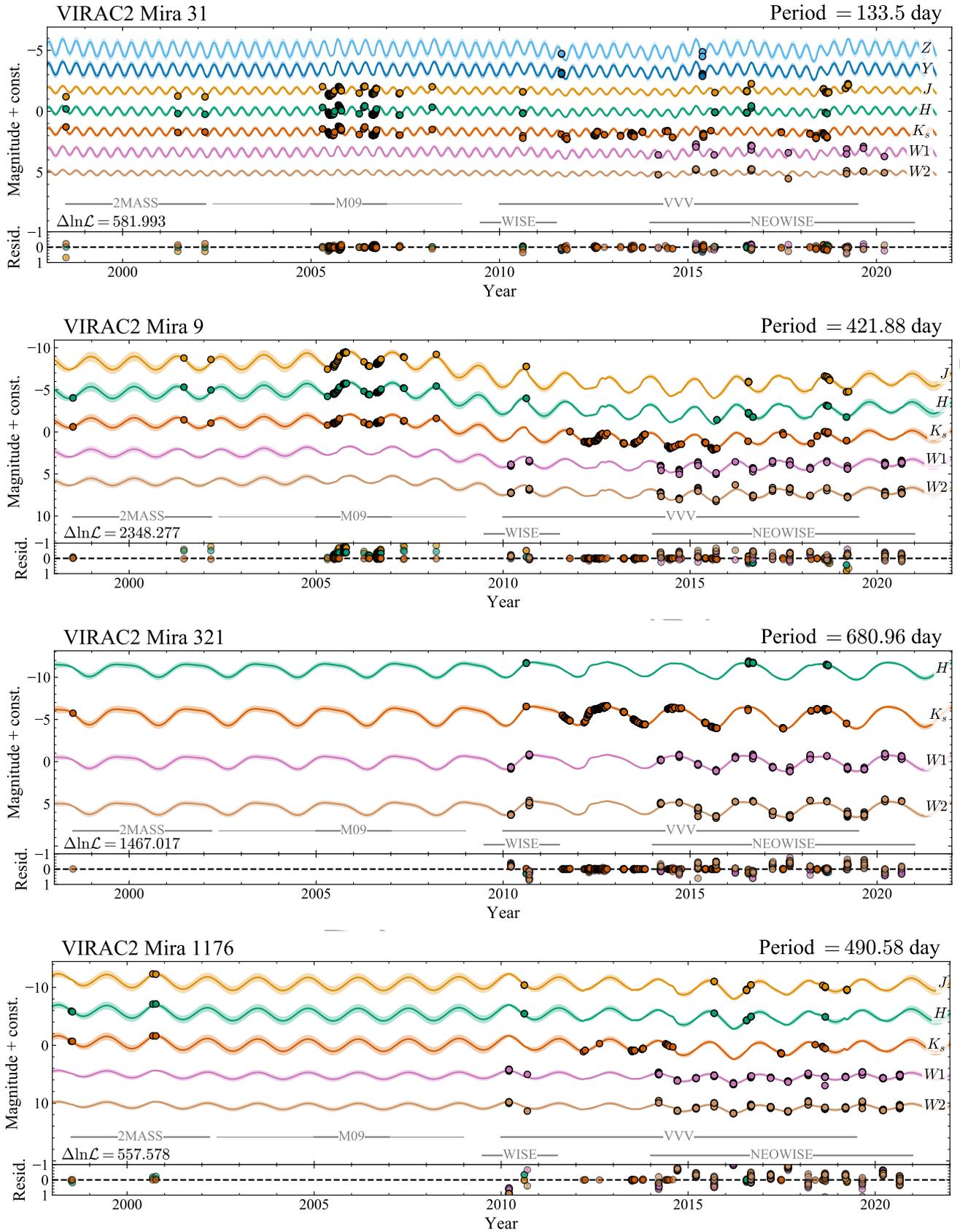
Figure 6. Example multi-band light curves along with a Gaussian process model. The top example has measurements in 7 photometric bands. The second example has observations from Matsunaga et al. (2009). The third and fourth examples are the dense and sparse long-period examples in the second and fourth panels of Fig. 3.
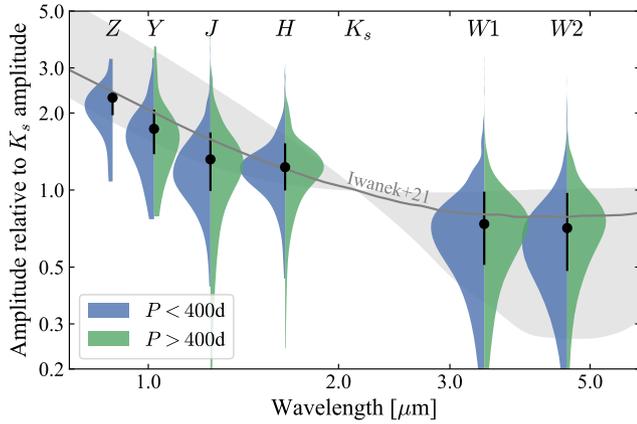
Figure 7. Measured light curve amplitude in the VVV $ZYJH$ bands and WISE $W1$ and $W2$ bands with respect to the K_s band amplitude. The black errorbars show the medians with $\pm 1\sigma$ estimated from the percentiles whilst the violins show the full distributions split by period (left blue < 400 day, right green > 400 day). The relation derived by Iwanek et al. (2021a) is shown in grey.

up to 7-band light curves for our sample which we fit using a 2d generalization of the CELERITE Gaussian process model (Foreman-Mackey et al. 2017) as described in Appendix A3. This is particularly useful for measuring the mean magnitudes in each band (properly corrected for phase variation) as well as measuring the amplitudes in each band. We show some of the example multi-band light curves in Fig. 6. We have chosen a case where all seven bands are observed, one where we have data from 2MASS, M09 and VVV, and then two examples from Fig. 3. It is somewhat evident from these examples that the amplitude of variability decreases with increasing wavelength. In Fig. 7, we show the distributions of the amplitudes in each photometric band relative to that in the K_s band. Our sample follows the O-rich relation derived by Iwanek et al. (2021a) which, it should be noted, was used as a weak prior on the amplitude ratios (see Appendix A3).

With the multi-band light curves in hand, we go through a final visual inspection stage of our sample. We assess three aspects: (i) whether the multiband light curve has perhaps produced spurious results (possibly due to contaminated WISE observations) in which case we resort to the results from the single K_s light curve fit, (ii) whether the light curve fits have evidence of some periodicity but no clear Mira-like oscillations in which case we flag the star as 'unreliable' and (iii) whether there is no clear evidence for periodicity in which case we remove the star from our sample. This visual inspection stage removes 42 stars from our sample and flags 91 as unreliable. ∼ 75 per cent of the unreliable stars have K_s amplitudes below 0.4 suggesting they are semi-regular variables (e.g. M09). This procedure suggests the contamination in our full sample is between 5 and 10 per cent.

Our final sample contains 1782 Mira variable candidates of which 1691 are deemed 'reliable', 342 have K_s amplitudes less than 0.4 so are potentially semi-regular variables and 272 stars are in the sample of M09 (209 have periods reported by M09). The catalogue is temporarily available at https://www.homepages.ucl.ac.uk/~ucapjls/data/mira_vvv.fits.

Table 1. Period–luminosity relations adopted for Fig. 9 and for estimating the distance to our Mira variable sample. The period–luminosity relations are $M = a + b(\log_{10} P - 2.3)$ for $\log_{10} P < 2.6$ and $M = a + 0.3b + c(\log_{10} P - 2.6)$ for $\log_{10} P > 2.6$ where P is the period in days. e gives the extinction coefficient $A_{K_s}/E(i - K_s)$ for the ith band used for computing the Wesenheit magnitudes.

System	Band	a	b	c	e
MW	J	−5.73	−3.45	−3.87	+0.483
MW	H	−6.46	−3.69	−4.75	+1.345
MW	K_s	−6.89	−4.04	−5.86	-
LMC	J	−5.93	−3.35	−6.70	+0.483
LMC	H	−6.67	−3.49	−6.81	+1.345
LMC	K_s	−7.01	−3.79	−6.93	–
LMC	[3.6]	−7.40	−4.05	−7.11	−2.180
LMC	[4.5]	−7.51	−3.85	−7.35	−1.660
LMC	[5.8]	−7.68	−3.80	−7.63	−1.522
LMC	[8.0]	−7.86	−3.87	−8.51	−1.900

3 PROPERTIES OF OUR MIRA VARIABLE SAMPLE

With the Mira variable sample in hand, we now turn to inspecting some of its properties.

3.1 Are they nuclear stellar disc members?

One crucial question regarding the presented sample is whether in fact the Mira variables are genuinely part of the NSD. As a comparison sample, we take all reliable VIRAC2 stars with unextincted K_s (denoted K_{s0}) between 11.2 and 11.9 (encompassing the asymptotic giant branch bump at the Galactic Centre distance). We use the $E(H - [4.5])$ interstellar extinction maps from Sanders et al. (2022). From Fig. 8, it is clear that the on-sky distribution of the Mira variable sample is flattened as per the comparison sample. However, as we will discuss shortly, this is in large parts due to the larger in-plane extinction that makes Mira variables faint enough for reliable detection in VVV. We can see from Fig. 8 that very approximately the dust is mostly a function of Galactic latitude. In the small inset in the central panel of Fig. 8 we display the absolute Galactic longitude (with respect to the location of Sgr A*) of our sample with $|b| < 0.4$ deg and amplitude > 0.4 along with the equivalent distribution of OGLE Mira variables with 2.5 deg $< |b| < 5$ deg (Soszyński et al. 2013). We see that our sample is more centrally concentrated than the nearly flat OGLE 'bar-bulge' distribution and has an exponential scalelength of 130 pc (assuming the population is at 8.275 kpc Gravity Collaboration et al. 2021). This is similar to the scalelength of ∼ 90 pc found in the models of Sormani et al. (2022) although we have not attempted to separate out the bar-bulge contamination. In the models of Sormani et al. (2022), the projected densities of the NSD and the background bar/bulge are equal along an elliptical contour intersecting $(|\ell|, |b|) \approx (1.5, 0)$ deg and $(|\ell|, |b|) \approx (0, 0.4)$ deg. The NSD is then dominant for smaller $|\ell|$ and $|b|$ increasing to around 80 per cent of the total projected density around the Galactic centre (see table 2 of Sormani et al. 2022). This implies that on-sky location is only a weak indicator of NSD membership.

A further check of the Mira variable candidates' possible NSD membership is if their distance distribution coincides with the Galactic centre distance of ∼ 8.275 kpc from Gravity Collaboration et al. (2021). Again the background bar/bulge density distribution is also expected to peak here but will have a higher line-of-sight dispersion. In Fig. 9 we show the Wesenheit magnitudes $(W_{Ks,x} = K_s - A_{K_s}/(A_x - A_{K_s})(x - K_s) = x - A_x/(A_x - A_{K_s})(x - K_s)$ for band x where A_x is the extinction in this band) computed using
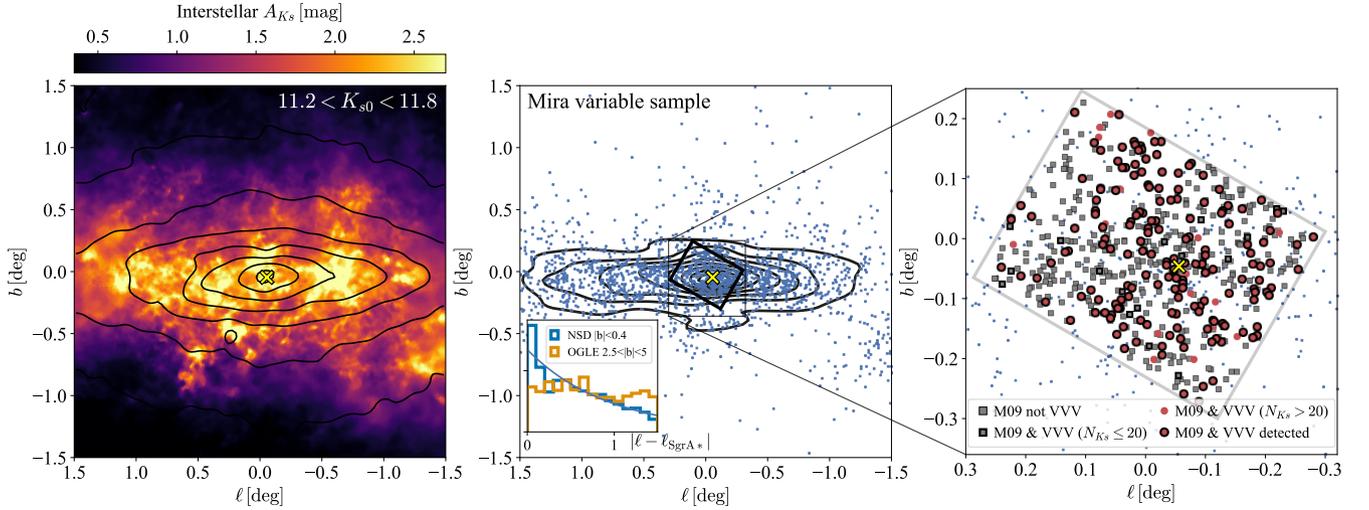
Figure 8. Nuclear stellar disc traced by Mira variables: the *left panel* shows the K_s interstellar extinction overlaid with the logarithmically-spaced density contours for stars with $11.2 < K_{s0} < 11.8$ mag which exhibit a disc morphology. The *central panel* shows the VVV Mira variable sample (including both reliable and unreliable identifications) with their corresponding density given by the contours. This sample also traces the NSD and approximately follows the distribution in the left panel but is subject to selection effects (see Section 3.2). The inset shows the distribution in absolute Galactic longitude (with respect to the location of Sgr A*) of our sample with amplitude > 0.4 and $|b| < 0.4$ deg in blue and the OGLE Mira variables between 2.5 deg $< |b| < 5$ deg from Soszyński et al. (2013) in orange. The line shows an exponential fit with scalelength 130 pc. In the *right panel* we show a zoom-in of the region surveyed by M09. Grey squares show Mira variables with periods from M09 not in VVV, grey squares outlined in black those with periods in M09 and VVV but with 20 or fewer epochs, red circles those with periods in M09 and VVV with more than 20 epochs but not detected picked up by our search, and finally red circles outlined in black are those in our final catalogue that are also in M09 with periods.
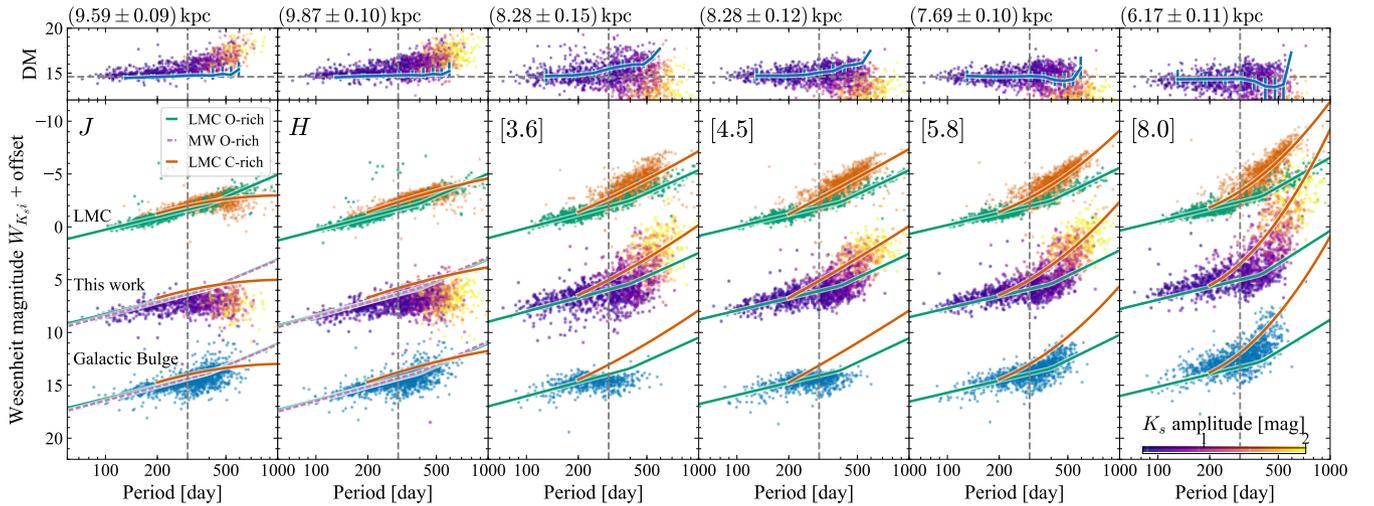

Figure 9. Period–luminosity relations for LMC Mira variables (green O-rich, orange C-rich), inner bulge Mira variables (blue) and our Mira variable sample (coloured by amplitude and limited to reliable sources with $\Delta K_s > 0.4$). In all panels we show the Wesenheit magnitudes $W_{Ks,i} = K_s - A_{Ks}/(A_i - A_{Ks})(i - K_s)$ to account for extinction (the LMC row uses the Wang & Chen 2019 coefficients whilst the other rows use the Fritz et al. 2011 extinction law applied to the O-rich model spectra described in Section 3.3). The green and orange lines are the fits to the LMC O-rich and C-rich populations and the pink-dashed lines are fits to solar neighbourhood O-rich Mira variables (Sanders, in prep., not shown in LMC row as they overlap the LMC relation). The small top panels show the distance modulus of our sample computed using the O-rich period–luminosity relations (MW for JHK_s and LMC otherwise, with the number above the plot reporting the median distance for stars with periods less than 300 days) and the blue errorbars show similar for the inner bulge sample. Note the truncation in the [3.6] and [4.5] distributions for the inner bulge sample is due to saturation.
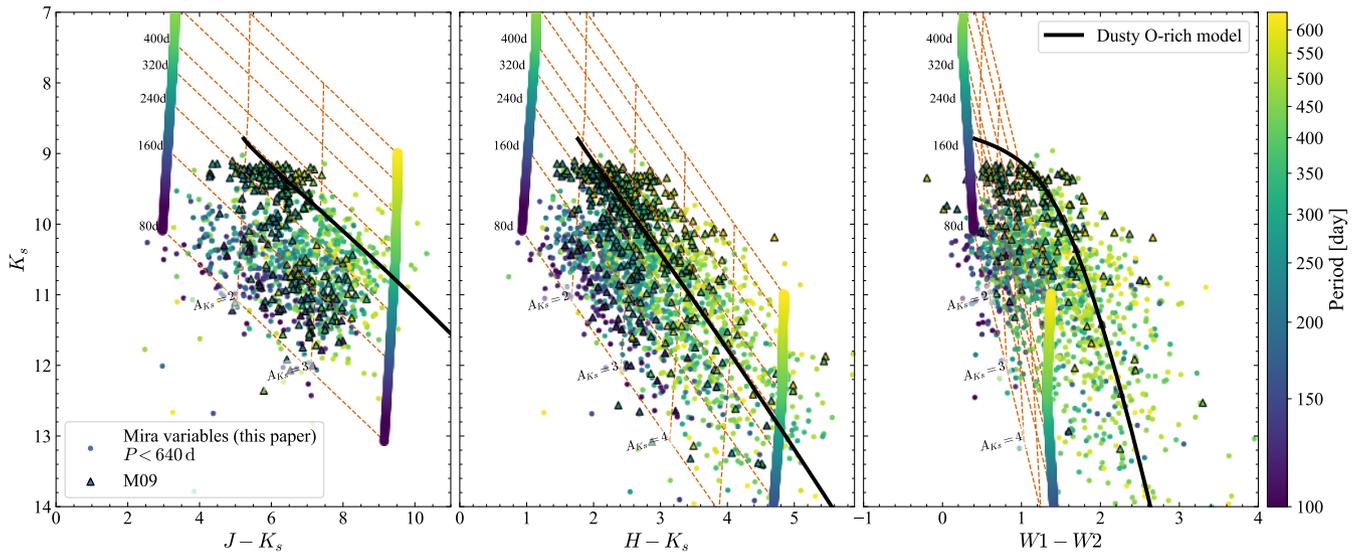
Figure 10. Colour–magnitude diagrams of the Mira variable sample presented in this work (dots, restricting to reliable sources with $\Delta K_s > 0.4$ and periods less than 640 day) with the sample from M09 shown in black outlined triangles (we use the matched VIRAC2 photometry for this sample where the SIRIUS photometry is unavailable). Both samples are coloured by the period. The mesh shows lines of constant period (diagonal) assuming the O-rich LMC period–luminosity relations from Table 1 and a Galactic Centre distance, and lines of constant extinction (vertical) assuming the extinction law from Fritz et al. (2011). The black line shows a sequence of dusty O-rich models of increasing opacity.

the (mean) VVV and GLIMPSE bands and using extinction coefficients, A_x/A_{Ks}, found from the median of a grid of O-rich AGB spectra (described in Section 3.3) combined with the extinction curve from Fritz et al. (2011). This accounts for the potential non-linearity of the extinction coefficients by using an estimate of the K_s band extinction from 2d $E(H - [4.5])$ interstellar extinction maps (Sanders et al. 2022). Wesenheit magnitudes are useful reddening independent measures that can account for reddening from both interstellar and circumstellar dust. However, the choice of the coefficient is key and any misalignment between the interstellar and circumstellar extinction vectors will give rise some spread in the Wesenheit magnitudes (Ita & Matsunaga 2011). In Table 1 we report the extinction coefficients $A_{Ks}/(A_x - A_{Ks}) = A_{Ks}/E(i - K_s)$ at the median A_{Ks} averaged over the model grid. In Fig. 9 we also compare with Mira variables in the inner Galactic bulge ($|\ell| < 2.5$ deg and $|b| < 2.5$ deg from OGLE and Gaia, Soszyński et al. 2013; Mowlavi et al. 2018) using the G-band amplitude cut from Grady et al. (2019, see Appendix C) and the LMC Mira variables from Soszyński et al. (2009) (again using the G-band amplitude cut).

Given the Wesenheit magnitude, we compute the distance using O-rich period–luminosity relations as given in Table 1 (Sanders, in prep.). For the Spitzer bands, we consider period–luminosity relations derived for the LMC whilst for the VVV bands we use relations derived for the solar neighbourhood (transformed from the 2MASS system to the VVV system, González-Fernández et al. 2018) from a sample of Gaia DR2 O-rich Mira variables defined using the G-band amplitude cut (Mowlavi et al. 2018; Watson et al. 2006, Sanders, in prep.). The distance modulus is shown in the top row in Fig. 9. The short period Mira variables trace a tight period–luminosity relation that is in approximate agreement with the expectation that the Mira variables are situated at the expected Galactic Centre distance and have a similar period–luminosity relation to the solar-neighbourhood

and LMC O-rich Mira variables. There is some discrepancy in the derived median distance (top panels of Fig. 9), particularly for J, H and [8.0] that could reflect shortcomings in the extinction correction (due to the high extinction for the NSD Mira variables, the Wesenheit magnitudes are quite sensitive to small differences in the coefficients $A_x/(A_x - A_{Ks})$), particularly as there is very little observed variation in the near-infrared period–luminosity relations for O-rich Mira variables (Whitelock et al. 2008; Goldman et al. 2019, Sanders, in prep.) although M09 find quite different period–luminosity relations in the Spitzer bands than Ita & Matsunaga (2011) do for the LMC data. In particular, the [8.0] extinction coefficient is quite unreliable as it is very sensitive to the source spectrum and the total extinction. Furthermore, the larger scatter about the period–luminosity relations could also reflect the increased presence of circumstellar dust for these (potentially significantly more metal-rich) stars.

For a different projection of the data, we display three colour–magnitude diagrams in Fig. 10 compared to a grid of LMC period–luminosity relations (Table 1, assuming the distance to the Galactic Centre and using the GLIMPSE bands as approximations of the WISE photometry) and lines of constant extinction (Fritz et al. 2011). No extinction correction has been applied to the data. We see again the sample is approximately consistent with being at the Galactic Centre distance behind approximately $A_{Ks} = 2 - 4$ magnitudes of extinction (which could be a combination of both interstellar and circumstellar dust). The WISE colour–magnitude diagram does not trace the extinction vector due to the increased dustiness of the Mira variables at long periods i.e. the misalignment of the interstellar and circumstellar extinction vectors.

Further evidence for the NSD membership of the presented sample can be found through their kinematics. We have proper motion data for all stars in the sample, which we will analyse in a follow-up to this paper. However, there is a subsample of our stars with maser
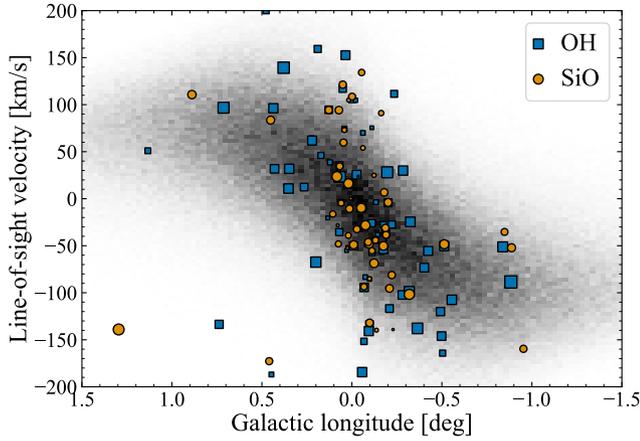
Figure 11. Line-of-sight velocities for our reliable Mira variable sample ($|b| < 0.4$ deg, $\Delta K_s > 0.4$) from maser observations. The background greyscale shows the density (linearly-scaled) from the Sormani et al. (2022) dynamical model (only the NSD component). The blue points show the OH maser observations and orange the SiO. The point size scales linearly with the period of each object.

observations (Engels & Bunzel 2015 for OH masers and Messineo et al. 2002, 2004, Deguchi et al. 2004 and Fujii et al. 2006 for SiO masers). We show the ℓ–v_{los} distribution of the stars with $\Delta K_s > 0.4$ and $|b| < 0.4$ deg compared with the dynamical NSD model from Sormani et al. (2022). In accord with the results of Habing et al. (1983) and Lindqvist et al. (1992), there is clearly net rotation in this sample reaching the ~ 100 km s^{-1} level in line with the dynamical model giving confidence that a high fraction of our long-period sample are indeed NSD members. However, the Mira variables with maser observations are limited to periods $\gtrsim 400$ days. Therefore, based on maser observations alone, we are unable to conclude anything on the NSD membership of the short-period Mira variables.

3.2 The sample selection function

As already highlighted, the presented Mira variable sample appears to be flattened on the sky and so traces the NSD. However, this effect is driven in part by a combination of observational limitations (e.g. arising from the magnitude range of the VVV survey) and the distribution of interstellar dust towards the NSD. VVV begins to saturate around $K_s \approx 11.5$. Consider an unreddened Mira variable with period ~ 100 day at the Galactic Centre. This is typically the lowest period reached by Mira variables and hence is the faintest source we might consider. This star will have $K_s \sim 9$ and will hence be too bright for VVV. However, as the reddening towards the Galactic Centre is significant, $A_{K_s} \approx 2$ mag, typically this type of source will have $K_s \sim 11$ and so be just measurable by VVV. In the top right panel of Fig. 12 we display the distribution of our sample in apparent K_s magnitude vs. period. Due to saturation effects, we have very few stars with $K_s \lesssim 10$. Comparison with the period-luminosity relations from Table 1 for a star at the Galactic centre distance reddened by different amounts demonstrates clearly the necessity of some reddening to make the Mira sample faint enough for reliable VVV observations. As we will discuss further, this can be due to either interstellar *or* circumstellar extinction.

We can assess the effect of incompleteness on our sample more quantitatively using the completeness calculations presented by

Sanders et al. (2022). These were based on artificial star tests requiring that the artificial stars are recovered in 20 per cent of the observations. An example of a typical result from this calculation is shown in the top left panel of Fig. 12. Combining these calculations with the 2d $E(H - [4.5])$ interstellar extinction map presented in Sanders et al. (2022), we find the on-sky completeness as shown in the lower left panel of Fig. 12. Clearly, we are only nearly complete in the high-density regions near the mid-plane. The completeness map lines up very well with the on-sky source density.

In the lower right panel of Fig. 12 we show the absolute K_s magnitudes computed assuming a distance of 8.275 kpc (Gravity Collaboration et al. 2021) and using extinctions derived in the next subsection (3.4). These extinction estimates only account for the interstellar extinction and are biased by our prior estimates from the 2d $E(H - [4.5])$ interstellar extinction map. We note that in this projection a high fraction of our sample are too faint to be consistent with the LMC/solar-neighbourhood period–luminosity relation despite the results of Fig. 9. Using these extinction estimates then suggests that the objects are more distant than the NSD. However, another explanation is that the average extinction estimates we are using here are inappropriate for our sample which may be more dust-obscured than the average bulge giant star in each region of the sky. The extinction varies significantly over very small scales in this part of the sky and we are biased towards fainter objects. It could also be that in high extinction regions we are losing the bulge/NSD red giants in $(H - [4.5])$ as they are too faint and the resulting extinctions are biased towards the extinction of more foreground objects (as discussed in Sanders et al. 2022, in the comparison of $E(H - [4.5])$ from giant branch stars compared with $E(H - K_s)$ from red clump stars). The faintness may also arise due to the stars having more circumstellar dust than expected. This may be the case at the long-period end ($\gtrsim 400$ day) but such dust appears comparatively rare in short-period objects (Ita & Matsunaga 2011, and subsection 3.4). Other effects, such as metallicity variation, may give rise to intrinsic differences in Mira variables independent of the dust properties although there is little evidence for this being such a significant effect in the near-infrared (Whitelock et al. 2008). Finally, as already noted, we are working at the bright limits of VVV and saturation could cause stars to be fainter than they are.

To test this discrepancy further, we follow Matsunaga et al. (2009) and estimate the total colour excess with respect to a period-colour relation (from Table 1). This estimate includes contributions from both interstellar and circumstellar extinction (or more precisely any additional circumstellar extinction relative to the reference population for the period-colour calibration). The points in the top right panel of Fig. 12 are coloured by the A_{K_s} implied by the $(K_s - [4.5])$ colour excess. For the bluer objects this colour excess estimate agrees with the 2d $E(H - [4.5])$ extinction map but for most sources it is higher (as seen approximately by comparing the colourbars of the upper and lower right panels of Fig. 12), even for short-period objects. Therefore, there is evidence that the 2d $E(H-[4.5])$ interstellar extinction map produces underestimates to these objects. For longer-period objects it is more difficult to say what is happening as it requires disentangling the interstellar and circumstellar extinction. These stars have evidence of significant circumstellar extinction but they are also younger objects, probably embedded in more dusty environments. We note that this mismatch between the Mira variable colour excesses and the interstellar extinction maps was also found by Nikzat et al. (2022) who attributed it to the Mira variables being at larger distances (possibly in the background disc) and so behind more dust than the bulge stars. This may well be the case for their
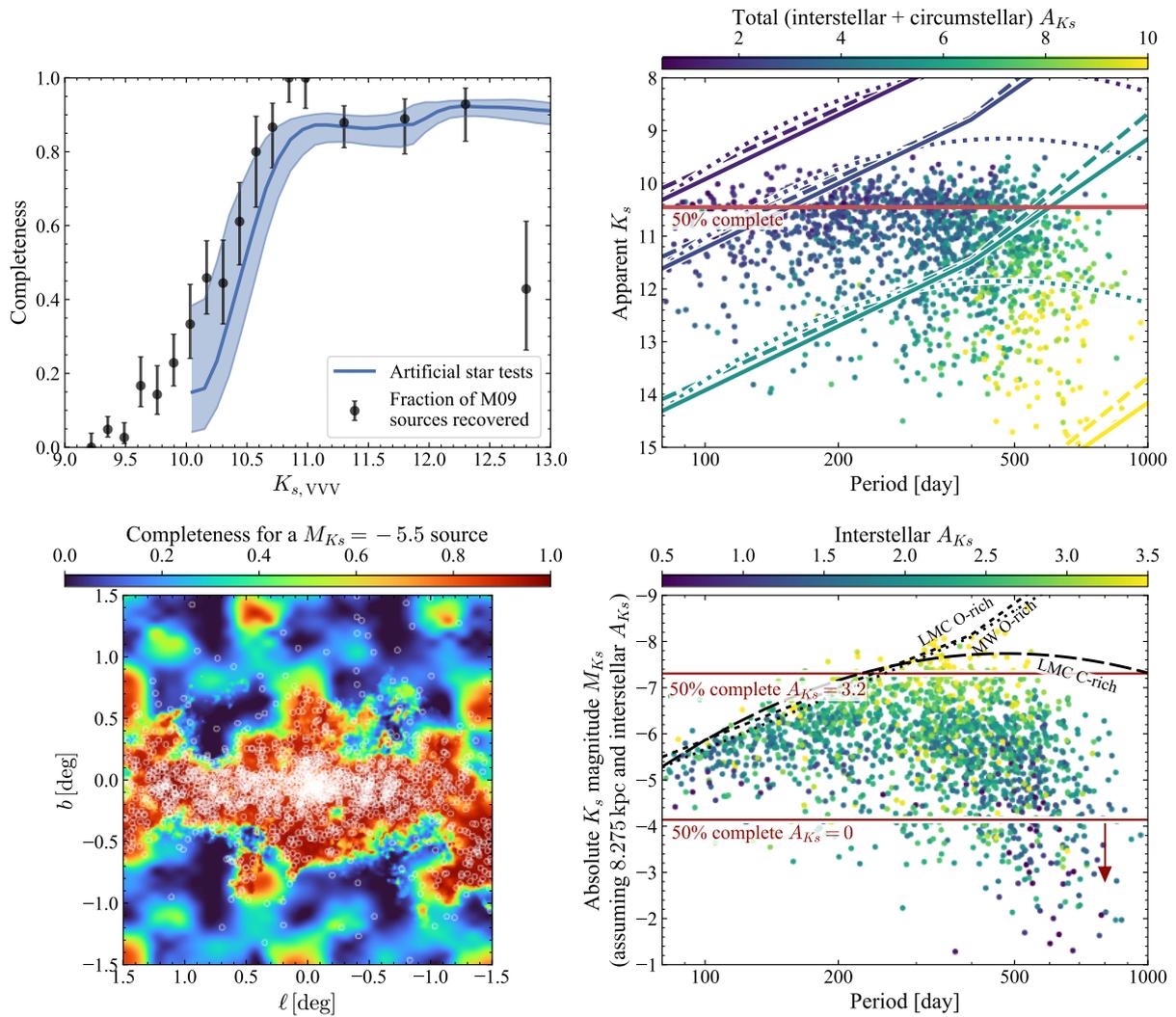
Figure 12. Selection effects for our sample: in the top left panel we show in blue a comparison of the completeness computed from artificial star tests (the line and shaded bracket give the median and $\pm 1\sigma$ over the region surveyed by M09) and the fraction of sources with periods (predominantly Mira variables) in M09 recovered by this work. The top right panel shows the apparent K_s magnitude vs. period for our reliable sample coloured by their extinction estimated from the colour excess with respect to the $K_s - [4.5]$ period colour relation (a proxy for the interstellar *plus* circumstellar extinction). The horizontal red line shows the 50 per cent completeness from the top left panel. Period-luminosity relations (as labelled in the lower right panel) are shown for stars at the Galactic centre distance extincted by different amounts ($A_{Ks} = 1, 2.3, 5, 10$ according to their colours). The bottom right panel shows the same sample in absolute magnitude vs. period (assuming all sources are at the Galactic Centre distance – more distant sources could also enter the selection). The K_s magnitudes are dereddened using solely the interstellar extinction estimates. The horizontal lines show the 50 per cent completeness limit assuming zero and 3.2 magnitudes (the 90th percentile of the sample) of extinction. The bottom left panel shows the completeness for sources with absolute magnitude of -5.5 with the sample overlaid in white. Note the correlation between the sample and the completeness map.

higher latitude sample, but Fig. 9 demonstrates that the distances of our stars are consistent with NSD/bulge membership.

Returning to what the extinction means for the completeness of our sample, we have marked with horizontal lines the values at which samples with different amounts of interstellar reddening would be 50 per cent complete i.e. we would typically see all stars fainter than these limits. We see that our sample truncates at around the limit for the 90th percentile of extinction $A_{Ks} = 3.2$ for our sample. We have no way to probe sources intrinsically brighter than this. Typically these would be the longer period sources, but we notice that many of the long-period sources actually fall significantly under the expected period–luminosity relations (either as a result of circumstellar dust or underestimated interstellar extinction) as already highlighted above. This is fortuitous for our purposes.

Finally, as a validation of the selection effects affecting the catalogue, the left panel of Fig. 12 compares the fraction of M09 sources with reported periods we recover together with the expectation from the artificial star completeness tests. We have approximated $K_{s,VVV}$ as the mean SIRIUS K_s magnitudes from M09 which are ~ 0.05 fainter than the K_s flux-means reported by M09. The artificial star tests are for constant sources and no adjustment has been made to consider how an artificial variable source might be recovered. However, the high degree of correspondence suggests our catalogue is as complete as can be expected given the quality of the photometry, and our algorithms for selecting and processing variable stars have not artificially removed any genuine Mira variables. Of the 549 M09 long period variables with periods we detect 169 (~ 30 per cent). 212 are in VIRAC2 but only 192 have more than the 20 epochs re-

quired to enter our high-quality catalogue (the missing 20 all satisfy the Wesenheit period cuts so enter the combined catalogue). Of the remaining 23 with a sufficient number of epochs but that we failed to detect, 3 fail our Wesenheit-period cuts (hence why we end up with an additional 40 sources in the combined catalogue), 2 stars are missed by the initial candidate selection (visible in Fig. 1), and then 7 fail the period cuts to be further processed using the Gaussian process model. 6 of the remaining 11 stars fall outside the period–amplitude selection (as shown in Fig. 5) and the other 5 fail the parallax and Wesenheit-period cuts.

Although the completeness is well understood, the contamination of the catalogue is much more difficult to assess. Through visual inspection we assessed the contamination at the 5 to 10 per cent level before removal of potential contaminants. One potential contaminant is young stellar objects. Our cuts on Wesenheit magnitude have sought to eliminate these intrinsically dimmer contaminants. We have checked whether any of our objects are in the young stellar object catalogues of Guo et al. (2021) and Guo et al. (2022). One source in our catalogue (VIRAC Mira 68) has been identified by Guo et al. (2022, VVV_PB_122) as a potential periodic outbursting young stellar object. This star is also in the M09 catalogue with an associated period and K_s amplitude of 0.55. It is one of the brightest objects in the Guo et al. (2022) catalogue so we therefore think it likely that it is a long-period variable but it is unclear.

With these considerations, there is significant scope for improving the completeness of the Mira variable NSD sample. It may be possible to utilise $(Z, Y, J, H, W1, W2)$ photometry without the reliance on K_s – unfortunately in the VVV reduction we are using a reliable K_s detection is a necessary requirement although this isn't a fundamental limitation. Furthermore, the presence of dark lanes in infrared images suggests we may be missing some Mira variables embedded in or shrouded by such thick dust that our searches need to go fainter to probe the full Mira variable population. In future, the PRime-focus Infrared Microlensing Experiment (PRIME)[4] and the JASMINE satellite (Gouda & JASMINE Team 2020)[5] are expected to provide better coverage of this area particularly at the brighter magnitude end.

3.3 Photometric characterisation of our sample

AGB stars are typically classified by their C and O compositions, with C-rich C stars having $[C/O] > 1$, O-rich M stars having $[C/O] < 1$ and the S stars having $[C/O] \approx 1$ (Höfner & Olofsson 2018). The $[C/O]$ ratio governs the dominant molecular species in the atmospheres and dusty circumstellar envelopes of AGB stars (e.g. silicates for O-rich and carbonaceous species, e.g. C_2 or C, for C-rich), and, as such, governs the observed colours, and the mass loss rate of AGB stars. The observed $[C/O]$ ratio is determined by the third dredge-up, which mixes more central C-rich material into the outer envelope, and is a function of both mass and metallicity of the star. Typically C-rich AGB stars are associated with younger and/or more metal-poor systems. For this reason, we typically find the bulge of the Milky Way occupied by O-rich Mira variables, whilst the outer disc has a greater number of C-rich stars (Blanco et al. 1984). Until the recent discovery of 5 C-rich Mira variables in the bulge by Matsunaga et al. (2017), it was believed the bulge consisted almost solely of O-rich Mira variables. However, it is not clear whether these stars are associated with younger or more metal-poor populations, or are
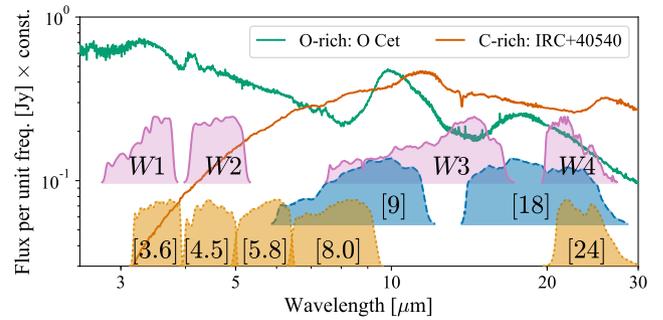
---

[4] http://www-ir.ess.sci.osaka-u.ac.jp/prime/index.html
[5] http://jasmine.nao.ac.jp/index-en.html

Figure 13. Mid-infrared spectra of typical O-rich (green) and C-rich (orange) Mira variables (o Cet – Mira itself, and IRC+40540) from Sloan et al. (2003). Also displayed are the relative spectral response functions for WISE (pink), AKARI (blue) and Spitzer IRAC/MIPS (orange). Note the strong silicate features in the O-rich spectrum at 10 and 18 μm and the SiC feature at 11 μm in the C-rich spectrum.

the result of binary evolution. It is important to classify the types of Mira variables we are dealing with, not just because it gives a reflection of the age and metallicity of population they belong to, but also because O-rich and C-rich Mira variables follow quite different period–luminosity relations.

For comparison to our Mira variable sample, we have run two grids of dusty AGB models, one O-rich and one C-rich. We utilize the DUSTY code (Ivezic & Elitzur 1997) using the analytic radiatively-driven wind solution. The source spectra are taken from the synthetic libraries of Aringer et al. (2016) and Aringer et al. (2019) where we use the solar mass models with $[C/O] = 0.55$, solar metallicity $Z = Z_\odot$ and $\log g = 0$ for the O-rich sources and the solar mass models with $[C/O] = 1.1$, $Z = Z_\odot$ and $\log g = -0.4$ for C-rich. These choices are similar to those of Lian et al. (2014) but as discussed by Aringer et al. (2009) and Kučinskas et al. (2005) the variation of the broadband infrared colours with mass, surface gravity and $[C/O]$ is weak. For O-rich circumstellar dust we take the O-rich interstellar warm silicate optical constants from Suh (1999) and for C-rich we take the amorphous carbon optical constants from Suh (2000). We run grids of models parametrized by the opacity at 10 μm, τ_{10}, (up to a maximum of $\tau_{10} = 10$ for the O-rich models and $\tau_{10} = 0.5$ for the C-rich), inner dust temperature T_{in} (between 600 K and 1400 K for O-rich and 600 K and 1800 K for C-rich) and the effective temperature T_{eff} of the source spectrum (see Goldman 2020, for other model grids). The radiatively-driven wind solution is self-similar (Elitzur & Ivezic 2001) allowing for simple rescaling for different central luminosities and gas-to-dust ratios. We have displayed the sequence of models with $T_{eff} = 2600$ K and $T_{in} = 600$ K in Fig. 10. We see that the circumstellar dust acts like the interstellar extinction vector in the $(J - K_s)$ and $(H - K_s)$ vs. K_s colour–magnitude diagrams. However, in $(W1 - W2)$ the models are aligned quite differently to the interstellar extinction vector and indeed it appears that the dusty models are necessary to match the data distribution.

As a further comparison to our sample, a large catalogue of O- and C-rich AGB stars was presented by Suh & Hong (2017) who compiled classifications from the literature which were primarily based upon maser observations (e.g. OH masers are associated with an O-rich AGB star) or low resolution spectroscopy (see Suh 2021, for a significantly expanded, more recent catalogue). As discussed in Suh (2021), the two types can also be more approximately separated with photometry. Colour selections are most effective in the near- and mid-infrared where there are a number of molecular features. In
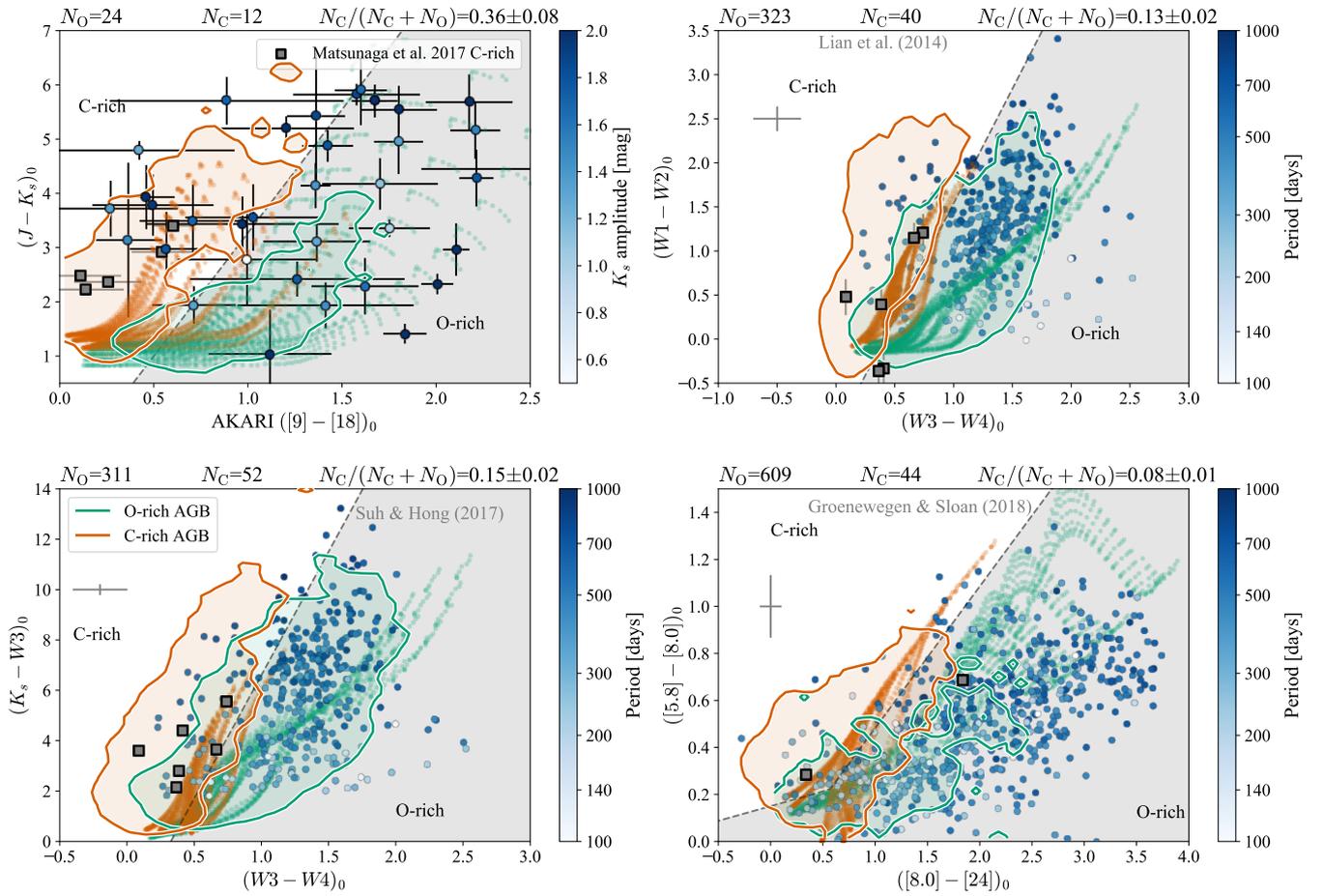
Figure 14. Compositions of the reliable Mira variables: four colour-colour diagrams in which O-rich and C-rich AGB stars separate. The O-rich silicate feature at $10\,\mu m$ is covered by AKARI [9], WISE $W3$ and GLIMPSE [8.0] allowing for separation of O- and C-rich. Blue circular points show the Mira variable sample presented here (90 per cent of the sample have uncertainties smaller than the displayed grey errorbars), grey squares are the OGLE C-rich Mira variables from Matsunaga et al. (2017), the green and orange shaded regions show the extent of the O-rich and C-rich AGB samples from Suh & Hong (2017), except for the lower right panel where Mira variables in the LMC are shown) and the green and orange points are theoretical models of dusty O-rich and C-rich AGB stars. The *top left* panel shows the distribution of the sample in dereddened $(J - K_s)$ against dereddened $([9] - [18])$ from AKARI coloured by amplitude. We show the line defined by Ishihara et al. (2011) to separate C-rich and O-rich and display the six OGLE Mira classified as C-rich by Matsunaga et al. (2017). The *top right* panel shows the distribution in dereddened WISE colours, as suggested by Lian et al. (2014), now coloured by period. The *bottom left* panel shows similar but for a combination of K_s and WISE magnitudes similar to that suggested by Suh & Hong (2017). The *bottom right* panel shows a combination of Spitzer IRAC/MIPS colours used by Groenewegen & Sloan (2018). Above each panel we display the number of Mira classified as C- and O-rich. From this, we estimate that $\lesssim 10$ per cent of our sample are C-rich Mira.

Fig. 13 we show the mid-infrared spectra of typical O-rich and C-rich Mira variables from the catalogue of Sloan et al. (2003). Several authors have suggested colour combinations in which the C-rich and O-rich Mira variables separate, which mostly rely on the strong silicate feature at $10\,\mu m$ which is covered by the AKARI [9] band and WISE $W3$. Lebzelter et al. (2018) demonstrated that a combination of Gaia and 2MASS colours could be employed to effectively separate Mira variables in the LMC. Unfortunately, for our purposes, Gaia photometry is unavailable for all but a handful of our Mira variables due to high extinction. Here we explore the following four near- and mid-infrared colour-colour selections as shown in Fig. 14:

(i) Ishihara et al. (2011) showed how AKARI photometric bands [9] and [18] covered the O-rich silicate features at 10 and $18\,\mu m$. We employ the selection in $(J - K_s)$ against $([9] - [18])$ where C-rich stars have redder $([9] - [18])$ at fixed $(J - K_s)$. This is colour-colour combination in which Matsunaga et al. (2017) searched for C-rich

Mira variables in the bulge. The line separating O- and C-rich is

$$(J - K_s) = 4.55([9] - [18]) - 1.27. \qquad (1)$$

(ii) Both Lian et al. (2014) and Nikutta et al. (2014) demonstrated that O-rich have redder $(W3 - W4)$ at fixed $(W1 - W2)$ than C-rich. The line separating O- and C-rich is

$$(W1 - W2) = 2.35(W3 - W4) - 0.84. \qquad (2)$$

(iii) In a very similar vein, Suh & Hong (2017) show using IRAS photometry the populations separate in $(K_s - [12])$ against $([12] - [25])$, again with O-rich redder in $([12] - [25])$ than C-rich. [12] and [25] are similar to $W3$ and $W4$ respectively. The line separating O- and C-rich is

$$(W3 - W4) = 0.102(K_s - W3) + 0.427. \qquad (3)$$

(iv) Groenewegen & Sloan (2018) demonstrated the clear separation of O-rich and C-rich AGB stars in the $([5.8] - [8.0])$ against
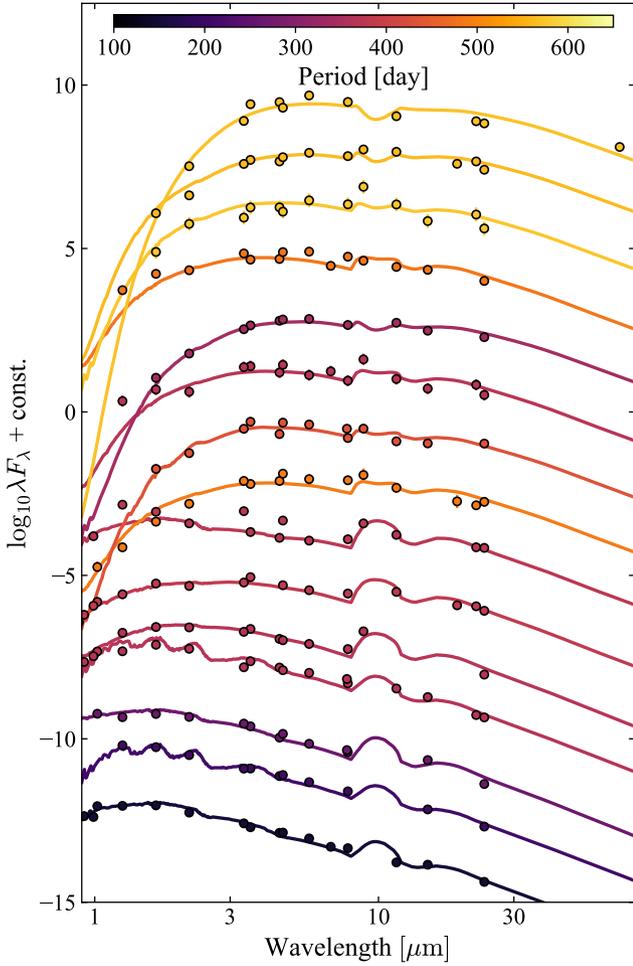
Figure 15. Multiband photometry fit with O-rich dusty AGB models. The stars are ordered such that their derived mass loss rates (and hence approximately the optical depths) increase upwards and are coloured by their period. The points are photometric measurements corrected by the best-fit interstellar extinction.

([8.0] − [24]) plane where [24] is from the MIPS instrument on Spitzer (see also Kastner et al. 2008). The main separating line is

$$([5.8] - [8.0]) = 0.603([8.0] - [24]) - 0.121, \quad (4)$$

although there is a slight break at the blue end.

All of the colour-colour selections rely on dereddened photometry. We use extinction coefficients computed as the median over the O-rich model grid at each star's total interstellar extinction using the extinction law from Fritz et al. (2011) and the total extinction is derived from the $H - [4.5]$ excess of all giant stars (Majewski et al. 2011; Sanders et al. 2022). Using the (interstellar *plus* circumstellar) extinction estimated from the period-colour relations (Matsunaga et al. 2009) does not significantly alter the results. There are two caveats when using WISE photometry: (i) AGB stars are bright in WISE bands so likely to be saturated. However, reliable magnitudes are still extracted for sources fainter than $(W1, W2, W3, W4) = (2, 1.5, -3, -4)$ provided we correct the photometry as per footnote[2]; (ii) the $W3$ and $W4$ angular resolutions are 6.3 and 12.0 arcsec respectively. In the highly-crowded Galactic centre regions it is likely $W4$ measurements are contaminated.

We have found that sources with $W4$ reduced chi-squared > 5 are likely blended as they lie at significantly higher magnitudes than expected from comparison with both models and other data samples. We therefore adopt this as a quality cut.

In Fig. 14 we show our Mira variable sample in the four highlighted colour-colour spaces. We also overplot the Suh & Hong (2017) AGB stars (those with WISE measurements greater than the previously quoted bright limits) and the two families of dusty AGB models. We observe that the majority of the sample are consistent with O-rich chemistry. Simply counting the number of stars in each region of the plots, we find (36 ± 8) per cent, (13 ± 2) per cent, (15 ± 2) per cent and (8 ± 1) per cent C-rich stars using the cuts (i), (ii), (iii) and (iv). There are very few stars with AKARI photometry and the uncertainties in the photometry are large, so we are more inclined to follow the GLIMPSE- and WISE-based cuts which suggest ∼ 10 per cent C-rich. Note though that the uncertainties in the photometry are significant (mostly coming from extinction uncertainties) and our simple cuts are not perfect. Indeed, we see from the AGB and LMC samples that there is contamination from O-rich within the defined C-rich region. Additionally, contamination in the longer wavelength bands, which might be expected for these crowded fields and large point-spread functions, could artificially move objects to redder colours and may be the cause of the disagreement between the dusty AGB models and the data (although this may also reflect shortcomings of the modelled dust composition). Finally, there is some sensitivity to the choice of extinction coefficients, particularly for [8.0] which is quite sensitive to source spectrum and total extinction, and so possibly explains the small disagreement between the data and the model spectra. In conclusion, we find that the fraction of C-rich Mira variables in our sample is ≲ 10 per cent although it appears there are a few good candidates for genuine C-rich bulge stars as found by Matsunaga et al. (2017).

3.4 Spectral fits

To further characterise the properties of the Mira variable sample, we fit the dusty AGB models described in the previous section to the broadband photometry. Based on the considerations above, we restrict to only considering the O-rich models. For each model we have computed the model flux $\tilde{f} = \lambda F_\lambda$ using the filters provided by the SVO filter service (Rodrigo et al. 2012; Rodrigo & Solano 2020). For each star, we then minimise

$$\sum_{i \in \text{bands}} \frac{(f_i - \tilde{f}_i(T_{\text{eff}}, T_{\text{in}}, \tau_{10}, A_0, \mathcal{N})^2}{\sigma_{fi}^2} + \frac{(A_0 - 0.7E(H - [4.5])^2}{(0.7\sigma_{E(H-[4.5])})^2}, \quad (5)$$

with respect to the model parameters $(T_{\text{eff}}, T_{\text{in}}, \tau_{10}, A_0, \mathcal{N})$. The model grid is parametrized by $(T_{\text{eff}}, T_{\text{in}}, \tau_{10})$: the effective temperature of the star, the temperature of the dust at the inner edge and the optical depth at $10\,\mu\text{m}$. $\mathcal{N}$ is a free normalization that can either be interpreted as distance or luminosity information. We extinct the model spectra using the extinction law from Fritz et al. (2011) normalized by the reddening at $\lambda = 2.149\,\mu\text{m}$, A_0. This procedure accounts for possible non-linearities in the extinction coefficients. A prior is placed on A_0 based on the 2d $E(H - [4.5])$ interstellar extinction maps from Sanders et al. (2022). However, as A_0 is free, it can in theory include contributions from both interstellar *and* circumstellar extinction, although we expect in large part the circumstellar extinction is handled by the dust intrinsic to the models (τ_{10}). Previously we have suggested the interstellar extinction towards these sources is underestimated. Relaxing the prior on A_0 produces degeneracies

with τ_{10} but has little impact on the derived mass loss rates. The data fluxes, f_i, are from VVV (Z, Y, J, H, K_s), DECAPS (z, y), GLIMPSE ([3.6], [4.5], [5.8], [8.0]), WISE ($W1, W2, W3, W4$), ISO ([7], [15]), AKARI ([9], [15]) and MIPS ([24], [70]), and are found using the zeropoints provided by the SVO filter service. The uncertainties σ_{fi} are a quadrature sum of the photometric uncertainties, the amplitude given the measured K_s amplitude and the Iwanek et al. (2021a) O-rich amplitude relation (many of the measurements are based on multi-epoch observations so the scatter is expected to be smaller than this but this will only marginally affect the results) and an error floor term, σ_0, to capture any additional variation. We interpolate our model grid fluxes for each choice of ($T_{\rm eff}, T_{\rm in}, \tau_{10}, A_0, \mathcal{N}$). After an initial fit, we remove any datum that differs by more than 5σ from the best-fit and repeat the fit. In Fig. 15 we show some example fits. The most pronounced spectral feature is the 10 μm silicate feature that transitions from in emission at low opacity to in absorption at high opacity (Suh 2021).

In Fig. 16 we show the distribution of our sample in colour vs. period space and period vs. mass loss rate. The expansion velocity and mass loss rate can be computed from the radiatively-driven wind model through scaling relations (Elitzur & Ivezić 2001) assuming a luminosity and gas-to-dust ratio. Here we assume all stars are located at 8.275 kpc (Gravity Collaboration et al. 2021) to convert the normalization $\mathcal{N}$ into a luminosity and we set the gas-to-dust ratio at $r_{\rm gd} = 45$ (Goldman et al. 2017) such that the expansion velocities approximately match the expansion velocities measured for the OH maser sources ($v_{\rm exp} \propto r_{\rm gd}^{-1/2}$ and $\dot{M} \propto r_{\rm gd}^{1/2}$ if other gas-to-dust ratios are required). Note that this procedure produces large expansion velocities at short period so it is likely in reality $r_{\rm gd}$ increases for shorter period objects. This introduces a factor of a few uncertainty in the mass loss rates. We see that at short period (< 400 day) the dereddened colours are flat with period, and the sources get significantly redder beyond 400 day period due to circumstellar dust (Whitelock et al. 2008). There is a small offset in the $(K_s - [5.8])_0$ colours of our low-period ($\lesssim 400$ day) Mira variables with respect to the O-rich LMC Mira variables possibly related to metallicity effects or underestimates of the interstellar extinction. The derived expansion velocities and mass loss rates have a similar structure to the period–colour diagrams where the distributions are relatively flat for < 400 day (at around the $3 \times 10^{-7} M_\odot$ year^{-1} level for the mass loss) before evolving rapidly for larger periods. The mass loss rate for the longest period Mira variables reaches $\sim 2 \times 10^{-5} M_\odot$ year^{-1}.

3.5 Period and age distribution

We close our investigation of the sample by returning to the main purpose of investigating Mira variables in the NSD – what is their age distribution and what does this tell us about the formation epoch of the Galactic bar? In Fig. 17 we show the period distribution of our sample alongside that of the LMC Mira variables and a reference Galactic bulge sample formed from all Gaia and OGLE Mira variables within a 5 deg×5 deg box centred on the Galactic centre. We observe that the overall shapes of the distributions are quite similar with an increase in Mira variables around 300 days. However, the Mira sample presented here has a broader long and short period tail suggesting more very young and very old stars than the LMC and the inner bulge. The LMC has a slightly broader long period tail than the inner bulge consistent with more recent star formation.

As noted in the introduction, there are relatively few studies of the period–age relation for Mira variables. Most studies are based on empirical period-kinematic calibrations from Feast & Whitelock (1987),

Feast & Whitelock (2000b), Feast et al. (2006, based on C-rich stars), Feast (2009) and Feast & Whitelock (2014) that are approximately calibrated against the age–kinematic relations observed in the solar neighbourhood. Wyatt & Cahn (1983), Feast & Whitelock (1987) and Eggen (1998) provide more theoretical investigations into the period–age relation, and produce relations that agree well with the kinematic calibrations (see Fig. 18). Both López-Corredoira (2017) and Nikzat et al. (2022) have fitted the collection of these calibrations with simple analytic forms (also shown in Fig. 18). There is a more recent theoretical investigation from Trabucchi & Mowlavi (2022) using the non-linear pulsation computations from Trabucchi et al. (2019). Their calibration predicts significantly shorter periods at fixed age than the kinematically-calibrated relations (see Fig. 18). Interestingly, this agrees well with the calibration from Grady et al. (2019) that was based upon Mira variables in LMC and Milky Way clusters. The study of Trabucchi & Mowlavi (2022) also highlights that there is a significant spread in age at each period as the Mira variables undergo thermal pulsations. Utilising this latter relation for stars in the Galactic bulge (Catchpole et al. 2016) would mean the Galactic bulge contains significant populations of stars with ages $\lesssim 1$ Gyr. The bulge is typically considered to be composed of old stars (Zoccali et al. 2003; Bovy et al. 2019; Hasselquist et al. 2020) although there is evidence for a small number of younger (~ 5 Gyr) stars (Bensby et al. 2013; Bernard et al. 2018). Possibly there are complications related to the period–metallicity relation (Feast & Whitelock 2000a). However, some old globular clusters e.g. NGC 5927 ($\gtrsim 10$ Gyr, Dotter et al. 2010; VandenBerg et al. 2013) have Mira variables with ~ 300 day periods (Feast et al. 2002) suggesting < 300 day Mira variables are associated with $\gtrsim 8$ Gyr old populations. Based on these considerations we use the kinematically-calibrated results. In Fig. 18 we display a simple analytic fit to the period–age (P–τ) results of

$$\tau = 13 \, {\rm Gyr} \frac{1}{2} \left(1 + \tanh \left[\frac{330 \, {\rm day} - P}{250 \, {\rm day}} \right] \right). \tag{6}$$

Using this relation, we display the age distribution of our Mira variable sample (restricting to those reliable stars with $|b| < 0.4$ and amplitudes $\Delta K_s > 0.4$), compared to the inner Galactic bulge sample and the LMC O-rich sample. We see that the NSD sample contains significant numbers of long-period stars indicative of recent star formation (Morris & Serabyn 1996). There are also significantly older stars than in the Galactic bulge sample although this may reflect incompleteness in the Gaia/OGLE samples at the short period end due to extinction. An interesting feature of the age distribution is the lack of stars around 8 Gyr. Visually this looks similar to the models of Baba & Kawata (2020, their figure 5) where an old nuclear bulge is included giving rise to an old peak before a younger peak due to the NSD formation. Another interpretation is that the NSD formation is contributing to the older peak and perhaps the younger peak is a secondary burst due to heightened gas accretion, or perhaps due to bar destruction and reformation at an early epoch. Either way, the distribution suggests an old ($\gtrsim 8$ Gyr) bar formation. Our sample is naturally contaminated with foreground Galactic bulge stars (e.g. the models of Sormani et al. 2022, suggest the relative on-sky density of bulge and NSD stars is only $\gtrsim 1$ for $|b| < 0.4$ deg) so it is difficult to conclude which stars are genuine NSD members and we reserve their full investigation to a follow-up publication.

4 CONCLUSIONS

We have described a methodology for discovering Mira variable stars from the VVV multi-epoch infrared data and presented a sample of
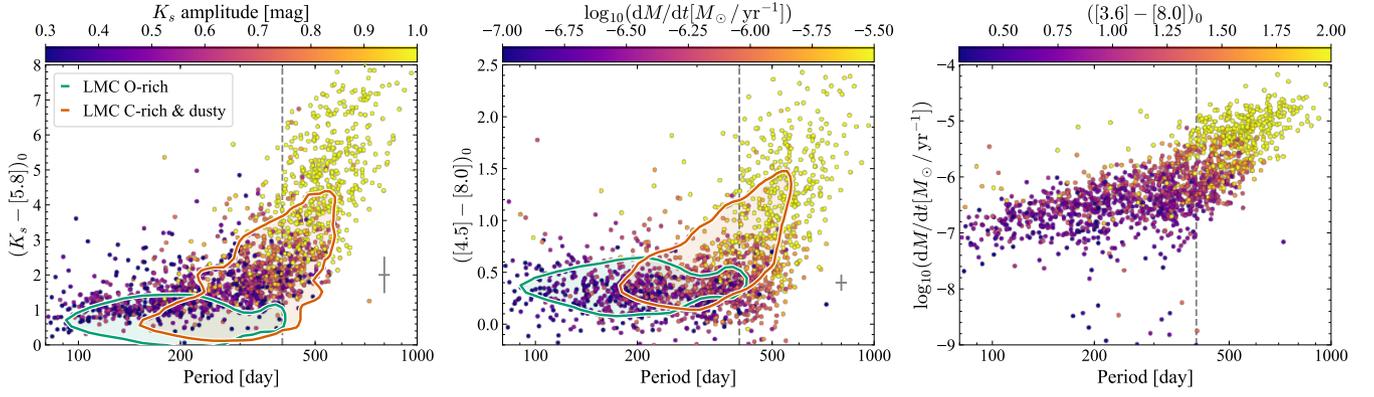
Figure 16. Period–colour diagrams for our reliable Mira variable sample and derived mass loss rates: the leftmost panel shows the distribution in period against $(K_s - [5.8])$ colour (dereddened using a 2D interstellar extinction map derived from the $H - [4.5]$ colour excess of giant stars) coloured by amplitude. 90 per cent of the sample have errors smaller than the displayed grey errorbar. We overlay the distribution of O-rich (green) and C-rich or dusty (red) LMC Mira variables using the classification from Lebzelter et al. (2018). The second panel shows similar but for $([4.5] - [8.0])$ colour and coloured by the modelled mass loss rate assuming O-rich composition. There is a clear transition around 400 days (marked by the vertical line) where the Mira variables become increasingly dust-dominated. The third panel shows the modelled mass loss rate against period coloured by dereddened $([3.6] - [8.0])$.
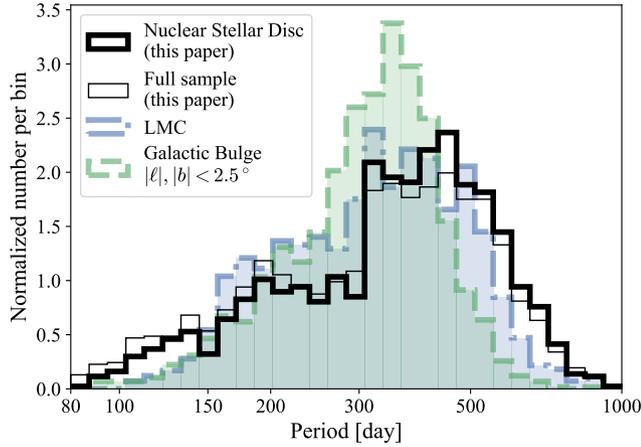
Figure 17. The distribution of periods of our Mira variables (the thick black line shows reliable sources with $|b| < 0.4$ deg and $\Delta K_s > 0.4$ whilst the thin black line shows the full reliable sample). We also display the period distributions of the LMC Mira variables (blue dot-dash) and those in the central Galactic bulge from Gaia and OGLE ($|\ell| < 2.5$ deg, $|b| < 2.5$ deg, green dash).

1782 Mira variable candidates across the NSD of the Milky Way. Our study was motivated by Mira variables being bright intermediate-age indicators that have the potential to characterise the NSD's star formation history and in turn the epoch of bar formation in the Milky Way.

We have demonstrated that although our sample spatially traces the NSD, it is subject to significant selection effects. In the absence of extinction, Mira variables at the Galactic Centre distance are too bright for VVV so we are only able to study Mira variables in highly-extincted regions. However, the completeness of the sample is in excellent agreement with results from artificial star tests so the impact of the selection effects can be modelled. Furthermore, we have demonstrated that:

(i) the sample is dominated by stars with oxygen-rich chemistry ($\gtrsim 90$ per cent),

(ii) the short-period ($\lesssim 400$ day) sample follows period–luminosity relations that locate the stars approximately at the distance of the Galactic Centre with period–Wesenheit relations based on K_s, [3.6] and [4.5] appearing to produce the most reliable distance measurements,

(iii) longer period variables have significant circumstellar dust producing increasingly red colours and high mass loss rates (up to $\sim 2.5 \times 10^{-5} M_\odot \, \mathrm{yr}^{-1}$) and fall well under the O-rich solar neighbourhood and O-rich LMC K_s period–luminosity relations suggesting large quantities of circumstellar dust,

(iv) and finally the age distribution shows two peaks at 10 Gyr and $\lesssim 8$ Gyr either of which we could tentatively associate with NSD formation although the contamination by bulge stars is expected to be significant.

We have only briefly touched upon what is possible using this catalogue. The VIRAC2 reduction of the VVV photometry also provides proper motion measurements for all stars. Although, as we have evidenced, the selection effects of our catalogue are well understood, modelling of kinematic data is often significantly simpler than modelling of spatial data as proper motion observations of bright stars are typically not subject to significant selection effects. This makes identifying NSD membership more straightforward. This will be the subject of the second paper on this sample. We also envisage the catalogue being useful for broader searches and studies of Mira variable stars. Indeed, an earlier version of the catalogue presented here has already been used as part of a long-period variable training set by Molnar et al. (2022).
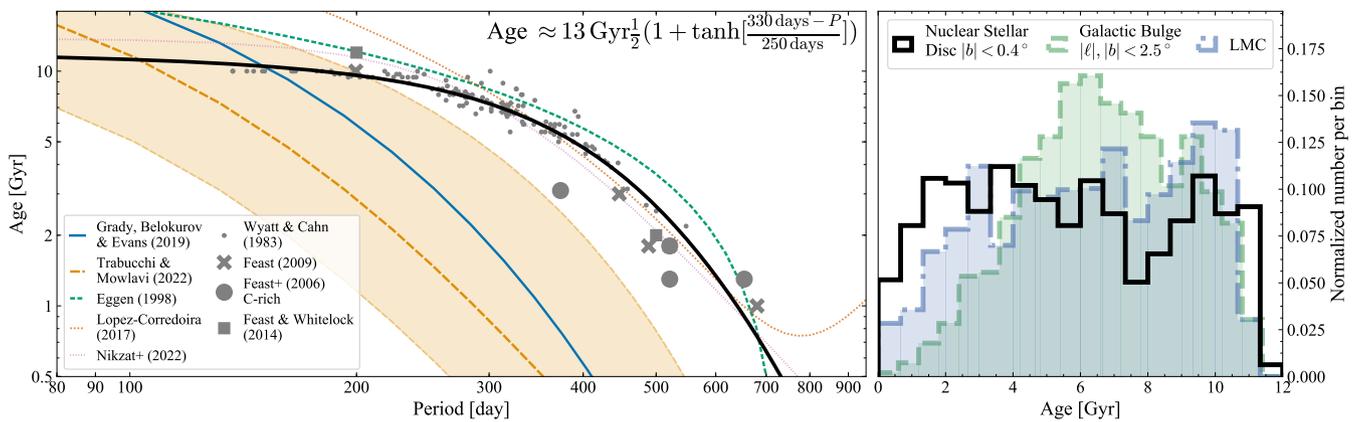
ACKNOWLEDGEMENTS

Figure 18. The age distribution of the nuclear stellar disc: the left panel shows the theoretical period–age relations whilst the right panel shows the derived age distributions assuming an 'average' of the theoretical results of age $\approx 13\,\mathrm{Gyr}\frac{1}{2}(1 + \tanh((330 - P/\mathrm{day})/250))$. The black distribution corresponds to those reliable Mira variables presented here in the midplane $|b| < 0.4°$ with $\Delta K_s > 0.4$, the blue distribution corresponds to Mira variables in the LMC and the green distribution corresponds to Mira variables in the inner Galactic bulge taken from OGLE and Gaia with $|\ell| < 2.5°$ and $|b| < 2.5°$. The long orange-dashed line is from table B1 of Trabucchi & Mowlavi (2022) and the corresponding orange shaded area gives the region over which their model density is > 25 per cent of the maximum.

DATA AVAILABILITY

The resulting catalogue of Mira variables will be made available via Vizier (temporary link available at https://www.homepages.ucl.ac.uk/~ucapjls/data/mira_vvv.fits).

REFERENCES

Alonso-García J., Mateo M., Sen B., Banerjee M., Catelan M., Minniti D., von Braun K., 2012, AJ, 143, 70
Ambikasaran S., 2015, Numerical Linear Algebra with Applications, 22, 1102
Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O'Neil M., 2015, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252
Aringer B., Girardi L., Nowotny W., Marigo P., Lederer M. T., 2009, A&A, 503, 913
Aringer B., Girardi L., Nowotny W., Marigo P., Bressan A., 2016, MNRAS, 457, 3611
Aringer B., Marigo P., Nowotny W., Girardi L., Mečina M., Nanni A., 2019, MNRAS, 487, 2133
Astropy Collaboration et al., 2013, A&A, 558, A33
Athanassoula E., 1992, MNRAS, 259, 345
Baba J., Kawata D., 2020, MNRAS, 492, 4500
Barbuy B., Chiappini C., Gerhard O., 2018, ARA&A, 56, 223
Bensby T., et al., 2013, A&A, 549, A147
Bernard E. J., Schultheis M., Di Matteo P., Hill V., Haywood M., Calamida A., 2018, MNRAS, 477, 3507
Bhardwaj A., et al., 2019, ApJ, 884, 20
Bittner A., et al., 2020, A&A, 643, A65
Blanco V. M., McCarthy M. F., Blanco B. M., 1984, AJ, 89, 636
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Blitz L., Spergel D. N., 1991, ApJ, 379, 631
Blommaert J. A. D. L., van der Veen W. E. C. J., van Langevelde H. J., Habing H. J., Sjouwerman L. O., 1998, A&A, 329, 991
Bovy J., Leung H. W., Hunt J. A. S., Mackereth J. T., García-Hernández D. A., Roman-Lopes A., 2019, MNRAS, 490, 4740
Braga V. F., Contreras Ramos R., Minniti D., Ferreira Lopes C. E., Catelan M., Minniti J. H., Nikzat F., Zoccali M., 2019, A&A, 625, A151
Catchpole R. M., Whitelock P. A., Feast M. W., Hughes S. M. G., Irwin M., Alard C., 2016, MNRAS, 455, 2216
Catelan M., Smith H. A., 2015, Pulsating Stars. Wiley
Chen Y. Q., Casagrande L., Zhao G., Bovy J., Silva Aguirre V., Zhao J. K., Jia Y. P., 2017, ApJ, 840, 77
Churchwell E., et al., 2009, PASP, 121, 213
Dalton G. B., et al., 2006, in McLean I. S., Iye M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62690X, doi:10.1117/12.670018
Deason A. J., Belokurov V., Erkal D., Koposov S. E., Mackey D., 2017, MNRAS, 467, 2636
Deguchi S., et al., 2004, PASJ, 56, 765
Dotter A., et al., 2010, ApJ, 708, 698
Eggen O. J., 1998, AJ, 115, 2435

Elitzur M., Ivezić Ž., 2001, MNRAS, 327, 403

Engels D., Bunzel F., 2015, A&A, 582, A68

Erwin P., Sparke L. S., 2002, AJ, 124, 65

Fakhouri H. K., et al., 2015, ApJ, 815, 58

Feast M. W., 1963, MNRAS, 125, 367

Feast M. W., 2009, in Ueta T., Matsunaga N., Ita Y., eds, AGB Stars and Related Phenomena. p. 48 (arXiv:0812.0250)

Feast M. W., Whitelock P. A., 1987, in Kwok S., Pottasch S. R., eds, Late Stages of Stellar Evolution. p. 33, doi:10.1007/978-94-009-3813-7_3

Feast M., Whitelock P., 2000a, in Matteucci F., Giovannelli F., eds, Astrophysics and Space Science Library Vol. 255, Astrophysics and Space Science Library. p. 229 (arXiv:astro-ph/9911393), doi:10.1007/978-94-010-0938-6_22

Feast M. W., Whitelock P. A., 2000b, MNRAS, 317, 460

Feast M., Whitelock P. A., 2014, in Feltzing S., Zhao G., Walton N. A., Whitelock P., eds, IAU Symposium Vol. 298, Setting the scene for Gaia and LAMOST. pp 40–52 (arXiv:1310.3928), doi:10.1017/S1743921313006182

Feast M. W., Glass I. S., Whitelock P. A., Catchpole R. M., 1989, MNRAS, 241, 375

Feast M., Whitelock P., Menzies J., 2002, MNRAS, 329, L7

Feast M. W., Whitelock P. A., Menzies J. W., 2006, MNRAS, 369, 791

Figer D. F., Rich R. M., Kim S. S., Morris M., Serabyn E., 2004, ApJ, 601, 319

Foreman-Mackey D., Agol E., Ambikasaran S., Angus R., 2017, AJ, 154, 220

Fraser O. J., Hawley S. L., Cook K. H., 2008, AJ, 136, 1242

Freytag B., Liljegren S., Höfner S., 2017, A&A, 600, A137

Fritz T. K., et al., 2011, ApJ, 737, 73

Fujii T., Deguchi S., Ita Y., Izumiura H., Kameya O., Miyazaki A., Nakada Y., 2006, PASJ, 58, 529

Gadotti D. A., et al., 2018, MNRAS, 482, 506

Gadotti D. A., et al., 2020, A&A, 643, A14

Gaia Collaboration et al., 2016, A&A, 595, A1

Gaia Collaboration et al., 2018, A&A, 616, A1

Gallego-Cano E., Schödel R., Nogueras-Lara F., Dong H., Shahzamanian B., Fritz T. K., Gallego-Calvente A. T., Neumayer N., 2020, A&A, 634, A71

Girardi L., 2016, ARA&A, 54, 95

Glass I. S., Evans T. L., 1981, Nature, 291, 303

Glass I. S., Matsumoto S., Carter B. S., Sekiguchi K., 2001, MNRAS, 321, 77

Goldman S., 2020, The Journal of Open Source Software, 5, 2554

Goldman S. R., et al., 2017, MNRAS, 465, 403

Goldman S. R., et al., 2019, ApJ, 877, 49

González-Fernández C., et al., 2018, MNRAS, 474, 5459

Gordon T. A., Agol E., Foreman-Mackey D., 2020, AJ, 160, 240

Gouda N., JASMINE Team 2020, in Valluri M., Sellwood J. A., eds, IAU Symposium Vol. 353, Galactic Dynamics in the Era of Large Surveys. pp 51–53, doi:10.1017/S1743921319007968

Grady J., Belokurov V., Evans N. W., 2019, MNRAS, 483, 3022

Grady J., Belokurov V., Evans N. W., 2020, MNRAS, 492, 3128

Gravity Collaboration et al., 2021, A&A, 647, A59

Groenewegen M. A. T., 2004, A&A, 425, 595

Groenewegen M. A. T., Sloan G. C., 2018, A&A, 609, A114

Guo Z., et al., 2021, MNRAS, 504, 830

Guo Z., et al., 2022, MNRAS, 513, 1015

Gutermuth R. A., Heyer M., 2015, AJ, 149, 64

Habing H. J., Olnon F. M., Winnberg A., Matthews H. E., Baud B., 1983, A&A, 128, 230

Hasselquist S., et al., 2020, ApJ, 901, 109

Hatchfield H. P., Sormani M. C., Tress R. G., Battersby C., Smith R. J., Glover S. C. O., Klessen R. S., 2021, ApJ, 922, 79

He S., Yuan W., Huang J. Z., Long J., Macri L. M., 2016, AJ, 152, 164

Höfner S., Olofsson H., 2018, A&ARv, 26, 1

Holl B., et al., 2018, A&A, 618, A30

Huang C. D., et al., 2018, ApJ, 857, 67

Huang Y., et al., 2020a, ApJS, 249, 29

Huang C. D., et al., 2020b, ApJ, 889, 5

Hunter J. D., 2007, Computing in Science Engineering, 9, 90

Ishihara D., et al., 2010, A&A, 514, A1

Ishihara D., Kaneda H., Onaka T., Ita Y., Matsuura M., Matsunaga N., 2011, A&A, 534, A79

Ita Y., Matsunaga N., 2011, MNRAS, 412, 2345

Ita Y., et al., 2004, MNRAS, 347, 720

Ivezic Z., Elitzur M., 1997, MNRAS, 287, 799

Iwanek P., et al., 2021a, ApJS, 257, 23

Iwanek P., Soszyński I., Kozłowski S., 2021b, ApJ, 919, 99

Jayasinghe T., et al., 2018, MNRAS, 477, 3145

Jayasinghe T., et al., 2019a, MNRAS, p. 2330

Jayasinghe T., et al., 2019b, MNRAS, 486, 1907

Kastner J. H., Thorndike S. L., Romanczyk P. A., Buchanan C. L., Hrivnak B. J., Sahai R., Egan M., 2008, AJ, 136, 1221

Keiner J., Kunis S., Potts D., 2009, ACM Trans. Math. Softw., 36

Kelly B. C., Becker A. C., Sobolewska M., Siemiginowska A., Uttley P., 2014, ApJ, 788, 33

Kučinskas A., Hauschildt P. H., Ludwig H. G., Brott I., Vansevičius V., Lindegren L., Tanabé T., Allard F., 2005, A&A, 442, 281

Lafler J., Kinman T. D., 1965, ApJS, 11, 216

Launhardt R., Zylka R., Mezger P. G., 2002, A&A, 384, 112

Lebzelter T., Mowlavi N., Marigo P., Pastorelli G., Trabucchi M., Wood P. R., Lecoeur-Taïbi I., 2018, A&A, 616, L13

Lebzelter T., et al., 2022, arXiv e-prints, p. arXiv:2206.05745

Lian J., Zhu Q., Kong X., He J., 2014, A&A, 564, A84

Lindqvist M., Habing H. J., Winnberg A., 1992, A&A, 259, 118

Lomb N. R., 1976, Ap&SS, 39, 447

López-Corredoira M., 2017, ApJ, 836, 218

Majewski S. R., Zasowski G., Nidever D. L., 2011, ApJ, 739, 25

Matsunaga N., Kawadu T., Nishiyama S., Nagayama T., Hatano H., Tamura M., Glass I. S., Nagata T., 2009, MNRAS, 399, 1709

Matsunaga N., et al., 2011, Nature, 477, 188

Matsunaga N., et al., 2015, ApJ, 799, 46

Matsunaga N., Menzies J. W., Feast M. W., Whitelock P. A., Onozato H., Barway S., Aydi E., 2017, MNRAS, 469, 4949

Menzies J. W., Feast M. W., Whitelock P. A., Matsunaga N., 2011, MNRAS, 414, 3492

Menzies J. W., Whitelock P. A., Feast M. W., 2015, MNRAS, 452, 910

Menzies J. W., Whitelock P. A., Feast M. W., Matsunaga N., 2019, MNRAS, 483, 5150

Merrill P. W., 1923, ApJ, 58, 215

Messineo M., Habing H. J., Sjouwerman L. O., Omont A., Menten K. M., 2002, A&A, 393, 115

Messineo M., Habing H. J., Menten K. M., Omont A., Sjouwerman L. O., 2004, A&A, 418, 103

Minniti D., et al., 2010, New Astron., 15, 433

Minniti D., Contreras Ramos R., Zoccali M., Rejkuba M., Gonzalez O. A., Valenti E., Gran F., 2016, ApJ, 830, L14

Molina C. N., Borissova J., Catelan M., Lucas P. W., Medina N., Contreras Peña C., Kurtev R., Minniti D., 2019, MNRAS, 482, 5567

Molinari S., et al., 2016, A&A, 591, A149

Molnar T. A., Sanders J. L., Smith L. C., Belokurov V., Lucas P., Minniti D., 2022, MNRAS, 509, 2566

Morris M., Serabyn E., 1996, ARA&A, 34, 645

Mowlavi N., et al., 2018, A&A, 618, A58

Nikutta R., Hunt-Walker N., Nenkova M., Ivezić Ž., Elitzur M., 2014, MNRAS, 442, 3361

Nikzat F., et al., 2022, A&A, 660, A35

Nishiyama S., et al., 2013, ApJ, 769, L28

Nogueras-Lara F., et al., 2019, A&A, 631, A20

Nogueras-Lara F., et al., 2020, Nature Astronomy, 4, 377

Omont A., et al., 2003, A&A, 403, 975

Ou J.-Y., Ngeow C.-C., 2022, AJ, 163, 192

Palmer D. M., 2009, ApJ, 695, 496

Pizzella A., Corsini E. M., Morelli L., Sarzi M., Scarlata C., Stiavelli M., Bertola F., 2002, ApJ, 573, 131

Press W. H., Rybicki G. B., 1989, ApJ, 338, 277

Price-Whelan A. M., et al., 2018, AJ, 156, 123

Ramírez S. V., Arendt R. G., Sellgren K., Stolovy S. R., Cotera A., Smith H. A., Yusef-Zadeh F., 2008, ApJS, 175, 147

Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning. Springer

Riebel D., Meixner M., Fraser O., Srinivasan S., Cook K., Vijh U., 2010, ApJ, 723, 1195

Rimoldini L., et al., 2019, A&A, 625, A97

Rodrigo C., Solano E., 2020, in Contributions to the XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society. p. 182

Rodrigo C., Solano E., Bayo A., 2012, SVO Filter Profile Service Version 1.0, IVOA Working Draft 15 October 2012, doi:10.5479/ADS/bib/2012ivoa.rept.1015R

Saito R. K., et al., 2012, A&A, 537, A107

Samus' N. N., Kazarovets E. V., Durlevich O. V., Kireeva N. N., Pastukhova E. N., 2017, Astronomy Reports, 61, 80

Sanders J. L., Smith L., Evans N. W., Lucas P., 2019, MNRAS, 487, 5188

Sanders J. L., Smith L., González-Fernández C., Lucas P., Minniti D., 2022, MNRAS, 514, 2407

Scargle J. D., 1982, ApJ, 263, 835

Schechter P. L., Mateo M., Saha A., 1993, PASP, 105, 1342

Schlafly E. F., et al., 2018, ApJS, 234, 39

Schödel R., Feldmeier A., Neumayer N., Meyer L., Yelda S., 2014, Classical and Quantum Gravity, 31, 244007

Schönrich R., Aumer M., Sale S. E., 2015, ApJ, 812, L21

Schultheis M., et al., 2020, A&A, 642, A81

Schultheis M., et al., 2021, A&A, 650, A191

Shahzamanian B., Schödel R., Nogueras-Lara F., Martínez-Arranz A., Sormani M. C., Gallego-Calvente A. T., Gallego-Cano E., Alburai A., 2022, A&A, 662, A11

Shappee B. J., et al., 2014, ApJ, 788, 48

Skrutskie M. F., et al., 2006, AJ, 131, 1163

Sloan G. C., Kraemer K. E., Price S. D., Shipman R. F., 2003, ApJS, 147, 379

Smith L. C., et al., 2018, MNRAS, 474, 1826

Sormani M. C., et al., 2022, MNRAS, 512, 1857âĂŞ1884

Soszyński I., et al., 2009, Acta Astron., 59, 239

Soszyński I., et al., 2013, Acta Astron., 63, 21

Stetson P. B., 1987, PASP, 99, 191

Suh K.-W., 1999, MNRAS, 304, 389

Suh K.-W., 2000, MNRAS, 315, 740

Suh K.-W., 2021, ApJS, 256, 43

Suh K.-W., Hong J., 2017, Journal of Korean Astronomical Society, 50, 131

Sutherland W., et al., 2015, A&A, 575, A25

Trabucchi M., Mowlavi N., 2022, A&A, 658, L1

Trabucchi M., Wood P. R., Montalbán J., Marigo P., Pastorelli G., Girardi L., 2019, MNRAS, 482, 929

Trabucchi M., Mowlavi N., Lebzelter T., 2021, A&A, 656, A66

VandenBerg D. A., Brogaard K., Leaman R., Casagrande L., 2013, ApJ, 775, 134

VanderPlas J. T., 2018, ApJS, 236, 16

VanderPlas J. T., Ivezić Ž., 2015, ApJ, 812, 18

van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science Engineering, 13, 22

Vassiliadis E., Wood P. R., 1993, ApJ, 413, 641

Virtanen P., et al., 2020, Nature Methods, 17, 261

Wang S., Chen X., 2019, ApJ, 877, 116

Waskom M. L., 2021, Journal of Open Source Software, 6, 3021

Watson C. L., Henden A. A., Price A., 2006, Society for Astronomical Sciences Annual Symposium, 25, 47

Wegg C., Gerhard O., 2013, MNRAS, 435, 1874

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

Whitelock P. A., Menzies J. W., Feast M. W., Nsengiyumva F., Matsunaga N., 2013, MNRAS, 428, 2216

Wood P. R., 2000, Publ. Astron. Soc. Australia, 17, 18

Wood P. R., 2015, MNRAS, 448, 3829

Wood P. R., Habing H. J., McGregor P. J., 1998, A&A, 336, 925

Wood P. R., et al., 1999, in Le Bertre T., Lebre A., Waelkens C., eds, IAU Symposium Proceedings Vol. 191, Asymptotic Giant Branch Stars. p. 151

Wright E. L., et al., 2010, AJ, 140, 1868

Wyatt S. P., Cahn J. H., 1983, ApJ, 275, 225

Xiong D. R., Deng L., Zhang C., 2018, MNRAS, 480, 2698

Yuan W., He S., Macri L. M., Long J., Huang J. Z., 2017a, AJ, 153, 170

Yuan W., Macri L. M., He S., Huang J. Z., Kanbur S. M., Ngeow C.-C., 2017b, AJ, 154, 149

Zechmeister M., Kürster M., 2009, A&A, 496, 577

Zijlstra A. A., Bedding T. R., Mattei J. A., 2002, MNRAS, 334, 498

Zinn J. C., et al., 2017, MNRAS, 468, 2189

Zoccali M., et al., 2003, A&A, 399, 931

APPENDIX A: TIME SERIES MODELLING

To extract a sample of Mira variables from the VVV VIRAC2 data, we use both standard periodogram methods as well as more flexible Gaussian process methods. In this appendix, we describe in detail our implementations of these methods. We describe our methods in terms of a general light curve of N magnitudes $\boldsymbol{y}$ with covariance matrix Σ (where we assume uncorrelated errors so $\Sigma = \sigma^2 I$) at times $\boldsymbol{t} = (t_1, t_2, \cdots, t_N)$.

A1 Periodogram methods

The standard method for characterising the periodicity of a variable source is using the Lomb-Scargle method (Lomb 1976; Scargle 1982), which has been generalised to account for a floating mean (Zechmeister & Kürster 2009). As discussed by VanderPlas (2018), this method is equivalent to fitting the first three terms of a Fourier series $K_s(t) = a_0 + a_1 \sin(\omega t) + b_1 \cos(\omega t)$ and so performs well for near-sinusoidal light curves. For non-sinusoidal, but periodic, light curves with long-term trends (as observed in some Mira variables), a general Fourier series plus polynomial of the form

$$K_s(t) = a_0 + \sum_{k=1}^{N_F} \left[a_k \sin(k\omega t) + b_k \cos(k\omega t) \right] + \sum_{k=1}^{N_P} c_k (t-t_0)^k, \quad \text{(A1)}$$

can be fitted to find a_k, b_k and c_k using linear least-squares (Palmer 2009). Here t_0 is an arbitrary zeropoint not fitted for. For computational speed, non-equispaced Fast Fourier transforms (NFFT) are used to evaluate the resulting sums over trigonometric functions (times polynomial terms, Press & Rybicki 1989; Keiner et al. 2009). These operate by 'extirpolating' the non-equispaced data onto a regular grid using sets of basis functions e.g. Lagrange polynomials (Press & Rybicki 1989) or a set of localised Gaussians where sparsity can be enforced as a further approximation (Keiner et al. 2009). We use Jake VanderPlas' implementation of the algorithms presented in Keiner et al. (2009)[6]. To avoid overfitting higher order Fourier terms, regularization is introduced by including a term $\theta^T \Lambda \theta$ in χ^2 where θ is the parameter vector, $\theta^T = (a_0, a_1, b_1, a_2, b_2, \cdots)$ (VanderPlas & Ivezić 2015). We choose $\Lambda = \lambda \text{Tr}\Sigma^{-1} \boldsymbol{k}^2 = \lambda \text{Tr}\Sigma^{-1}(0, 1, 1, 4, 4, \cdots)$, which minimises the curvature of the Fourier fits. As a default, we set $\lambda = 0.01$. No regularization is employed for the polynomial coefficients c_k.

When finding the best-fitting period, we use a regular grid of test frequencies from the ASTROPY autofrequency routine which implements the guidelines in VanderPlas (2018, Astropy Collaboration et al. 2013; Price-Whelan et al. 2018). The best-fit period is that with

---

[6] https://github.com/jakevdp/nfft

the lowest χ^2 and the uncertainty is estimated by the local curvature of χ^2. The covariance of the Fourier coefficients are found by summing the covariance at the best period with the covariance from linear propagation of the period error (calculated by finite-differencing the Fourier coefficients on the period grid).

A2 Gaussian process models

For quasi-periodic light curves such as those of Mira variables, we require a more flexible fitting procedure that can account for stochastic variations. Several authors (He et al. 2016; Zinn et al. 2017; Yuan et al. 2017b) have demonstrated the power of Gaussian process models for Mira variable light curves. A Gaussian process is a prior on the space of functions $y(\boldsymbol{x})$, where a finite set of N random variables $\{y_i\}$ with uncertainties $\{\sigma_i\}$ indexed by D-dimensional points $\{\boldsymbol{x}_i\}$ is distributed as a multidimensional Gaussian distribution (Rasmussen & Williams 2006). The Gaussian process is characterised by a kernel $\mathsf{K}_{ij} \equiv k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which describes the degree of correlation between points. Here we will use stationary kernels i.e. those that depend only on the Euclidean distance $\tau_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|$. Kernel parameters are optimally set by maximising the likelihood of $\{y_i\}$ given $\{\boldsymbol{x}_i\}$.

Despite their simplicity, Gaussian processes are ill-suited for large datasets due to the requirement of performing the $O(N^3)$ matrix inversion. However, for one-dimensional data Kelly et al. (2014), Ambikasaran (2015) and Foreman-Mackey et al. (2017) have demonstrated that for kernels composed of sums of stationary exponential kernels, the kernel matrix K is semi-separable and the matrix inversion can be evaluated with an $O(N)$ algorithm. The algorithm is implemented in the CELERITE python package (Foreman-Mackey et al. 2017). In the CELERITE scheme, the kernel must be positive-definite and of the form

$$k(\tau) = \sum_{j=1}^{J} \frac{1}{2} \left[(a_j + i\, b_j)\, e^{-(c_j + i\, d_j)\tau} + (a_j - i\, b_j)\, e^{-(c_j - i\, d_j)\tau} \right],$$
(A2)

where a_j, b_j, c_j and d_j are real numbers and τ the separation between points. Positive definiteness is ensured by requiring $|b_j d_j| < a_j c_j$ for all j. Foreman-Mackey et al. (2017) show how a number of popular kernels can be approximated with this kernel. In particular, both damped random walk (or the Ornstein-Uhlenbeck kernel) $k_{\mathrm{DRW}}(\tau)$ ($b_j = d_j = 0$) and (quasi-)periodic kernels are permitted. For periodic kernels, Foreman-Mackey et al. (2017) suggest working with the damped harmonic oscillator kernel $k_{\mathrm{SHO}}(\tau)$ described by the power spectral density

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0 \omega_0^4}{(\omega^2 - \omega_0^2)^2 + \omega_0^2 \omega^2 / Q^2},$$
(A3)

where S_0 is related to the amplitude of oscillation, ω_0 the frequency and Q is the quality factor describing the damping. These parameters are related to a_j, b_j, c_j and d_j.

We have found it useful to work with a kernel of the form

$$k(\tau) = \sum_{i}^{N_O} k_{\mathrm{SHO},i}(\tau)\, k_{\mathrm{DRW},i}(\tau) + \sum_{j}^{N_E} k_{\mathrm{DRW},j}(\tau).$$
(A4)

The first term gives a sum of damped harmonic oscillators – we choose to multiply each SHO term by a DRW and initialize Q as a large value. Although the DRW multiplying the SHO duplicates the parameters in the SHO we have found it helps with convergence of the models. The second term is a pure random walk. We additionally include a white noise term in the kernel so $\mathsf{K} \leftarrow \mathsf{K} + \sigma_w^2 \mathsf{I}$. Such a

model is appropriate for Mira variables as it encompasses (i) multiple periodic signals, (ii) slow changes in the period and (iii) stochastic changes. For small datasets, He et al. (2016) opt for a hybrid Fourier and Gaussian process method where the residuals with respect to a Fourier series fit are described by a Gaussian process with a DRW kernel. We find this method gives similar results to only using a Gaussian process with a SHO plus DRW kernel, but is more sensitive to the initial choice of period and does not allow the small changes in the period (without employing a more complex kernel).

A3 Multi-band light curves

For the modelling of multi-band light curve data, the Gaussian process models can be generalized to include effective wavelength as a second dimension by using a 2D kernel (e.g. Fakhouri et al. 2015, for supernova light curve modelling). This method produces correlations in the phase across different photometric bands. However, we then lose the ability to use the rapid algorithms such as CELERITE, applicable for 1D. For higher dimensionality data, Ambikasaran et al. (2015) have described how covariance matrices can be inverted approximately by hierarchical factorisation in $O(N \log^2 N)$ time. However, the overheads become costly for periodic kernels. A cheaper method is to combine all the individual photometric bands into a single light curve using a band-dependent scale and shift. The combined light curve can then be considered as a 1D problem and modelled with a CELERITE kernel (Gordon et al. 2020). The disadvantage of this procedure is we are unable to introduce a smooth correlation scale in the wavelength dimension to perhaps reflect the fact that variability in close bands are related.

We utilise the same kernel as given in equation (A4) and maximise the likelihood to find the kernel parameters additionally fitting for the white noise σ_w, the mean μ and scaling α (relative to the K_s band) in each photometric band. Furthermore, we place priors on the scaling amplitudes α based on the O-rich amplitude ratios derived from synthetic spectral energy distributions from fits to multi-band light curves by Iwanek et al. (2021a, reported relative to I band amplitude). We use their table 1 to find the mean $\mathrm{amp}(x)/\mathrm{amp}(K_s)$ and variance $\sigma^2_{\mathrm{amp}(x)/\mathrm{amp}(K_s)} = (\mathrm{amp}(I)/\mathrm{amp}(K_s))^2 \Big(\sigma^2_{\mathrm{amp}(x)/\mathrm{amp}(I)} + (\mathrm{amp}(x)/\mathrm{amp}(K_s))^2 \sigma^2_{\mathrm{amp}(K_s)/\mathrm{amp}(I)} \Big)$. Due to the uncertainty in $\mathrm{amp}(K_s)/\mathrm{amp}(I)$ from Iwanek et al. (2021a) this is quite a generous prior.

APPENDIX B: BLENDED PHOTOMETRY

In our early Mira variable candidate lists, we found significant numbers of variable sources clumped in period and with on-sky distributions tracing the VIRCAM footprint. One of these sources is shown in Fig. B1, which illustrates the following discussion and which we will discuss in full below. The most notable of these spurious sources had periods of 100 or ~ 215 days (and its multiple at ~ 430 days and were concentrated in the region $0 < \ell < 1.5\,\mathrm{deg}$ and $-1.5\,\mathrm{deg} < b < 0$. The likely explanation for a clumping in period is due to aliases arising from the cadence of the observations. However, the periodogram of an example light curve with the magnitudes replaced by noise (VanderPlas 2018) did not show peaks at the observed periods. Inspecting some example light curves of these spurious sources revealed that many of these sources had almost 'bimodal' light curves, where the magnitude was one of two values. This bimodality corresponded to two distinct sequences in magnitude against seeing (and in turn, magnitude against DoPhot χ^2). For low seeing observations,

the magnitudes were consistently fainter. These spurious sources regularly had other VIRAC detections within 1 arcsec, and furthermore, the two distinct magnitude sequences corresponded to the epochs when the neighbour was detected or not. For low seeing, the blended source is resolved and a fainter magnitude assigned, whilst higher seeing results in all the light being attributed to a single source. Curiously, periodicity in the seeing, possibly due to seasonal variations, is then transferred into periodicity of the source. This is confirmed by taking a periodogram of the seeing which has exactly the observed peaks of 100, ~ 215 and ~ 430 days.

The described issues with blended sources are a limitation of the employed photometric reduction procedure. Schechter et al. (1993) demonstrate that both DoPhot and DaoPhot (Stetson 1987) exhibit the same systematics with blending, so the problem is likely to occur if we employed a different photometric algorithm. However, a cleaner route around the problem is to use a list of sources from the best seeing epochs to perform forced photometry on the poorer seeing epochs. Furthermore, the source list can be supplemented with the Gaia, DECAPS or Pan-STARRS catalogues. However, here we will adopt a simpler procedure by a posteriori correcting any trends in magnitude with seeing when the source is suspected to be contaminated by blending. It should be said that the problem was more significant in early versions of the VIRAC2 catalogues that did not include information on whether detections were astrometric outliers. With this information the problem has been somewhat alleviated.

For each candidate variable source, we take the raw light curve and remove those observations with $\chi^2 > 10$ for $K_s < 13.2$ (the basic quality cut described in Section 2). We then check if there are any Gaia, DECAPS or reliable VIRAC sources within 1 arcsec of the source (for Gaia and DECAPS we require two sources within 1 arcsec – the source itself and potential neighbours). Here, reliable means non-duplicate, detected at least 10 times and detected in at least 20 per cent of the covering observations as explained in Section 2.1. We then separate the light curve observations into those without neighbours detected at the same epoch, and those with different combinations of neighbours detected (e.g. if two neighbours, there are four groups: 1. no neighbours, 2. neighbour 1, 3. neighbour 2 and 4. neighbours 1 and 2). For each group of observations, we fit a simple Gaussian process model to 1/seeing against magnitude using an exponential squared kernel and a white noise term (using the basic solver in George, Ambikasaran et al. 2015). The observations are then shifted by the fitted model trend and the overall magnitude is selected as the model value at the 10th percentile in seeing for the group with the most neighbours detected (provided there are more than four observations in this group, in which case we use the group with the second most neighbours detected etc.). We select the best seeing observations with most neighbours for the overall shift as this is anticipated to produce the most accurate photometry for a source. However, this is a likely source of systematic biases in the magnitudes.

In Fig. B1 we show an example blended source in the VIRAC2 catalogue (at (RA,Dec)=(267.64689, −28.36941) deg). This source has median magnitude $K_s = 13.5$ and a single neighbour 0.52 arcsec away with $K_s = 14.2$. The observations where the neighbour is detected separate clearly in magnitude and seeing from those where no neighbour is detected. In the worse seeing observations, the neighbour is not detected and the magnitude is measured as ~ 13.4 mag, whilst the better seeing observations allow resolution of the two sources and a magnitude measurement of ~ 13.7 mag. The two clumps separate in seeing around the 'Nyquist limit' (2× pixel size) of 0.678 arcsec for VIRCAM. Interestingly, the aperture photometry shows the opposite trend in photometry with seeing. For poor seeing,
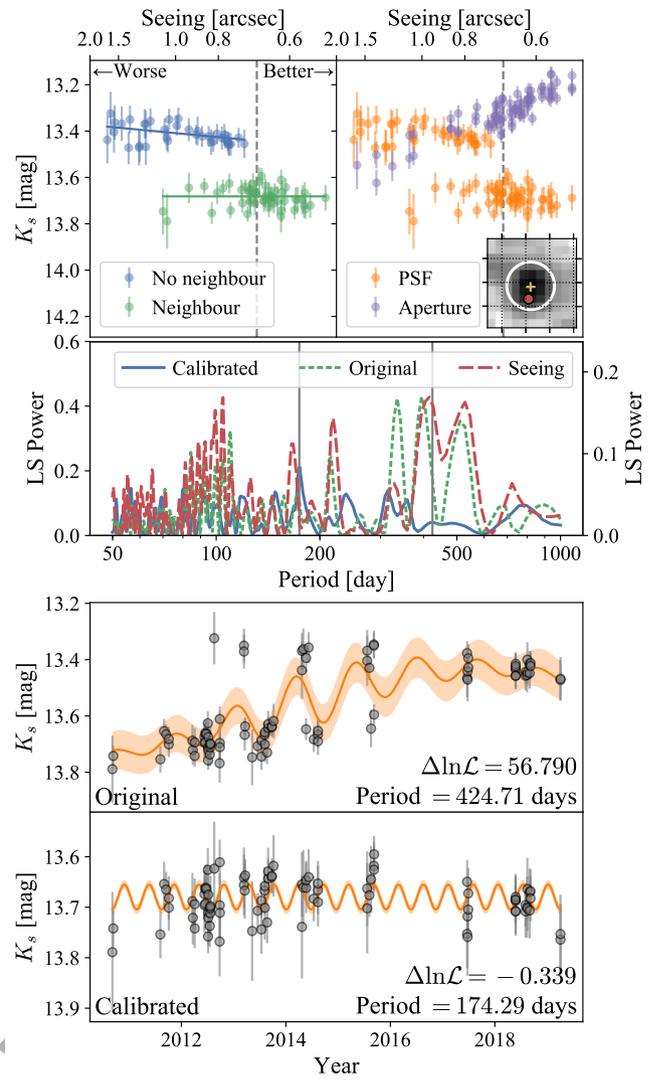

Figure B1. Example of a light curve for a blended source showing spurious variability and periodicity. The *top two panels* show the variation of magnitude with (inverse) seeing: in the *left panel* we colour the points by whether a neighbour (green) or no neighbour (blue) is detected within 1 arcsec at the same epoch (see small deep-stack image inset of source [yellow cross] and neighbour [red circle] with a white 1 arcsec circle) and display the mean trend models adopted; the *right panel* shows the variation of PSF (orange and corresponding to all points in the left panel) and aperture (purple) photometry with seeing. The *second panel* shows periodograms of the calibrated light curve (blue solid), original light curve (green short-dashed) and seeing (red long-dashed) (note the seeing periodogram corresponds to the right axis). The *bottom two panels* show the light curves with Gaussian process models where the two periods are marked on the periodogram panel.

more light falls outside the aperture and for good seeing, less light. In both cases, this is not corrected for correctly. The original light curve (third panel) shows the bimodal distribution of magnitudes, which are seemingly well fitted by a periodic signal with ~ 430 days. This period corresponds to the highest peak in the Lomb Scargle periodogram. The errors introduced by the seeing produce a clear periodicity. This can be seen plainly by the periodogram of seeing which traces the light curve periodogram. Correcting the light curve by the procedure described above removes the 430 day peak and reduces the amplitude of variability. The model now fits a period of

174 days which may still be a systematic error (the peak corresponds to the slight wing in the seeing periodogram) but this periodicity is not deemed significant.

APPENDIX C: COMPARISON WITH KNOWN MIRA PERIODS

We validate our procedure for measuring the periods of Mira variables using those previously discovered and studied. Unfortunately, the VVV overlap with known Mira variables is relatively small. Mira variables are intrinsically very red $(G - K_s) \sim 6$ and those in the bulge can be heavily extincted. This means the Mira variables are typically at the bright end of infrared surveys and the faint end of optical surveys. Non-linearity in the VVV photometry begins brighter than $K_s \sim 11.5$ mag and the photometry completely saturates for ~ 9 mag meaning the bright well-studied Mira variables in low extinction regions are too bright for VVV. Similarly, those Mira variables that are well covered by VVV are typically too faint for the optical surveys. However, a sufficient number overlap for testing purposes, particularly as we are fortunate to have the infrared Mira variable catalogue of M09. For cross-matching to other catalogues, we employ a cross-match radius of 0.4 arcsec. For OGLE, Gaia and VSX, we also cross-match to 2MASS with a 1 arcsec radius and ignore the source if 2MASS $K_s < 9$ (as it is likely saturated in VVV) or VVV and 2MASS K_s differ by more than 2 magnitudes. In these cases any VVV cross-match is likely erroneous.

C1 Matsunaga et al. (2009) Mira variable catalogue

M09 used the SIRIUS near-IR camera on the Infrared Survey Facility (IRSF) 1.4 m telescope to image in JHK_s a 20×30 arcmin^2 area around the Galactic centre. The authors first select sources with 3 times more variable photometry than the median expectation as a function of magnitude, and then measure the periods by fitting sinusoids in a least squares sense (equation (A1) with $N_F = 1$) to identify 1364 long period variable candidates, of which 549 were assigned periods. The majority of the sources with periods are likely Mira variables, although as acknowledged by M09 the exact definition of a Mira variable from infrared photometry is awkward. For certainty in their analysis, these authors consider stars non-Mira variables if the amplitude in any of J, H or K_s is less than 0.4 although this removes only ~ 10 per cent of the stars. The M09 sample is most similar to the VVV data so ideal for checking the VVV results, except that (i) the M09 photometry is non-linear brighter than $K_s = 9$ mag so many brighter Mira variables are better studied than in VVV and (ii) variability and periodicity is characterised in J, H and K_s (although only all three for ~ 20 per cent of the data) whilst in VVV we must essentially rely on K_s alone (there is more limited coverage from J and H photometry). Of the 549 long period variables with periods in M09, we successfully cross-match 212 in VIRAC2, of which 192 have more than 20 valid K_s epochs in their cleaned light curve.

C2 OGLE long period variables in the bulge

The Optical Gravitational Lensing Experiment (OGLE) is a long term variability study of the Magellanic Clouds and the Galactic bulge and disc from mainly I (and some V) photometry using the 1.3m Warsaw telescope at the Las Campanas Observatory. The OGLE-II and OGLE-III surveys took place between 1997 and 2009 and produced catalogues of long period variables in both the LMC and the Galactic bulge (Soszyński et al. 2009, 2013). Periods are measured

in the range $5 - 2000$ days using a $N_F = 3$-term Fourier series least-squares fit which is subtracted and iteratively repeated to measure up to five periods (of which three are reported). From the Galactic bulge catalogue of Soszyński et al. (2013), 6528 Mira variables were discovered, the majority of which lie within the VVV footprint. Although the I band is near optimal for surveying unextincted Mira variables, within the Galactic bulge its use becomes limited in high extinction regions ($|b| \lesssim 2$ deg). However, the dense time-sampling of the OGLE survey makes the period determination highly reliable. Of the 6528 Mira variables, we successfully cross-match 43 in VVV, of which 39 have more than 20 valid K_s epochs.

C3 Gaia DR2 long period variables

A catalogue of long period variables (LPVs) with periods was produced as part of the Gaia DR2 data release (Mowlavi et al. 2018). Based on an average of 26 observations per star over 668 days, LPV candidates were identified using a random-forest classifier trained on photometric attributes, of which red colours and long term variability are the most useful (Holl et al. 2018; Rimoldini et al. 2019). A further cut to retain LPVs with G-band amplitudes $(5 - 95\%)$ greater than 0.2 mag was applied, producing a list of $\sim 150,000$ candidates. The generalized Lomb-Scargle method was employed to search for periods in the range 10 to 1000 days and only those $\sim 90,000$ LPVs with periods > 60 days have published periods. Comparison with all-sky ASAS-SN observations and OGLE observations in the LMC and bulge demonstrated the good recovery of the Gaia periods for overlapping sources although due to the scanning strategy periods around ~ 190 days and below 120 days are susceptible to aliasing.

We limit the Gaia DR2 LPV catalogue to likely Mira variables using the cut on G-band amplitude from Grady et al. (2019): $\mathrm{Amp}(G) = \log_{10}(\sqrt{N_{obs}}\sigma_{\bar{I}_G}/\bar{I}_G) > -0.55$ (where $\bar{I}_G$ is the mean G flux and $\sigma_{\bar{I}_G}$ its error). Mowlavi et al. (2018) also provide frequency uncertainties Δf which we transform to period uncertainties under a linear approximation: $\Delta T = \Delta f/f^2$. Cross-matching to VVV, we find 39 matches of which 35 have more than 20 valid K_s epochs.

C4 VSX

The AAVSO International Variable Star Index (Watson et al. 2006) is a compilation of variables initially constructed from the General Catalogue of Variable Stars (Samus' et al. 2017). It includes all variables from the ASAS-SN Variable Stars Database (Shappee et al. 2014; Jayasinghe et al. 2018, 2019b,a) We downloaded all variables labelled as type 'M' or 'M:' ('M:' are uncertain Mira classifications) on 5th August 2020, removed objects from M09, OGLE-III and Gaia which we have already considered (the full Gaia LPV catalogue was not included in the VSX catalogue version we used), and cross-match to VIRAC2. We successfully cross-match 11 Mira variables with periods to sources in VVV, of which 10 have more than 20 valid K_s epochs.

C5 Mira variable period recovery

In Fig. C1 we show the periods from the four literature sources plotted against the periods measured from the VVV data. We employ our Gaussian process method using a range of different initial period guesses and kernels (as described previously). We only show results for light curves with more than 20 valid K_s epochs and with a K_s amplitude greater than 0.25 mag (corresponding to our lower selection limit in Fig. 5). We have found that light curves outside
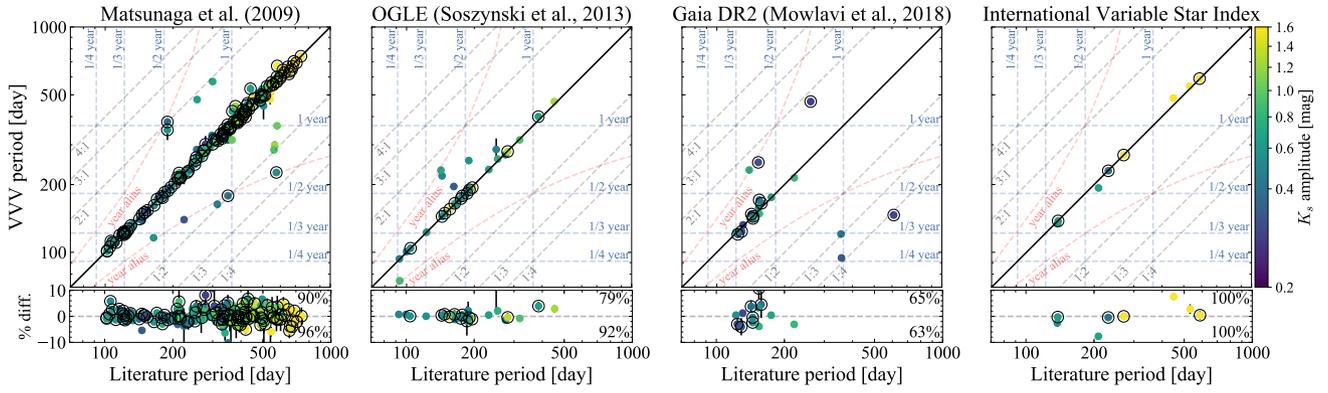
Figure C1. Period recovery using K_s-band VVV data for known Mira variables observed by VVV. We show all stars with periods provided by Matsunaga et al. (2009), OGLE (Soszyński et al. 2013), Gaia (Mowlavi et al. 2018) and the International Variable Star Index (VSX), and with more than 20 valid K_s epochs and a K_s amplitude of > 0.2 in VVV (circled points have more than 100 valid epochs). The points are coloured by K_s amplitude. The black solid line is $1 : 1$ whilst dashed grey lines are $4 : 1$, $3 : 1$, $2 : 1$, $1 : 2$, $1 : 3$ and $1 : 4$. The red dashed lines are $1/(1/\text{Period} \pm 1/365\,\text{days})$ – likely aliases given the VVV observing strategy. Horizontal and vertical blue dashed lines show fractions of a year. The upper (lower) numbers in the lower panels give the percentages of sources with > 20 (> 100) K_s epochs for which our period agrees within 25 per cent with the literature period.

these cuts have quite poor period recovery and this motivates our choice of epoch number cut for the full catalogue. From Fig. C1, we find a high fraction of measurements lie on the $1 : 1$ line, particularly those circled sources with > 100 valid K_s epochs. Likely alias frequencies for VVV light curves are $n\,\text{year}^{-1}$ ($n = 1, 2, 3, \cdots$, due to the observing strategy), and linear combinations of the true frequency and $1\,\text{year}^{-1}$ i.e. the measured period is $1/(1/P \pm 1\,\text{year})$ for true period P. Also, we might expect for near sinusoidal Mira light curves, the measured period may be mis-measured as a harmonic of the true period. We see there is a tendency for the VVV period to lie along the alias lines. However, in general the period recovery is very good. We deem periods to agree if they are within 25 per cent of each other (as this encompasses all sources on the one-to-one in the M09 comparison). We find upwards of ~ 90 per cent agree with the literature samples for those sources with > 100 K_s epochs and this decreases to only $\gtrsim 80$ per cent for those with > 20 K_s epochs. For the Gaia DR2 sample the agreement is less good (~ 65 per cent). Restricting to only those in the bulge region ups this to ~ 77 per cent possibly. We also see two sources have year periods from Gaia. This may be genuine and then missed by VVV or they could be aliases in Gaia. Only 14 of the cross-matched DR2 LPVs are in the new Gaia DR3 LPV catalogue from Lebzelter et al. (2022) of which we find 75 per cent matching periods or 80 per cent restricting to those with > 100 K_s epochs. This suggests some of the matched Gaia DR2 LPVs are spurious or possibly contaminating YSOs in the VVV disc region (Mowlavi et al. 2018).

period relation (as traced approximately by the M09 sources) in all projections.

APPENDIX D: WESENHEIT MAGNITUDE–PERIOD CUTS

In Section 2.2 we employ a series of cuts for isolating reliable Mira variables in our sample. The primary cuts are in delta log-likelihood with respect to a non-periodic model, period–amplitude space (as shown in Fig. 5), parallax signal-to-noise and Wesenheit magnitude–period space (as Mira variables are known to follow period–luminosity relations). In Fig. D1 we display the Wesenheit magnitude–period cuts described in Section 2.3. There is a population of contaminants around $K_s - 1.328(H - K_s) \approx 13$ that is particularly prominent around 1 year periods. These sources appear to fall significantly under the expected Wesenheit magnitude against
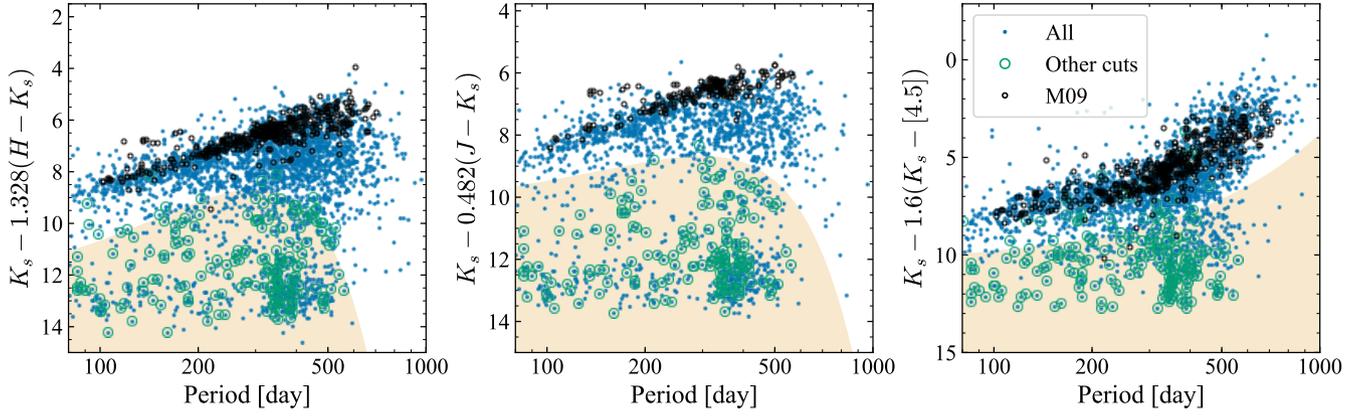
Figure D1. Illustrations of the cuts performed to isolate the likely Mira variable stars. Each panel shows a different Wesenheit magnitude–period distribution. The blue points are the full set of Mira variable candidates without any Wesenheit magnitude–period cuts applied (they still satisfy the period–amplitude, parallax and delta log-likelihood cuts described in Section 2.3). The orange shaded regions show the selection of sources we exclude. In each panel, the green circles show the sources that satisfy the other two cuts. The black points are the data from Matsunaga et al. (2009).