Determining periods of Mira Variables using the VVV sky survey

Kylar Greene

May 21, 2019

Abstract

A string-length method of calculating light curve periods of sources with sparse time series sampling is presented and its application to determine the period of Mira variables based on catalog data is analyzed. The estimated period which creates the shortest path length between points in phase space is the statistically most likely period for the data. By applying constraints based on physical properties of Mira variables, such as minimum periods of 80 days, and maximum magnitude amplitudes of 10, periods can be reliably extracted from multi-epoch catalogs with more than 7 observations. The accuracy of the calculated periods increase quickly as more observations are included. The described methodology is applied to the publicly available Ks-band multi-epoch catalog from the VVV survey. A list of known Mira variables are cross-matched with this multi-epoch catalog, resulting in 618 cross-matched sources. From the 618 cross-matches, 227 contain more than 5 data points. 90 of these 227 sources showed significant variability. For 41 of the 90 sources, periods could be estimated using the string-length method.

1 Introduction

We study the Milky Way galaxy for many reasons, but perhaps the most important reason is to understand the structure of similar galaxies and their evolution. Even in nearby galaxies it is a challenge to spatially resolve individual stars, and even more challenging to measure their motions. Within the Milky Way, we are provided with a unique opportunity to investigate the dynamics of stellar populations as they can be resolved and their positions pinpointed. The Bulge Asymmetries and Dynamical Evolution (BAaDE) project's primary goal is just that; to study the gravitational potential in the bulge and bar region of the Milky Way. The BAaDE project surveys approximately 28,000 stars for SiO maser emission, forming a resulting data base of stellar positions and velocities. Velocities are extracted by measuring the frequency-shift of the SiO maser line and using Doppler shift calculations. The locations of the sources on the sky are known to a precision of greater than $0.''1$, but the distance along the line-of-sight to most sources is unknown. [1]

In fact, this is a common issue in astronomy as the night sky is viewed as a 2D surface with no radial depth. The distance measurement techniques used are commonly arranged in a 'distance ladder', as specific techniques are only useful within certain ranges of distances. The first and most accurate method of distant measurement is trigonometric parallax [2]. Such observations measure the apparent shift of a source's position on the sky with respect to the background stars and then use trigonometric relations to calculate how far the source must be. The measurements are generally taken half a year apart when the Earth is on either side of the Sun, in an effort to see the greatest apparent positional shift of the target. This technique becomes more difficult with distant objects as the farther away the target is, the smaller the parallax angle becomes. Currently, the Gaia space telescope offers the farthest viable optical distance calculations using trigonometric parallaxes, reaching distances up to 9.19 kpc. Comparing that to the distance between the Earth and the Galactic Center

of 8 kpc [3], Gaia is able to measure positions and angular shifts of optically bright and compact sources as far away as the center of the Galaxy with a 10% or better accuracy [4].

For distances larger than a few kpc, the next method available on the distance ladder is using Cepheid variable stars. These stars are typically identified as yellow supergiants, whose brightness changes very regularly with periods between 1 and 40 days. The period of oscillation in the brightness has been shown to correlate with the absolute magnitude of the star. Then with a measurement of the apparent brightness and application of the distance modulus, a distance can be derived. The distances calculated with Cepheid variable measurements have been calibrated with parallax measurements where possible, and are accurate indicators of line-of-sight distance effective out to 3×10^5 kpc, well beyond the size of the Milky Way galaxy. [2]

Then by using these two methods, distances to optically bright, compact sources or super giants should, in principle, be measurable anywhere in the Milky Way. But this is not the case for the inner Milky Way as like most spiral galaxies, it contains a great amount of dust. Interstellar dust is very efficient at absorbing and scattering light, a phenomena referred to as extinction. Extinction affects shorter wavelengths more than longer ones due to the dust particle size, which causes difficulties in the distance measurements made by Gaia as it is an optical instrument and thus has difficulties reaching deep inside the bulge and plane regions. Similarly, the Cepheid period-luminosity relation is strong in the optical, while these objects are about three magnitudes dimmer in the infrared (IR) with less accurate models present for the period to luminosity relation in that regime. Therefore, other methods using longer wavelength measurements must be developed to determine intra-galatic distances for objects located in the central region of the Galaxy.

Stepping away from the optically observed Cepheids, there is another type of varible star; the Mira variables. These objects are late type evolved variable stars whose periods can range between 100 and 1000 days with amplitudes in their magnitude brightness of up to ± 5 [5]. These objects are especially bright in the IR range, making them more easily detectable through the interstellar medium (ISM) present in the center of the Galaxy. Because of this, these objects are excellent candidates to investigate the possible existence of a Period-Luminosity (PL) relationship similar to that established for Cepheid variables, allowing them to be used as distance indicators in the bulge of the Milky Way.

There is one major obstacle in establishing a PL relationship for Miras: their very long periods. It is much too expensive in regard to telescope time to monitor each single object for years, so little data is directly available for the periods of most Miras. However, as sky surveys become more common and available, they may offer a solution to this problem. Sky surveys like Wide-field Infrared Survey Explorer (WISE) [6], Deep Near Infrared Survey of the Southern Sky (DENIS) [7], and the Vista Variables in the Via Lactea (VVV) [8] systematically take observations of the sky. These surveys typically collect data over a few years, and provide more detailed source lists as the data is taken, reduced, and released. Once the sky survey catalogs are available, a list of known objects can be-cross matched with the catalog, based on position. If multi-epoch data can be obtained, periods can be determined. Once periods are derived, they can be plotted against luminosity for a subset of sources with known distances, and then a P-L relationship can be deduced. With the PL relationship, distances can then be estimated to all Mira stars with periods by using their apparent magnitude, thus hopefully allowing distances to be estimated to the bulk of the BAaDE stars.

2 Mira Variable Mechanics

2.1 Evolution

Before a discussion of period calculation for Mira variables begins, a discussion on stellar evolution is useful. The aim of this discussion is not to provide a detailed description

of stellar evolution, but to illustrate where on the Hertzsprung-Russell (HR) diagram the Mira variables reside. For more rigorous detail, please see Carroll and Ostile, Modern Astrophysics, Ch. 13.

Mira variables are evolved stars with main-sequence masses roughly between 0.8-6 $M_\odot$. After the formation of a protostar, stars of any mass will enter the main sequence where hydrogen undergoes fusion into helium. As the hydrogen fuses, a helium core forms. Eventually hydrogen burning stops in the core, and for stars whose masses are less than 6 $M_\odot$ a shell of hydrogen fusion forms around the inert helium core. For stars with masses greater than 6 $M_\odot$, their mass and temperature support very deep convective cells in the star, preventing the helium core from forming. This is why the mass of Mira variables are less than 6 $M_\odot$, they require the inert helium core and fusing hydrogen shell for later steps of evolution [5].

Once the helium core forms and hydrogen fusion stops there, a hydrogen shell undergoing fusion forms around the core and the star enters the sub-giant branch (SGB) of the HR diagram. While in the SGB, the star sees a decrease in temperature and an increase in radius. This allows the abundance of H- ions to increase in the atmosphere which increases the opacity of the photo-sphere. Because of the opacity increase, deep surface layer convective cells form and the luminosity of the star increases as energy is carried out by the convective cells. The star is now considered to be a red giant and can be found in the red giant branch (RGB) of the HR diagram.

The next step of evolution toward a Mira star is when the red giant undergoes its first helium flash. The first helium flash is when the helium core ignites and undergoes fusion. The helium flash is caused when the conditions in the core of the star are appropriate to allow the triple alpha process to occur in the inert helium core. Mass is also being expelled out from the star into a surrounding circumstellar envelope. After the helium flash, the star enters the horizontal giant branch (HGB) of the HR diagram. Here an inert carbon core forms with shells of hydrogen and helium undergoing fusion.

After the helium flash, the star expands but then slowly contracts. This contraction causes the star to move once-more on the HR diagram to the asymptotic giant branch (AGB). In the AGB of the HR diagram some of the stars experience more helium shell flashes, expelling more matter into the circumstellar envelope. These are referred to as thermal pulses, and the star is considered to be in the thermal pulse asymptotic giant branch (TP-AGB) of the HR diagram. This is where Mira variables are found, with typical masses between 0.8-6 $M_\odot$ and typical luminosities of approximately 3000 $L_\odot$. [2]

2.2 Pulsation

There are two different sources of variability for Mira variables; thermal pulses and acoustic pulses. They occur on very different time scales. Thermal pulses generally occur on time scales of 10^4 years [2] while acoustic pulses occur between 100 and 1000 days [5]. Both are discussed here as they are both important to understanding Mira variables, but in practice the acoustic observations are those causing the periodical variations discussed in this thesis.

2.2.1 Thermal Pulsation

The thermal pulses originate from interactions between the hydrogen and helium burning shells which surround the inert carbon core. The helium shell, which is closer radially to the core, is not always experiencing fusion as the temperature and pressure is too low. The hydrogen shell is experiencing fusion and raining helium down onto the helium shell. The helium shell accumulates mass this way and its temperature/pressure increases. Eventually the temperature and pressure will be great enough for fusion to occur in the helium shell. Once the helium shell ignites, the hydrogen shell and the rest of the stellar atmosphere will expand greatly from the increased radiation pressure. The expansion is great enough that the hydrogen shell burning will cease as its pressure and temperature are too low to support

fusion. Eventually the helium shell burning will stop, and the star will contract. As the star contracts gravitational potential energy is converted to heat and the temperature increases. Eventually conditions become appropriate to allow for the hydrogen shell to undergo fusion and the process repeats itself.

During these thermal pulses mass is ejected into a circumstellar envelope surrounding the star. Because of the large convective cells found in TP-AGB/Mira variables, the ejected material has many heavier elements present. Specifically, there are various shells in the envelope containing SiO, OH, and H_2O where maser emission is observed. The envelopes also contain dust and thereby obscure optical light, making them difficult to observe in the optical.

2.2.2 Acoustic Pulsation

The observed variability in Mira variables and variable stars in general originate from standing sound waves resonating in the star's interior. The stars can oscillate in either radial or non-radial modes of pulsation. In radial pulsation models, different modes are possible for the standing waves. The most common radial mode for pulsating stars is the fundamental mode, where the matter of the star is all moving in the same radial direction. The radial mode for stars with much longer periods, such as Miras, is either in the fundamental or the first overtone, although this is not currently entirely clear and is a topic of ongoing research [2]. The first overtone is a mode where there will be a sphere of mass within the star which does not see any movement, and the matter on either side of this sphere is moving in opposite directions.

In general, pulsating stars act as thermodynamic heat engines. The force which drives the pulsations originates from the opacity of the stellar atmosphere. Under compression the opacity increases which causes the pressure to increase as well as the stellar radius. Under expansion the opacity decreases which causes the pressure to decrease as well as the stellar radius [9]. These interactions occur in the partial ionization zones of the star. For example, under compression, a portion of the work done further ionizes the atmosphere. This is the necessary valve effect for a heat engine model, where under compression the opacity must increase.

Recent work by Da-run and Li-cai (2013) [10] or by Wood (2007) [11] analyze non-radial modes of pulsation for red giant type stars, like Miras. The key difference in this approach is that the modeling techniques no longer assume hydrostatic equilibrium and concentric mass shells, and by definition do not uniformly move in and out at all points on the surface at the same time. These models still consider the pulsations a result of standing sound waves, but rather than looking at the various radial overtones, the oscillations are described by a set of p and f modes. The p modes represent the portion of the solution which deals with the restoring force due to pressure, and the f modes represent the portion of the solution dealing with the motion across the stellar surface. Both articles by Wood and Da-run and Li-cai consider not only various p and f mode solutions, but also look at the influence convective cells have on stellar pulsation. By varying the efficiency of convection, the pulsation period at a given luminosity will change because a majority of energy from the core of a TP-AGB is carried by convection. This changes where the partial re-ionization zones occur and the pressure solutions for the p modes.

3 The VVV Dataset

The long periods of Mira Variables make monitoring campaigns challenging and costly in terms of observing time. However, due to their long periods, monitoring with relatively sparse sampling (30 days or less often) is sufficient to derive their periods. This is the specialty of the Visible and Infrared Survey Telescope for Astronomy (VISTA) Variables in the Via Lactea survey (VVV). Beginning observations in 2006 the VVV survey aims

to catalog 528 square degrees including the bulge of the Milky Way and nearby disk at 5 IR wavelengths (Z, Y, J, H, and Ks Bands) in various epochs. The resulting catalog will contain approximately 25 observational periods over a few years time in the Ks band (2.2 μm) of 10^9 sources, with approximately 10^6 of these sources expected to demonstrate variability. The magnitude sensitivity of the catalog is expected to allow for sources down to 18^{th} magnitudes while still possessing a signal to noise ratio of at least 3:1. The positions of the sources are calibrated against the respective 2MASS positions, leading to very small errors ($<0.''1$) in positional measurement. [8]

The survey is conducted with the VISTA telescope, a 4.1 meter telescope, with the VISTA InfraRed CAMera (VIRCAM) near-IR camera attached. The area observed extends in Galactic coordinates from $-10^\circ < l < 10^\circ$ and $-10^\circ < b < 5^\circ$ in the bulge region and $-65^\circ < l < -10^\circ$ and $-2^\circ < b < 2^\circ$ in the disk [8]. The goal of this Honors Thesis is to investigate if periods can be systematically extracted for the BAaDE sources, and the VVV survey thus provides an excellent data base to start.

At this time, the thesis is based on the first and only multi-epoch Ks-band data release which represents less than 10% of the overall DR2 general catalog [8]. Once the full multi-epoch catalogs are released, improvements to the accuracy of the light curves may be made. In addition to the multi-epoch data, there is an associated main catalog with magnitude and position measurements in all five bands. The currently released multi-epoch catalog contains about 30% of the sources in the main catalog, primarily because only a fraction of all tiles were included in the first multi-epoch release [8]. The currently available multi-epoch catalog has on average five measurements per source, and a maximum of 22 measurements. The measurements tend to be grouped in two main populations, separated by approximately 250 days. The uneven distribution of data points is a result of the observing strategy of the VVV survey.

The VVV multi-epoch data may still contain calibration issues and as such the data must be treated cautiously. For example, about 5% of the sources have a data point approximately 300 days after the first observation with a magnitude difference of 5. This data was not considered as this may be a pipeline associated effect. Another example is that 5% of the sources in the multi-epoch catalog have data points separated in time by a few minutes but with a large magnitude difference of 5. This is unlikely a real phenomena but probably a result of a failing source detection algorithm for specific, bright sources where the detectors go into saturation. Such data were also not considered in our analysis.

4 Data Aggregation and Selection

The BAaDE catalog contains 28,000 sources selected based on their IR colors, chosen such that the presence of SiO masers is likely [12]. The original IR catalog used for the selection was the Midcourse Space eXperiment (MSX), which has a positional uncertainty of about $1''$-$2''$ [13]. The positions of the BAaDE sources have been cross-matched with the 2MASS catalog, providing positional uncertainties less than $0.''1$. A positional cross-match between the BAaDE and the VVV main catalog was therefore made utilizing the 2MASS positions.

For a VVV source to be considered a cross-match with a BAaDE source, we first required that both the declination and right ascension were no more than $3.''6$ different. This value was chosen based on the reported error in the data base, which is indicative of the pipeline-reduction error and not the true positional uncertainty. This results in some sources having multiple matches. This is resolved by then finding the absolute separation and discarding sources with separations greater than $5''$. If sources still have multiple matches, the one with the smallest absolute separation is considered.

Once the list of cross-matched objects in the main VVV catalog was obtained, the objects names were extracted and used to pull out the correct sources from the multi-epoch catalog (which did not contain positional information directly). This resulted in 618 unique sources. After selecting only sources which had more than 5 observations with data,

227 sources remained. Sources were then further selected by hand, removing objects that showed little variability, and unreliable data as discussed in Sect. 3. This left a final set of 90 Mira variables with multi-epoch VVV data, for which we performed period analysis (Sect. 5).

5 Period Calculation

The data selected from the VVV multi-epoch catalog is sparsely populated, with each object having on average 8.53 data points. Dworetsky (1983) discusses a method to find the period of randomly, and sparsely observed spectroscopic binaries. By plotting the data in phase space, the 'length of string' can be calculated, which is the length of the line connecting all points in phase space. The most likely period corresponds to when the length of string is minimized [14]. Here, we extend the work of Dworetsky (1983) by applying this method to sparse observations of Mira variables.

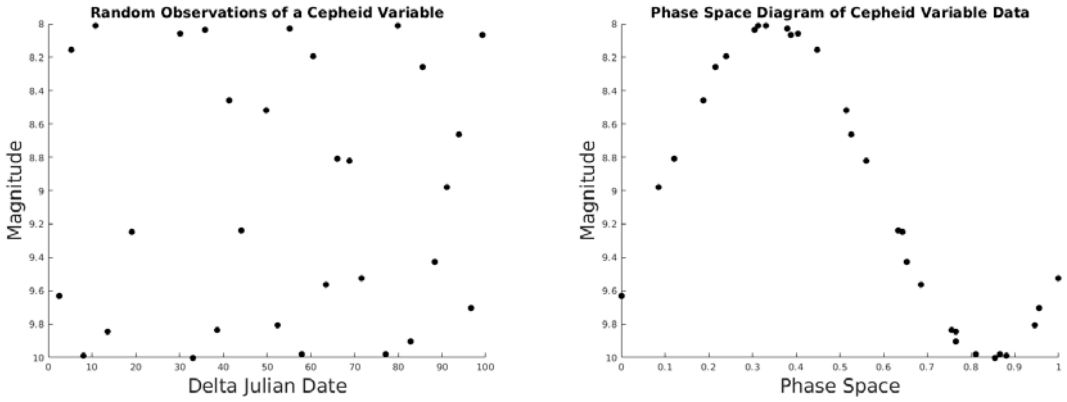
5.1 Process Description

To calculate the length of string, the data are first plotted in phase space, based on an assumed period. The period is constrained to be between 80-1000 days, covering the known period range for Miras [5]. For each of our 90 targets, phase space data were constructed within this period range with a step size of 0.092 days. The length of string was subsequently calculated. For each object, the period corresponding to the shortest length of string was recorded.

The string length method as discussed above does not include any information about the amplitude of the variations. To retrieve this information, a least-square fit algorithm was used, with the period as a fixed parameter and the amplitude varying between 0-10 magnitudes with steps of 0.04 magnitudes. A chi-squared method would be preferred, but the errors reported in the VVV catalog are not reporting observational error but rather fitting parameter error introduced in the data reduction pipeline. In addition, the reduced fit also fits for average brightness, with values between 0-20 magnitudes. After running the least-square tests using the string length calculated period, the fitted curve which provides the smallest value corresponding to the inaccuracy of the fit is assumed to be the correct period curve.

5.2 A Cepheid Example

To verify the accuracy of the method described in Sect. 5.1, a sample of observations of a Cepheid variable star with a known period of 6.28 days was used as a test case. In Figure 1a, the data is plotted in terms of amplitude as a function of time, and it is difficult to discern a period by eye. Figure 1b shows the same data, plotted in phase space for a period of 6.2952 days.



(a) *(b)*

Figure 1: Figure (a) is time series optical magnitude data as a function of date. Figure (b) is the data plotted as a function of phase. The phase plot is generated every time a new period is tested. The plot which returns the shortest distance possible between points is assumed to be the most accurate period calculation. This specific phase plot corresponds to a period of 6.2952 days.

Once the most likely period is calculated, the least-square fit finds the best fit for a sinusoidal curve with a period of X days, amplitude of Y magnitudes, and an average magnitude of Z magnitudes.

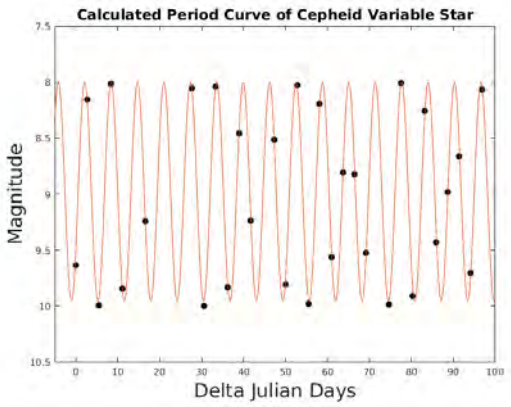


Figure 2: Plot of observations of the Cepheid variable with the most likely period and curve parameters generated by the reduced fit algorithm.

The previously determined period of 6.28 days is based on much more data. The relative error of using the string length method is 0.2%. This example therefore demonstrates that the string length method works very well when the time series data is well sampled. In Sect. 6, this method will be applied to more coarsely sampled data.

6 Results

Table 2 lists the calculated periods utilizing the method outline in Sect. 5, for 81 of the 90 cross-matched samples. The 9 sources whose periods could not be calculated, while they show variability, have to few data points for the string length method to calculate a meaningful period or have data which is a result of error in pipeline reduction. Additionally, 4 of the 81 periods calculated are considered to be an 'open fit' where the most likely period is more than 1000 days based on the data. Of the remaining 77 sources, 41 produced a good fit with the fitted curve providing a less than one standard deviation in the least-square fit value. The standard deviation is calculated from the fit parameters in the process characterization discussed in Sect. 8, assuming that the fit parameter follows a normal

distribution. These 41 sources will, in the following discussion, be considered to have accurate periods. Sect. 6.1 - 6.5 discusses a few specific objects, selected as to illustrate the diversity of the VVV data for our sample.

6.1 Peaks: ad3a-00210

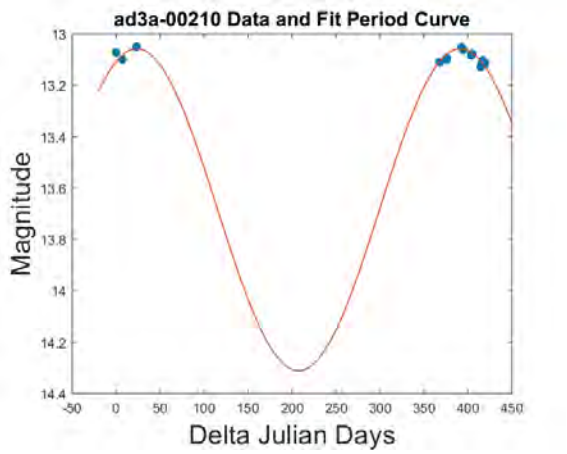


Figure 3: ad3a-00210 demonstrating the best fit of 'peak' data distributions.

The VVV data for the BAaDE source ad3a-00219 is an excellent example of a data set with the measurements focused around the peaks. 15 of the 90 sources in Table 2 have this data structure, flagged as 'P' in the Table. The prevalence of this time structure is likely due to human bias, as the 227 sources that were found to have more than 5 entries in the multi-epoch data were selected by hand (Sect. 4). It is not due to the VVV detector not having sufficient sensitivity. While overall it is poorly sampled, the maxima/minima corresponding to the peaks are generally very well sampled. The period calculation method is not impacted negatively by this time structure and can still reliably calculate periods for objects which demonstrate this pattern.

6.2 Well Sampled: ad3a-21113

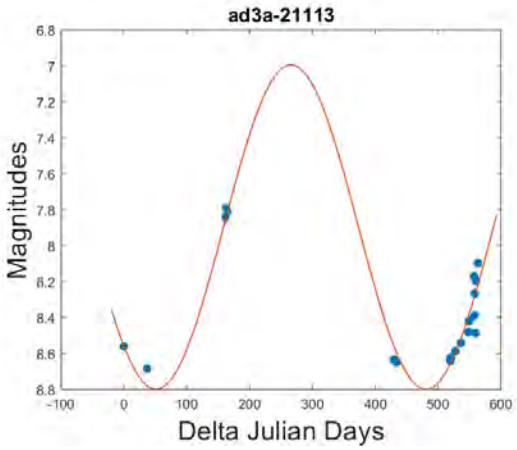


Figure 4: ad3a-21113 demonstrating well sampled data from the VVV catalog.

40 sources have more well sampled data, amongst them ad3a-21113. As most of the sources in the multi-epoch data do not have such well sampled time series, it is interesting to see this object as an example of what the future of the work with the VVV catalog may look like. ad3a-21113 has 19 observations with a good distribution along the time axis. Sources in

the final release of the VVV multi-epoch catalog will typically have at least 25 observations [8]. Here we can see that the calculated period and curve aligns well with the time series data, and should be considered an accurate period calculation.

6.3 Open Fits: ae3a-00450

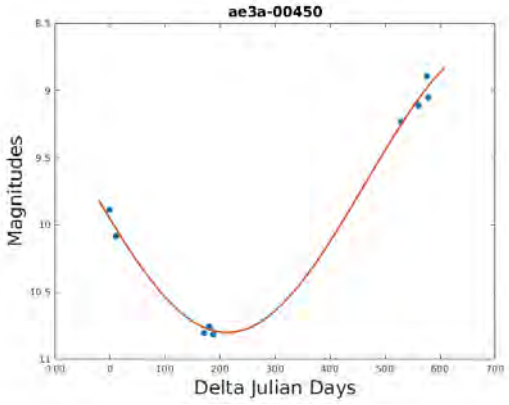


Figure 5: ae3a-00450 demonstrating an open fit from the VVV catalog.

Figure 5 demonstrates a trend found in four sources where there is clear variability, but there is not a long enough time sampling to sufficiently sample the light curve to derive a period. Therefore, for sources like this the string length method will always assert that the maximum value of the fitted period is the most likely one. This trend only occurs when less than 50% of the possible light curve is represented in the data. These sources are included in the analysis as when more epochs of the VVV survey are released with more data points, a more accurate period should be able to be extracted.

6.4 Variability with Bad Curve: ad3a-05288

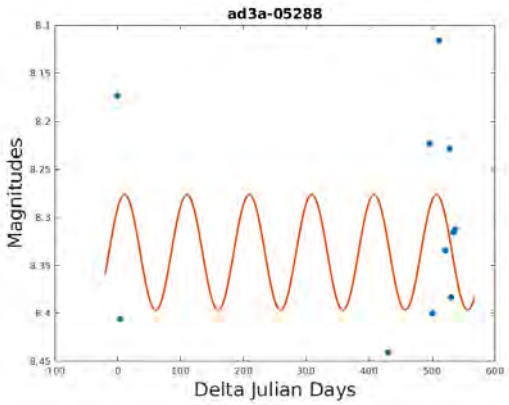
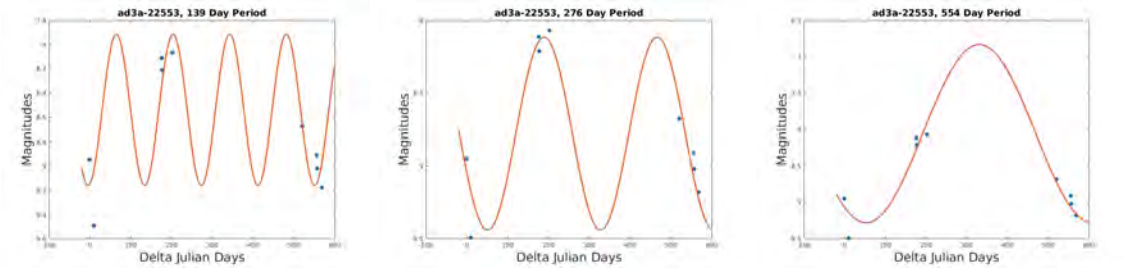


Figure 6: ad3a-05288 demonstrating variability with poor period curve from the VVV catalog.

In 15 cases, the reduced fit period curve calculation fails to produce a curve which agrees well with the data points, although the string length method calculates a closed period for the data. In cases where this occurs, the data also appears to be 'noisy'. Because of difficulties of estimating VVV data errors, discussed in Sect. 9.2, the data for sources like the one illustrated in Fig. 6 are currently considered unreliable.

6.5 Period Multiples: ad3a-22553



(a) *(b)* *(c)*

Figure 7: ad3a-22553 demonstrating a break down of the string length method, where multiple periods that are multiples of each other are all plausible solutions based on string length and least-square fit value.

In 3 cases, the string length method experiences difficulties determining what the most likely period is. For the case of ad3a-22553, the string lengths corresponding to the 139, 276, and 554 day periods are 3.0160, 3.0404, and 3.0717 respectively. These differences are very small compared to other sources, so each of the three potential periods were used to find the least-square fit period curve. The fit parameter for the three different periods are 0.4472, 0.1682, and 0.2196. Then, based on the fit parameter the 276 day period is selected to be the most likely period. This period calculation should not be considered completely confident though, as the difference in both the string length and fit parameter are very small between the 276 day and 554 day periods.

One characteristic about this phenomena is that it only occurs for multiples of the periods. For example, the 554 day period is approximately 4 times as long as the 139 day period and the 276 day period is approximately 2 times as long as the 139 day period. This phenomena is most likely due to undersampling of the data.

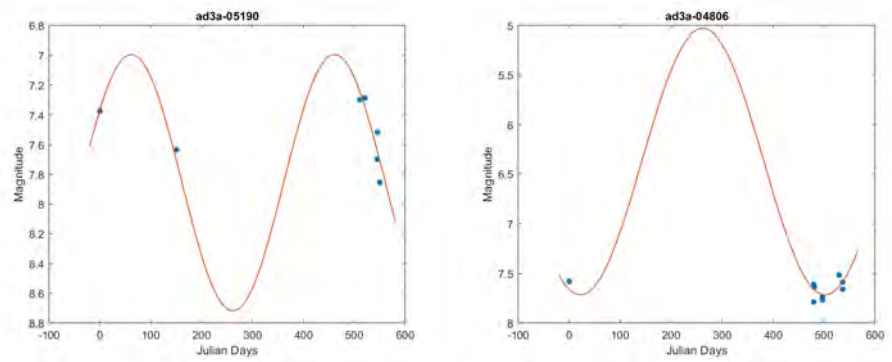
7 Comparing sources to AAVSO data

AAVSO (American Association of Variable Star Observers) are a group of both amateur and professional astronomers who focus on observing variable stars and making the results easily accesible to the public for scientific use. To further assess the accuracy of the string length method when applying it to the VVV data for the BAaDE sources, we cross-matched the 90 objects with the AAVSO data base [15]. Five cross-matches were found, and Sect. 7.1-7.2 discuss the results from the string length method applied to the AAVSO data and VVV data, respectively. Table 1 describes the results of the cross-matching.

Table 1: VVV Period Calculation Results

BAaDE	VVV Name	AAVSO Name	Est Period	AAVSO Period	Error
ad3a-05190	J175853.25-265932.62	OGLE-BLG-LPV-155144	399.7	436.6	0.084
ad3a-04806	J180359.85-273820.97	OGLE BUL-SC35 V438	479.8	474.0	0.012
ad3a-01994	J175749.91-300023.76	OGLE-BLG-LPV-143704	115.3	417.2	0.733
ad3a-06122	J174529.77-243031.99	OGLE-BLG-LPV-029519	NA	484.5	NA
ad3a-01441	J175814.08-305823.28	OGLE-BLG-LPV-148010	126.6	301.2	0.579

7.1 The good: ad3a-05190 and ad3a-04806

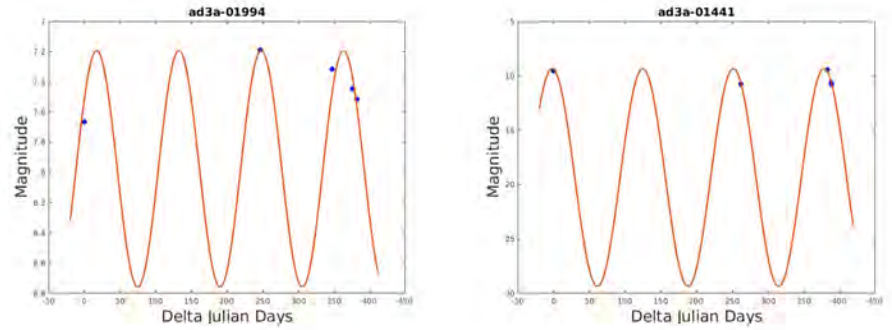


(a) ad3a-05190 *(b) ad3a-04806*

Figure 8: Two selected sources of VVV data with corresponding AAVSO data.

These two sources have VVV data sets (Fig. 8) with a data sampling pattern commonly occurring in the VVV data. This structure is characterized by most of the data taken approximately 500 days after the first observation. ad3a-04806 is also indicative of the known 'peaks' structure where much of the data is centered around minima/maxima of the stars variability (Sect. 5). Both sources have period curves which agree well with the time series data, and when compared to the AAVSO period, are within 9% error.

7.2 The bad: ad3a-01994, ad3a-06122, and ad3a-01441



(a) ad3a-01994 *(b) ad3a-01441*

Figure 9: Two selected sources of VVV data with corresponding AAVSO data. Note that these two sources are under sampled, and figure 9b represents a unphysical answer due to a very large amplitude.

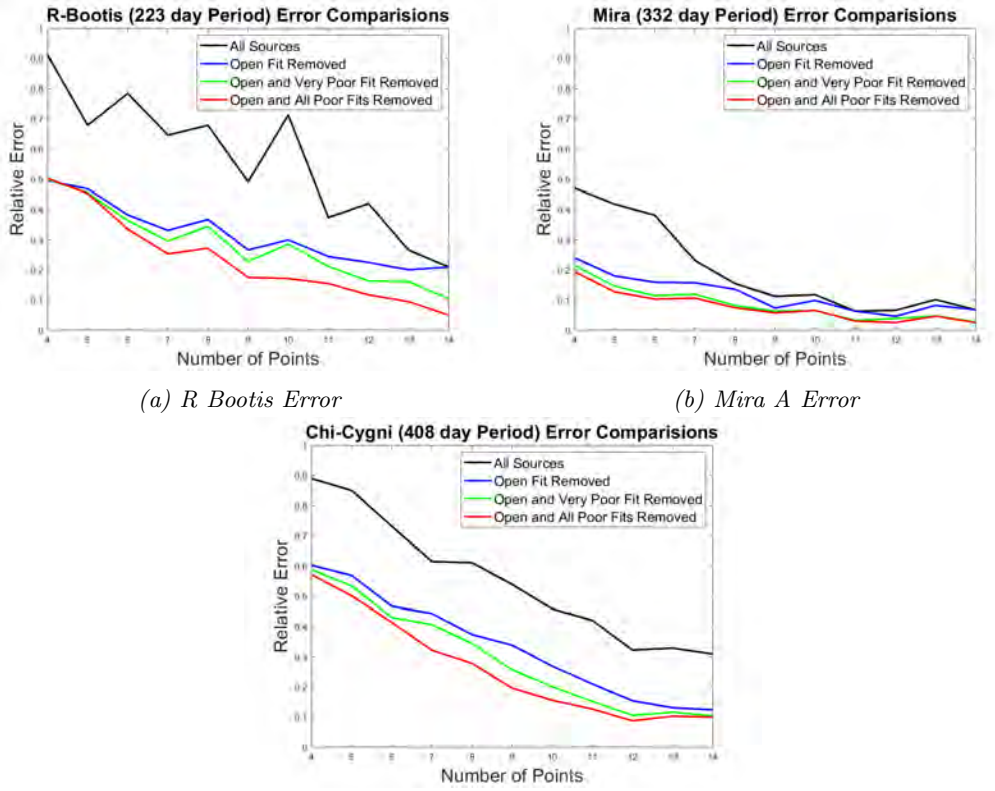
While the periods calculated for ad3a-01994 and ad3a-01441 are much different from the known AAVSO periods, both objects have very few observations which likely causes the difficulties of deriving the period. In the case of ad3a-01994, the fitted period matches well with the given data and the fit curve parameters are minimized, thus without the AAVSO data this period estimate would be adopted based on the string-length method results. Perhaps it could be argued that such a short period is only possessed by a minority of all Mira's, thereby requiring additional points for verification. Additionally, the sampling of data for this object may not be frequent enough to be able to calculate a reliable period, as discussed in Sect. 6.5. In the case of ad3a-01441, the fitted period also matches very well with the given data, but this fit can be considered inaccurate because the best fit provides a very large amplitude, which is unlikely based on its period (Sect. 9.1). In the case of ad3a-06122, the data from VVV is too coarsely sampled to produce a meaningful period.

This is because the largest difference in amplitude for the observations is 0.5 magnitudes, and that the difference in time of the observations is much shorter (<120 days) than the other cross-matched objects (>500 days). It is included here in order to demonstrate that the cross-match was successful for five objects.

8 Process Characterization

There is little overlap between the currently available VVV data base and AAVSO data, as the AAVSO is mostly optical and is restricted to brighter objects. Additionally, most AAVSO observers are in the northern hemisphere while the VVV is based out of the southern hemisphere. Because of this, it is not possible to directly compare our full sample to the periods estimated by the string-length method. Therefore, in this section we attempt to characterize how accurate the string-length method is based on the number of data points available. This is done using AAVSO data with a large number of data points available, and simply varying how many points are included in the period fitting. Three well known and well documented sources were selected to test against: R Bootis, Mira A, and Chi Cygni. These sources all have a rich set of V band observations in the AAVSO data base. Additionally, the three sources were selected as they have 223 day, 332 day, and 408 day periods respectively, which is typical of the range of periods most Mira variables have [5].

To simulate the conditions found in the VVV catalog, the data acquired for the three sources is within a 550 day range of observations, similar to the average range of dates for VVV observations. To better recreate the conditions of the VVV catalog, most observations taken in the middle of the range (200-400 days) were removed to mimic the two typical groups of observations found for sources in the VVV catalog. Then, using a uniform random number generator, observations were randomly selected from the AAVSO catalog. Tests using between 4 to 14 observations were performed, and for each value 100 different sets of randomly picked data points were generated. This results in a data base of 3300 tests.



(a) R Bootis Error *(b) Mira A Error*

(c) Chi Cygni Error

Figure 10: Results of three tests to quantify the accuracy of the period calculation using a small amount of data points. The black line shows the relative error using all 100 trials for each point value. The blue line shows relative error removing only trials who report periods of more than 1000 days. The green line shows relative error removing trials with least-square fit values greater than the mean plus one standard deviation. The red line shows relative error removing trials with least-square fit values greater than the mean.

From figure 10 it is seen that as the number of observations increases, the relative error of the period decreases quickly. Additionally, it demonstrates that the selection process performed as discussed in Sect. 5 provides an effective method for reducing the error of calculated periods for sources with less than 10 observations. There is an overall discrepancy between the magnitude of the relative errors between Mira A and the other two sources, R Bootis and Chi Cygni. Upon inspection of the reported errors from AAVSO the measurement errors tend to be less for Mira A when compared to the other two, and Mira A has fewer data outliers. This will cause the overall error of this specific testing regime to be lower as the process is sensitive to measurement error.

9 Errors

9.1 Nonsymmetric Curves

The initial tests of the string-length method described in Sect. 4-7 are promising. However, not only measurement errors will affect the accuracy of the period. Here we discuss an additional error source, associated with assumptions made in the period estimates. Most importantly, the period variation of Mira variables is not always symmetric. Fig. 9 illustrates three different light curve shapes for the Mira variables X Cam, Mira A, and Chi Cygni [16].

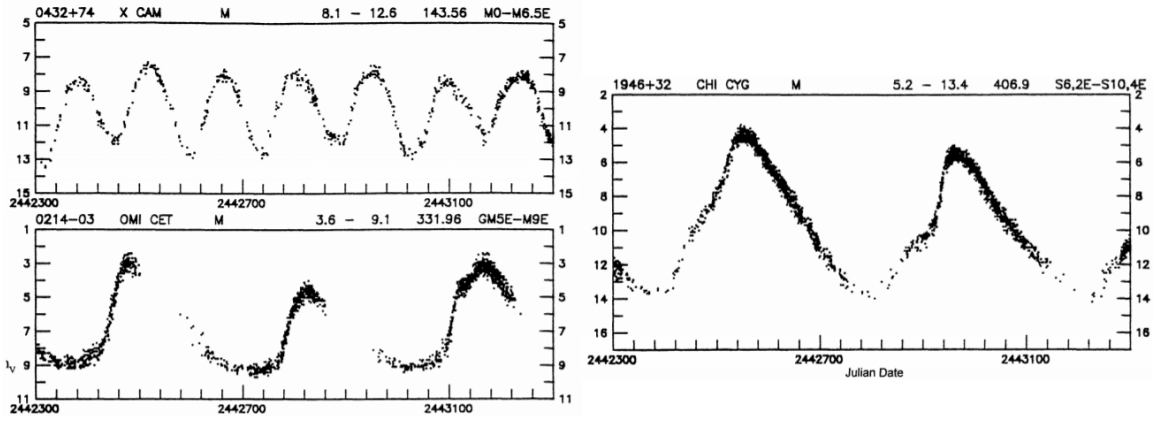


Figure 11: Period curves of X Cam (a. top left), Mira A (b. bottom left), and Chi Cygni (c. right), demonstrating the different shape of the period curves for Mira variables with different periods.

The graph of X Cam is more typical of Mira variables with periods less than 200 days, demonstrating symmetrical light curves with small amplitudes. Mira A is more typical of Mira Variables with periods between 200 and 350 days, having a much steeper rising branch of the light curve with a larger amplitude. Finally, Chi Cygni is more typical of Miras with periods greater than 350 days, having steep rising and falling branches, and occasional 'bumps' or standstills in the rising branch portion of the light curve. [16]

By using a more varied set of light curve shapes, more accurate periods may eventually be derived. Both the 'string length' method and the least-square fits assume that the light curves are perfect sinusoids, and do not consider deviations from perfect curves. This effect becomes less significant as more data points are acquired for the source, but for sources with few observations can perturb the calculated period by approximately 10% in worst case scenarios. Future work will investigate how to improve this systematic error by attempting to calculate periods and curve parameters by fitting nonperfect sinusoids to the data.

9.2 Measurement Errors

At this time, the full set of errors to be considered for the VVV data has not been established. The errors related to the magnitudes reported are generally very small, usually of the order of 0.01 magnitudes. This value is much to small to be true measurement errors, and are most likely errors involved in the fitting process from the data reduction pipeline. Sources which reported large errors in this parameter were not included in this study. Due to the lack of true measurement errors, traditional statistical methods such as chi-squared tests were not able to be performed.

10 Conclusions

The string length method combined with a least-square fit curve selection proves to be an promising method to extract periods of Miras from the VVV. The technique reliably extracts accurate periods for sources with more than 7 observations as shown in Sect. 8. Further confirmation for the method is demonstrated by cross checking the VVV sources with the AAVSO catalog in Sect. 7. Of the four AAVSO sources which period curves are derived, two are found to have calculated periods with errors less than 9%. By further characterizing the process using 3300 test samples of well observed objects in Sect. 8, it is shown that having at least 7 points provides a reliable fit.

The content of this study is not limited to the VVV survey. Future work may take these techniques described and apply to any future catalog. Specifically, this study may be done once more looking for Miras in the VVV once the full multi-epoch catalog is available. With the full multi-epoch catalog more Miras will be found in the VVV as only 20% of tiles are

included in the current multi-epoch catalog [8]. This will also increase the accuracy of the calculated periods as there will be significantly more data available in the full multi-epoch catalog. Currently a typical selected source has 8 observations while the final multi-epoch catalog will have 25 [8].

References

[1] L.O. Sjouwerman, Y.M Pihlström, R.M. Rich, M.J Claussen, M.R. Morris, and the BAaDE collaboration. Stellar SiO masers in the Galaxy: The Bulge Asymmetries and Dynamic Evolution (BAaDE) survey. *Proceedings IAU Symposium*, 2017.

[2] Bradley W. Carroll and Dale A. Ostlie. *An Introduction to Modern Astrophysics*. 2nd (international) edition, 2007.

[3] Z. Malkin. The current best estimate of the Galactocentric distance of the Sun based on comparison of different statistical techniques. *arXiv e-prints*, February 2012.

[4] Herbert J. Kramer. Gaia (global astrometric interferometer for astrophysics) mission, 2012.

[5] H.J. Habing. Circumstellar envelopes and asymptotic giantbranch stars. *The Astronomy and Astrophysics Review*, 7(2):97–207, Jun 1996.

[6] E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, T. N. Gautier, III, I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes, and C.-W. Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *Astrophysical Journal*, 140:1868–1881, December 2010.

[7] N. Epchtein. Denis: a deep near-infrared southern sky survey. *Experimental Astronomy*, 3(1):73–76, Mar 1994.

[8] R.K. Saito et al. VVV DR1: The first data release of the Milky Way bulge and southern plane from the near-infrared ESO public survey VISTA variables in the Vía Láctea. *Astronomy & Astrophysics*, 2012.

[9] Bradley W. Carroll and Dale A. Ostlie. *An Introduction to Modern Astrophysics*. 2nd (international) edition, 2007. Section 14.2, page 491.

[10] D.-r. Xiong and L.-c. Deng. A Theoretical Probe for Excitation Mechanisms of Sun-like and Mira-like Oscillations of Stars. *Chinese Astronomy and Astrophysics*, 37:1–7, January 2013.

[11] P. R. Wood. Convection and pulsation in red giant stars. In F. Kupka, I. Roxburgh, and K. L. Chan, editors, *Convection in Astrophysics*, volume 239 of *IAU Symposium*, pages 343–348, May 2007.

[12] S. Capen, L. Sjouwerman, and M. Claussen. MSX And IRAS Two-Color Diagrams For Late-Type Stars. In *American Astronomical Society Meeting Abstracts #213*, volume 41 of *Bulletin of the American Astronomical Society*, page 471, January 2009.

[13] Eddie Hilburn. Infrared properties of stars in the bulge asymmetries and dynamical evolution survey. 2018.

[14] M. M. Dworetsky. A period-finding method for sparse randomly spaced observations of 'How long is a piece of string?'. *Monthly Notices of the Royal Astronomical Society*, 203:917–924, June 1983.

[15] S. Kafka. Observations from the AAVSO International Database, 2019.

[16] J. A. Mattei. Introducing Mira Variables. *Journal of the American Association of Variable Star Observers (JAAVSO)*, 25:57–62, 1997.

Table 2: VVV Period Calculation Results

BAaDE Name	VVV Name	Median	Amp	Est Period	S Val	# Obs	Flags
ad3a-00794	J174732.21-324639.05	8.657	0.782	88.189	1.029	5	
ad3a-01184	J174209.61-313908.59	12.104	1.222	467.543	0.0751	8	A
ad3a-01189	J174714.43-313822.60	8.738	1.864	325.113	0.0592	5	P
ad3a-01441	J175814.08-305823.28	19.360	10.000	126.741	0.065	5	*
ad3a-01994	J175749.91-300023.76	7.976	0.782	115.332	0.057	5	*
ad3a-02246	J171740.91-294309.49	6.012	2.866	467.635	0.049	6	A,P
ad3a-02564	J174610.18-292509.03	6.693	2.786	193.999	5.253	8	P
ad3a-02607	J174613.69-292335.41	13.387	0.060	350.231	0.014	10	
ad3a-03347	J174726.28-285533.15	NaN	NaN	NaN	NaN	10	T
ad3a-04515	J174306.59-280135.01	18.397	10.000	493.673	0.284	6	P
ad3a-04754	J174411.06-274357.52	11.062	3.347	499.838	0.048	5	A,P
ad3a-04806	J180359.85-273820.97	6.373	1.343	479.780	0.049	9	A*,P
ad3a-04938	J173653.79-272620.35	7.134	0.501	101.070	0.017	5	A
ad3a-05190	J175853.25-265932.62	7.856	0.862	399.732	0.053	7	A*
ad3a-05230	J172757.81-265413.36	5.932	2.224	248.377	0.008	6	A,?
ad3a-05288	J174229.66-264549.86	8.337	0.060	99.138	0.103	11	
ad3a-05582	J175523.29-255956.74	NaN	NaN	NaN	NaN	10	T
ad3a-05918	J174443.11-250333.48	7.455	0.461	275.796	0.140	6	A
ad3a-06046	J173229.38-244331.64	10.261	0.301	475.732	0.050	7	
ad3a-06122	J174529.77-243031.99	NaN	NaN	NaN	NaN	6	*
ad3a-06154	J174332.67-242549.35	11.623	0.782	451.901	0.089	7	A
ad3a-06318	J181501.04-235858.17	8.056	0.661	461.930	0.037	6	A
ad3a-06485	J175815.02-233329.18	8.497	0.341	98.309	0.055	7	
ad3a-06566	J180741.03-232142.04	17.154	10.000	509.683	0.069	7	P
ad3a-06742	J180247.32-230556.22	14.990	0.100	491.097	0.034	11	
ad3a-06901	J181404.20-225149.99	11.102	0.741	158.576	0.079	6	A
ad3a-08084	J175501.56-192549.37	8.297	1.062	514.099	0.018	6	A
ad3a-08135	J180125.00-191358.48	10.460	1.383	445.921	1.110	13	
ad3a-18528	J132015.48-645407.29	14.869	5.150	106.314	21.690	9	
ad3a-18994	J141650.13-630258.44	6.893	0.782	164.925	0.109	8	P
ad3a-19083	J141533.53-625330.18	5.651	1.583	171.181	0.148	8	P
ad3a-19279	J130727.46-623333.83	12.104	2.786	1000	0.421	10	O
ad3a-19296	J125806.61-623208.60	8.096	1.503	474.167	0.015	6	A
ad3a-19648	J143427.61-614900.56	16.834	10.000	501.034	0.049	5	P
ad3a-19974	J131358.05-610959.93	11.503	0.220	434.788	0.006	7	A
ad3a-20113	J150111.92-605020.19	7.495	0.661	433.867	0.199	10	A
ad3a-20160	J131621.46-604314.46	8.498	0.862	441.780	0.173	6	A
ad3a-20529	J150754.64-595132.25	6.894	0.140	397.892	0.075	8	
ad3a-20592	J151451.52-594206.83	11.303	0.421	422.734	0.037	6	A
ad3a-20656	J145524.80-593210.18	7.776	0.301	472.327	0.055	11	A
ad3a-20675	J150731.51-592951.05	8.377	0.501	1000	0.811	22	O
ad3a-20741	J145457.92-591725.85	13.547	3.026	412.613	0.308	6	A
ad3a-20924	J151002.30-584024.32	11.503	0.180	358.972	0.280	12	
ad3a-20943	J150637.36-583723.50	8.297	0.381	429.891	0.368	10	A
ad3a-21046	J150926.74-581927.63	7.495	0.180	88.741	0.047	7	A
ad3a-21107	J151159.72-580940.07	8.537	2.064	138.886	0.083	6	A,P
ad3a-21113	J151513.46-580850.29	7.896	0.902	429.911	0.123	19	A
ad3a-21218	J150343.81-575059.64	6.894	0.341	362.100	0.241	7	
ad3a-21359	J152414.21-571850.28	9.058	0.501	476.192	0.448	12	A
ad3a-21469	J153436.93-565705.87	10.421	1.102	429.911	0.327	12	A
ad3a-21535	J153917.86-564103.09	9.499	1.824	394.948	0.476	13	A

Table 2: VVV Period Calculation Results

BAaDE Name	VVV Name	Median	Amp	Est Period	S Val	# Obs	Flags
ad3a-21803	J153049.81-553002.53	7.535	0.100	251.229	0.017	10	A
ad3a-21942	J152644.51-545614.88	7.214	0.822	505.727	0.145	12	A,?
ad3a-21982	J153748.97-544557.55	7.375	0.461	324.653	0.385	8	A
ad3a-21997	J153726.93-544335.70	7.015	1.102	393.843	0.232	7	A
ad3a-22014	J154640.98-543842.20	6.774	0.301	1000	0.016	6	O
ad3a-22113	J155139.25-541850.61	10.862	0.301	160.692	0.082	9	A
ad3a-22148	J153850.98-541143.24	8.898	3.828	434.972	50.394	7	
ad3a-22279	J160129.29-534017.51	12.473	1.371	512.811	0.338	6	A
ad3a-22420	J160608.38-530813.79	8.377	0.701	508.671	0.139	9	A
ad3a-22553	J161017.58-523635.50	8.778	0.661	276.440	0.168	9	A,?
ad3a-22907	J161020.23-511449.90	NaN	NaN	NaN	NaN	6	
ad3a-23026	J160730.43-504546.96	12.866	2.144	335.602	0.077	8	A,P
ad3a-23124	J161847.97-502308.81	9.178	1.463	108.707	17.889	16	T
ad3a-23261	J160927.06-494742.23	11.022	0.301	316.004	0.439	8	
ad3a-23595	J161613.80-482326.13	8.417	0.581	177.990	0.218	14	A
ad3a-23606	J161726.87-482105.69	7.736	0.220	428.347	0.107	9	
ad3a-23672	J162438.64-480644.61	9.659	0.301	91.685	0.732	8	
ad3a-24131	J162841.39-464734.16	8.176	1.463	392.739	0.156	6	A
ad3a-24642	J164820.86-452553.66	10.862	0.541	432.119	0.775	8	
ad3a-24665	J164853.11-452144.56	9.178	0.381	130.145	0.099	7	
ad3a-24683	J165023.94-451628.53	10.727	0.576	408.197	0.021	5	A
ad3a-25501	J170917.03-413541.23	12.665	1.222	107.695	0.050	7	
ad3a-25514	J170521.87-413142.02	11.263	0.501	137.046	0.142	7	
ad3a-25618	J170602.80-405944.61	9.820	0.461	169.249	0.278	14	
ad3a-26204	J170425.30-382636.21	8.096	0.902	361.824	0.133	6	
ad3a-26265	J171421.98-381030.73	7.695	0.782	332.013	0.055	7	P
ad3a-26328	J171702.84-375635.73	NaN	NaN	NaN	NaN	11	T
ad3a-26336	J170537.67-375433.48	11.984	0.782	367.253	0.113	8	A
ad3a-26537	J172805.62-370729.50	8.457	0.902	376.914	0.056	8	A
ad3a-26591	J172319.13-365441.77	NaN	NaN	NaN	NaN	16	T
ad3a-26601	J172333.75-365123.69	NaN	NaN	NaN	NaN	16	T
ad3a-26842	J172633.26-355931.87	10.782	0.421	231.815	1.219	16	T
ad3a-26925	J171851.29-354417.57	13.788	4.549	341.950	4.602	7	P
ae3a-00437	J151400.80-591508.67	11.022	0.180	373.969	0.138	10	
ae3a-00450	J155604.18-535044.07	9.6994	1.102	1000	0.027	10	O
ce3a-00002	J172549.40-342020.59	NaN	NaN	NaN	NaN	10	
ce3a-00302	J155433.15-535129.89	NaN	NaN	NaN	NaN	10	T

*Table 3: Flag Legend: A is considered an accurate period. P represents sources with 'Peak' data structure. O represents sources with open fits. T represents sources with the 'tuple' data phenomena. ? represents sources where it is difficult to determine which period multiple is correct. * represents sources also found in AAVSO*