Draft version September 3, 2021
Typeset using LaTeX **twocolumn** style in AASTeX631

Characterizing The Variability of Long Period Variable Stars in the r $\leq$ 2 kpc Solar Neighborhood With Gaia and ZTF.

Katiya C. Fosdick [1,2]

Mentors:

Luis Henry Quiroga Nuñez [2,3] And Lorant O. Sjouwerman [2]

[1] University of Wisconsin-Madison
[2] National Radio Astronomy Observatory (NRAO)
[3] University of New Mexico

ABSTRACT

Periods of variability are a key characteristic of long-period variable (LPV) stars. Periods are potentially related to an LPV's initial mass, mass loss rate, metallicity, circumstellar envelope thickness, and luminosity. We develop fitting algorithm combining Jiménez-Esteban et al. (2021) style least squares regression fits and the Lomb Scargle Periodogram to fit periods to 36,000 ZTF-Gaia crossmatched LPVs in the solar neighborhood (r $\leq$ 2). We take advantage of the normalization of Gaia DR2 and ZTF DR3 light curves and long time baselines to reduce the errors on our fit periods. We successfully fit periods to 93% of our sources, and identify a sample of sources with periods consistent to Gaia published frequencies of variability. We use a sample of BAaDE-Gaia-ZTF crossmatched LPVs with consistent periods to reproduce a log period - k-band apparent magnitude (PLR) relationship in the solar neighborhood. We find the PLR for LPVs in the solar neighborhood is flat, and hypothesize that the solar neighborhood PLR is dependent on a LPV's metallicity and circumstellar shell thickness.

Keywords: Long Period Variables — Periods — Stellar Astronomy — Stellar Evolution — TP-AGBs — Period-Luminosity Relation— Solar Neighborhood — Gaia — Zwicky Transient Facility — BAaDE — Lomb Scargle Periodogram — Least Squares Regression Fit — Sinusoidal Fit — Period Fitting Algorithm — Variability

1. INTRODUCTION

Thermally Pulsing Asymptotic Giant Branch (TP-AGB) stars with initial masses between 0.5 to 8 $M_\odot$ (Höfner & Olofsson 2018). TP-AGBs are a class of evolving giant stars defined by their location at the the top of the Asymptotic Giant Branch on the Hertzsprung-Russell Diagram. TP-AGB stars, that are fueled by nuclear fusion in their hydrogen and helium shells, are characterized by high mass loss rates of 10^{-8} to $10^{-4}\frac{M_\odot}{yr}$ (Pols 2012).

Over the course of a TP-AGB star's time on the AGB, the star will lose 50-70 % of its mass (Wood et al. 2004). The dust and gas expelled by TP-AGBs form a diffuse circumstellar envelope around them that can extend several hundred $R_\odot$ (Pols 2012). TP-AGB stars are described for abundant molecules found in their circumstellar envelopes, such as oxygen-rich TP-AGBs or carbon-rich TP-AGBs. TP-AGB stars are present in the solar neighborhood ($r \leq 2$ kpc), at a conservative density of about 25 stars/kpc (Jura & Kleinmann 1989).

TP-AGB stars are also defined by their variability in luminosity. TP-AGB stars have been characterized as 'unstable,' but their variation in brightness is better described as cyclic fluctuation between the influence of radiation pressure and gravity on the star. These fluctuations are caused by thermal pulsing as helium burns in the star (Pols 2012). The changing balance between these forces causes the circumstellar envelope of the TP-AGB star to regularly pulsate. This causes the luminosity of a TP-AGB to vary cyclically with some period of variability. TP-AGBs are often classified using the length of their period of variability. TP-AGBs with peri-

kfosdick@wisc.edu

ods longer than about 100 days (50 days conservatively) are classified as long-period variable stars (LPVs).

A LPV's period of variability is hypothetically related to several physical characteristics of the star. Pols (2012) and Höfner & Olofsson (2018) suggests that a LPV's period is strongly related to the star's initial mass and its mass loss rate. Additionally, Höfner & Olofsson (2018) and Dr. Sjouwerman (priv. comment) suggest that a LPV's period may be related to the thickness of the star's circumstellar envelope and it's metallicity. One of the most well studied and well known relationships is that between a LPV's period and its average luminosity.

The period-luminosity relationship or log period-absolute magnitude relationship, both of which we will refer to as the PLR, have been historically studied in the Large and Small Magellanic Clouds (LMC and SMC): low metallicity (Z $\approx$ 0.5 solar) environments where the distances to observed stars are known, and the effects of dust extinction can be assumed to be the same. Whitelock et al. (1991), Matsunaga & IRSF/SIRIUS Team (2007), and Whitelock et al. (2008) show well fit log period - k-band absolute magnitude relationships for Miras[1], a type of LPV, in the LMC. These relationships are critical in characterizing the populations of LPVs in a given galaxy. The PLR is well studied because it allows the luminosities, and thus distances to objects, to be extracted from periods fit to photometric light curve data. If a PLR existed for LPVs in the solar neighborhood, we could use LPVs as standard candles to extract distances for nearby objects in the Milky Way. However, as reported by Ita et al. (2004) and Quiroga-Nuñez et al. (2020), the log Gaia period - k-band absolute magnitude relationship for Oxygen-rich LPVs in the solar neighborhood is flat.

The Quiroga-Nuñez et al. (2020) PLR consists of sources from their BAaDE-Gaia sample. The BAaDE sample, described in Sjouwerman et al. (2018), consists of 28,000 LPVs that are identified as oxygen rich TP-AGBs through their positions on the IRAS two color–color diagram (2CD) (Quiroga-Nuñez et al. 2020). The identification of SiO masing molecules in 70% of these stars corroborates their classification as TP-AGBs. Quiroga-Nuñez et al. (2020) crossmatched TP-AGBs in the BAaDE Survey and 2nd Gaia Data release (DR), Gaia DR2. BAaDE-Gaia sample sources are constrained to the Milky Way's disk and are LPVs with oxygen-rich circumstellar envelopes. Quiroga-Nuñez et al. (2020) calculated K-band absolute magnitudes for the BAaDE-Gaia sample from Two Micron All Sky Survey (2MASS) IR photometry and Capitanio et al. (2017) and Lallement et al. (2019) extinction maps. The periods used in the Quiroga-Nuñez et al. (2020) PLR are calculated by inverting frequencies of variability fit to photometric light curve data by Gaia. It is a well known problem that frequencies with errors larger than 20% of their value can not be inverted to produce accurate periods. We calculate Gaia periods for BAaDE-Gaia sources where this inversion is possible, when the frequency error published by Gaia for a source is less than 20% of the published frequency. Regardless, some Gaia periods have large errors and inaccuracies caused by effects like aliasing. To truly study the PLR for BAaDE-Gaia LPVs in the solar neighborhood, we need to identify a sub-sample of BAaDE-Gaia sources with accurate Gaia periods with relatively small errors.

We develop an algorithm to fit periods to BAaDE-Gaia source light curves to confirm and dispute periods published by Gaia. We take advantage of Zwicky Transient Facility (ZTF) light curve data from the 3rd ZTF data release (DR3) available for a selection of BAaDE-Gaia sources to add more data points to Gaia DR2 light curves, and improve the accuracy of our fits. In designing our algorithm, we take advantage of two well-tested algorithms. The Lomb-Scargle periodogram (LSP), thoroughly described in VanderPlas (2018), is a power spectrum that identifies the frequencies of sinusoidal signals present in unevenly sampled light curve data. The frequencies can be inverted to produce a period of variability. Additionally, we take advantage of a least squares regression (LSR) sinusoidal fit and error formula published by Jiménez-Esteban et al. (2021), which uses long time baselines to reduce the error on fitted periods.

2. SAMPLE

This project involves two sub-samples of Gaia DR2 long period variables (LPVs). These samples were created with the purpose of testing our period fitting algorithm and investigating the relationship between the log-period and k-band absolute magnitude relationship in the solar neighborhood. The larger of our two samples is a sample of 36,000 Gaia DR2 stars that are classified by the Gaia survey as long period variables. These sources were crossmatched with ZTF DR3 objects. Both surveys provide light curves for each star in several optical bands. Of these 36,000 sources, 17,548 or 49% have published Gaia DR2 fit frequencies of variability. 7,362 or

[1] LPVs include two subclasses of TP-AGB stars: Mira and Semi regular variable stars. Thin-shelled LPV stars that are visible in optical bands and experience pulsations larger than 2.5 mag in the V-band are classified as Mira variables or Miras. Miras have longer periods of variability than semiregular variable stars.

20% of these sources have frequency errors that are small enough that their Gaia fit frequency can be inverted to extract a Gaia period. We will refer to this sample of sources as our "ZTF-Gaia Crossmatched" sample.

The second sample of interest, a subsample of the BAaDE-Gaia sample, is a 137 source subsample of the 36,000 ZTF DR3 and Gaia DR2 cross listed long period variables and the BAaDE sample. BAaDE-Gaia sources are characterized by their good Gaia DR2 parallax values, which allow accurate distances to these stars to be calculated. Additionally, BAaDE-Gaia stars are crossmatched with 2MASS survey and Capitanio et al. (2017) and Lallement et al. (2019) paper sources, which provide k-band IR photometry and extinction maps. IR photometry and extinction maps are used to calculate k-band absolute magnitudes for each sample star through methods described in Section 3.4. 86 or 63% of these stars have published Gaia DR2 fit frequencies of variability. 49 sources or 36% of the sample have fit frequencies that can be inverted to extract a Gaia period. We will refer to this Gaia-BAaDE-ZTF crossmatched sample of sources as our "BAaDE-Gaia" sample, though it is truly a subsample of the original BAaDE-Gaia sample defined in Quiroga-Nuñez et al. (2020), which is not crossmatched with sources in ZTF.

3. METHODS

3.1. *ZTF and Gaia Data Crossmatching*

We used two different cross matching methods to form the ZTF-Gaia Crossmatched and BAaDE-Gaia samples. The Gaia survey identified 150,000 LPVs from over 1.6 billion sources in Gaia DR2. We used a 3" radius to successfully crossmatch 36,000 of these Gaia-Classified LPVs with ZTF DR3 objects, to form the ZTF-Gaia Crossmatched sample. We applied the same cross matching process to Quiroga-Nuñez et al. (2020)'s BAaDE Gaia Sample to form our 137 source ZTF-crossmatched BAaDE-Gaia sample.

3.2. *Period Fitting Algorithm*

To fit periods to our sample of variable LPV stars, we designed a period fitting algorithm that combines least square regression (LSR) fitting and Lomb-Scargle Periodograms (LSP) through the use of the astropy.timeseries.LombScargle and scipy.optimize python classes. Applying our algorithm to Gaia DR2 and ZTF DR3 light curves, and a light curve produced by normalizing and combining data from both surveys, we take advantage of more data points and the long time interval between the collection of Gaia DR2 and ZTF DR3 light curves, known as time baselines, to produce relatively small errors on our calculated periods.

To fit periods to Gaia DR2 and ZTF DR3 light curves, we first must select light curves collected in several available bands published by both surveys. ZTF DR3 published light curve data in the i (702.710 - 888.374 nm), g (408.668 - 552.196 nm), and r (560.040 - 731.669 nm) bands (Rodrigo et al. 2012). Gaia DR2 published light curves in one wide G band (330 - 1050 nm), as well two narrow bands produced by separating the G band into its red and blue parts: the Gaia DR2 RP ("red part", 630 - 1050 nm) and BP ("blue part", 330 - 680 nm) bands (Evans et al. 2018).

Physically, a LPV star will vary in brightness differently when observed in different bands. A LPV may shine at its peak luminosity in one band several days or weeks before this peak luminosity is observed in another band. A time shift would be required to normalize and accurately compare two light curves of the same object in dramatically different bands, as this light is produced in different regions of one LPV star. Our period fitting algorithm relies on normalizing and combining ZTF DR3 and Gaia DR2 light curves to form a single light curve, to which we fit a period. To avoid accounting for a time shift, we opted to normalize and combine ZTF DR3 and Gaia DR2 data in similar bands only. By default, we fit a period of variability using the ZTF DR3 r band and Gaia DR2 RP band light curves, which are redder bands where any time shifts in variability present will be less pronounced. We fit periods to these light curves as long as there were more than 5 data points in the r-band light curve. If this condition was not satisfied, but the g-band light curve for the same star had more than 5 points, we fit a period of variability using the ZTF DR3 g-band and Gaia DR2 BP-band light curves. If neither of these conditions were met, we did not attempt to fit a period of variability to the source and flagged these stars with the flag "too few points fit not attempted" in the Fosdick_2021_LPV_Periods table published in Appendix B.

The core method used in our period fitting algorithm is a least-squares regression sinusoidal fit inspired by period-fitting methods described in Jiménez-Esteban et al. (2021). The sinusoidal function fit to our light curve data is described in Equation 1 The scipy.optimize.curve_fit() function allows for the mean amplitude, M, amplitude A, time shift, t_0, and period of variation, P, parameters in Equation 1 to be fit to light curve data. Equation 2, reproduced from Jiménez-Esteban et al. (2021) is used to produce an error, ΔP on the fitted P parameter. Here, t_{final} and $t_{initial}$ are the times of the first and last points of our fitted light curve. As shown from the functional form of Equation 2, the error on our fit period P will be minimized as the length of our fit period of variability decreases and if the time difference between the first and last collected points in the fitted light curve is large. Generally, P will be minimized as the number of points in the fitted light curve increases, allowing a better fit to be produced.

$$m(t) = M + A\sin(\frac{2\pi(t-t_0)}{P}) \tag{1}$$

$$\Delta P = \frac{P^2}{2(t_{final} - t_{initial})} \tag{2}$$

To increase the accuracy of our LSR fit periods and avoid complications caused by aliasing, we took advantage of a feature of the scipy.optimize.curve_fit() function which allows user provided 'first guesses' of the values of the fitted M, A, t_0, and, P parameters to influence values fit by the function. Assuming that a fitted light curve is approximately sinusoidal in form, the first guess parameters M, A, and t_0 are relatively straightforward to calculate directly from light curve data, as done in Equations 3-5, where m is an array of the apparent magnitudes of our light curve, and t is an array of the times at which these apparent magntude readings were collected. To calculate a first guess fitted period, we produce a Lomb-Scargle Periodogram of the light curve being fit for a period. We removed signal frequencies of variation returned by the LSP larger than $f = 0.02\ \frac{1}{\text{days}}$ under the conservative physical assumption that LPVs will not have periods smaller than 50 days. The remaining frequency associated with the maximum power value in the LSP, f_{LSP}, is inverted to produce our first guess period, as shown in Equation 6. Various first guess periods are included in multiple columns of the Fosdick_2021_LPV_Periods table.

$$M^* = mean(m) \tag{3}$$

$$A^* = \frac{max(m) - min(m)}{2} \tag{4}$$

$$t_0^* = min(t) \tag{5}$$

$$P^* = \frac{1}{f_{LSP}} \tag{6}$$

The following steps are used to fit a period to each sample LPV for which we attempt to fit a period using either the r and RP band light curves or g and BP band light curves. First, we use a LSP and Equations 3-6 to produce first guess parameters for both the ZTF DR3 and Gaia DR2 light curves separately. Then, these first guess parameters are used with the scipy.optimize.curve_fit() function to fit the sinusoidal function in Equation 1 to the LPV's ZTF DR3 and Gaia DR2 light curves separately to produce a set of fit M, A, t_0, and P parameters for each light curve. These fit parameters are used to normalize the ZTF DR3 and Gaia DR2 light curves, by applying Equation 7 to each individual point in the light curves. Once normalized, the data from the Gaia and ZTF light curves are combined to form a final normalized light curve. Equations 3-6 are used to return first guess parameters for the normalized light curve. Finally, the LSR fit of Equation 1 is applied to the normalized light curve to produce the final fitted period returned by fit parameter P. Equation 2 is used to provide an error on this fitted period. This process is reproduced visually for Gaia source 2207771934780480000 in Figures 1a - 1h and Figure 2. The period fit to source 2207771934780480000 is given in Table 1. These figures and the fitting parameters produced during the implementation of our period fitting algorithm can be reproduced for any source in our sample by running the PeriodFit() and PeriodFancyFit() functions included in Appendix A.

$$m_{Norm}[i] = \frac{m[i] - M}{A} \tag{7}$$

3.3. *Known Problems*

In the development of our period fitting algorithm, we tested our algorithm on the ZTF-Gaia Crossmatched sample and several other sources in the Gaia Archives. We compared our fit periods of variability to extracted Gaia periods and found several problems with our period fitting algorithm. We will present examples of these problems here. Further study will be needed to identify all problems present with our period fitting algorithm and the full extent of these problems in our published sample of LPV periods.

One problem became evident when several of our sources were fit nonphysical periods of hundreds of thousands of days by our period fitting algorithm. We believe that these sources may have been mislabeled as LPVs by Gaia, and may not be variable stars. One of these stars was Gaia DR2 Source 2021849536746341376. Additionally, the Gaia DR2 light curve of source 2021849536746341376, shown in Figure 3b, may to contain poor quality data. Regardless, we expect our period fitting algorithm to only be able to fit periods approximately up to four times longer than the length of the time baseline of a source's normalized light curve.

Aliasing is a particularly well known problem of LSR fitting algorithms and the LSP (VanderPlas (2018), Jiménez-Esteban et al. (2021)). Aliasing describes the phenomenon where under or over sampled light curve

Fosdick 2021 Fit Period and Error	Gaia Period and Error
170 ± 7 days	175 ± 32 days

Table 1. Period fitting results for Gaia DR2 source 2207777193478048000

data of a variable source can be fit well to several sinusoids with relatively small errors. This effect causes several periods of variability to be accurate to our data, that are often integer multiples or rational fractions of each other. Physically, a LPV star will only have one primary period of variability, and potentially, a few undertone frequencies, at which the star will pulsate (Wood 2000).

An example of aliasing is presented by Gaia source 4979075051323208832. This source is not cross listed in ZTF DR3 data release and is not found in the ZTF-Gaia Crossmatched or BAaDE-Gaia Samples. Rather, this was a source in the Gaia Archives for which we were able to extract a period with a relatively small error of 65 ± 1 days. We intended to use this source to check that our period fitting algorithm could reproduce results obtained by Gaia fitting. However, when we fit the Gaia DR2 RP-band light curve of this object, we found a non-consistent period of 280 ± 64 days. Our period was corroborated by the LSP, which we used to extract a period of 268 days, though we did not develop a method to extract errors on periods fit by the LSP. We present this light curve as a potential example of aliasing in the fit frequencies of variability published to the Gaia Archives. Evidence of aliasing between periods extracted from Gaia and periods published by the AAVSO may suggest that aliasing is prevalent in the Gaia Archives.

To avoid aliasing in our fit periods, we applied a physical frequency cut of $f = 0.02\ \frac{1}{\mathrm{days}}$ to the Lomb Scargle Periodogram before extracting our first guess fit parameters. Before this was done, we noticed a significant sample of LPVs that were fit to periods of 20 days, which is physically improbable. However, even with this physically justified frequency cut in place, the LSP is susceptible to the effects of aliasing. Though our fit period and Gaia period are consistent with each other for Gaia source 4138630692585295744, Figure 5 demonstrates the potential for aliasing in the LSP that provides our first guess fitting parameters. Here, our normalized light curve is well fit by periods of about 70 and 140 days. We expect aliasing to be one of the most prevalent problems with our period fitting algorithm. Figure 6 suggests that aliasing impacts our fitted periods or Gaia periods for a large portion of ZTF-Gaia Crossmatched sources with Gaia periods.

3.4. *K-Band Apparent Magnitudes*

K-band absolute magnitudes are calculated for BAaDE-Gaia sample sources from 2MASS IR photometry data and some survey extinction maps. We use a Python function developed by Quiroga-Nuñez et al. (2020) to execute this calculation.

3.5. *Majority and Outlier Sample Selection*

Our majority samples are characterized as samples containing 66% of the total sources around the 1:1 ratio line formed when we compare our fit periods and Gaia periods. Sources in these majority samples were selected by forming two $m = 1$ slope parallel lines, Equations 8 and 9, around the 1:1 ratio line and manually adjusting the b intercept until 66% of the sample was contained between these lines. The 34% of sources that did not fall within these lines were placed into our outlier samples. A Python function, MajorityCut(), that automates this selection process is included in Appendix A.

$$y = x + b \tag{8}$$

$$y = x - b \tag{9}$$

4. KEY RESULTS

Applying our period fitting algorithm to each of our samples described in Section 2, we fit periods to 33,492 or 93% of our ZTF-Gaia Crossmatched sample sources and 132 or 96% of our BAaDE-Gaia sample Sources. We have successfully fit periods of variability to sources for which Gaia did not report a frequency of variability. We recognize that problems with our fitting algorithm, identified and discussed in Section 3.3, will produce outlier periods of variability such that our fit period for a small selection of these sources may be inaccurate. By normalizing and combining ZTF DR3 and Gaia DR2 into a single light curve with more data points than Gaia DR2 light curves, our period fitting algorithm should successfully increase the accuracy of fitted periods. Generally, the long time baseline produced in our normalized and combined light curves successfully reduced the errors reported on fit periods.

In addition to publishing combined light curve fit periods of variability and their errors for each source in the Fosdick_2021_LPV_Periods table in Appendix B, we publish fitting parameters, data statistics, and intermediate periods fit over the course of our period fitting algorithm. We publish several first guess periods, which

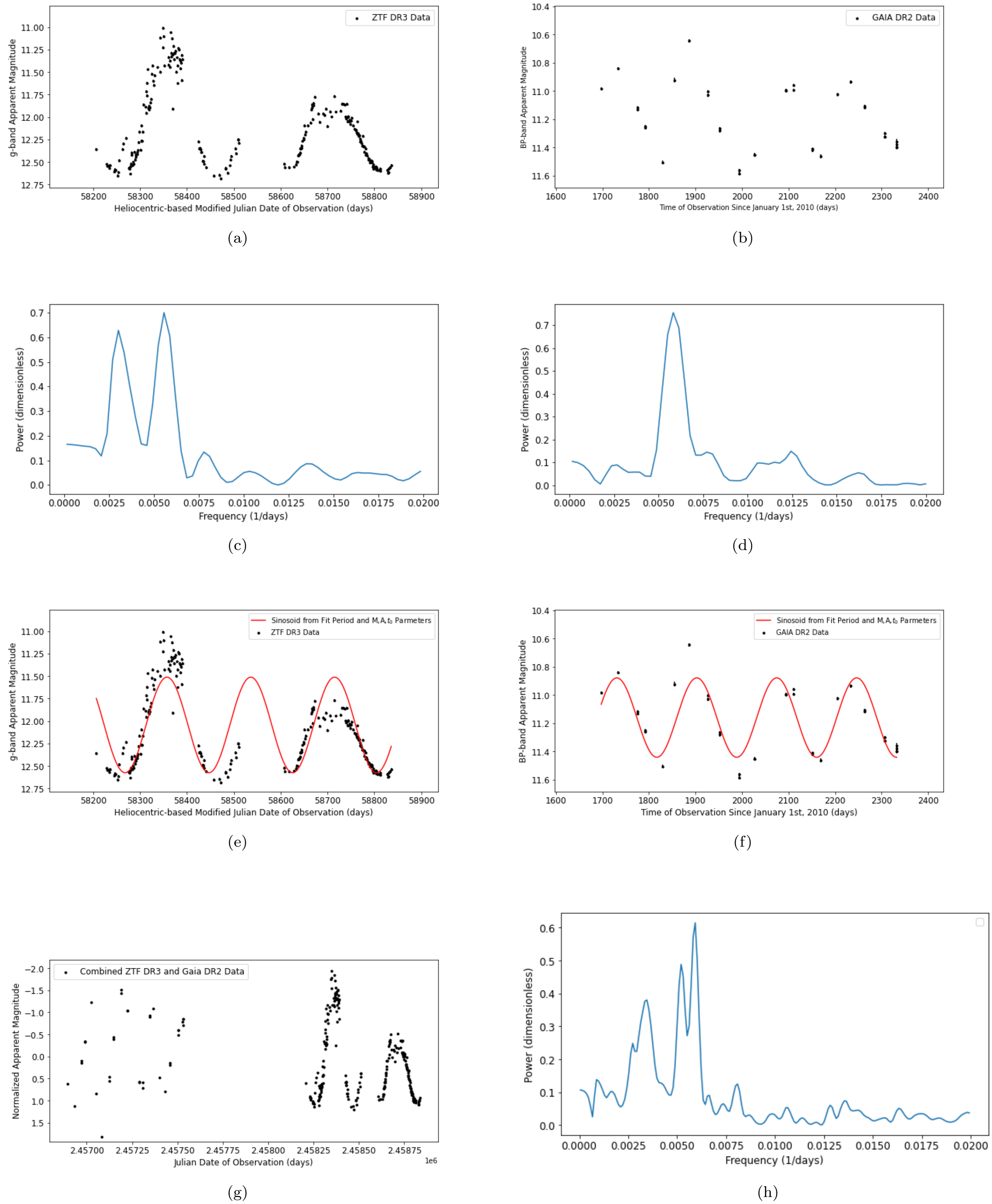


Figure 1. Figures 1a-1h detail steps of our PeriodFit() process for Gaia DR2 source 2207777193478048000. These figures can be reproduced using the PeriodFancyFit() function. Figures 1a and 1b are source 2207777193478048000's ZTF DR3 (1a) and Gaia DR2 (1b) light curves. Figures 1c and 1d are the Lomb-Scargle Periodograms of the ZTF DR3 (1c) and Gaia DR2 (1d) light curves. Figures 1e and 1f show the sinusoid fit to each light curve through our LSR fit of Equation 1. Figure 1g is our normalized light curve, and Figure 1h is the result of applying the LSP to the normalized light curve.

.

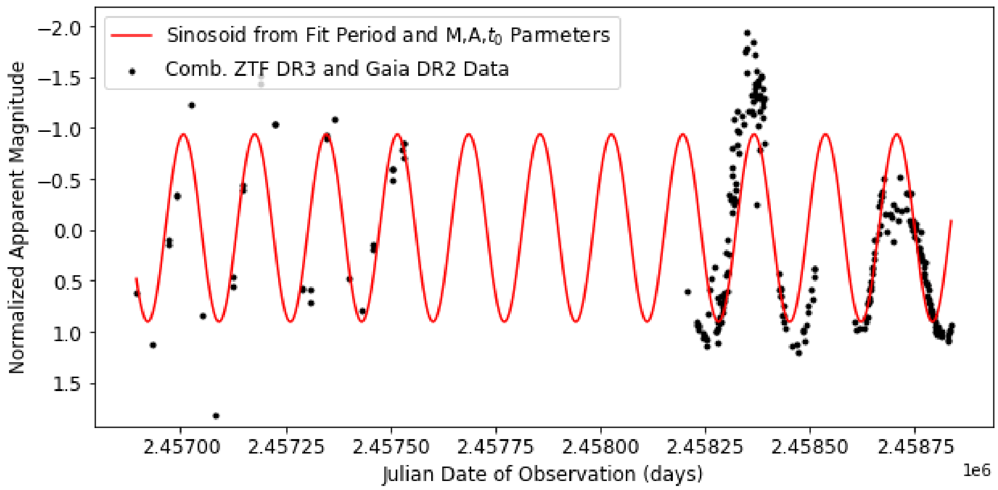


Figure 2. LSR sinusoidal fit of normalized light curve of Gaia DR2 Source 2207777193478048000. The period and period error extracted from this fit are published in Table 1.

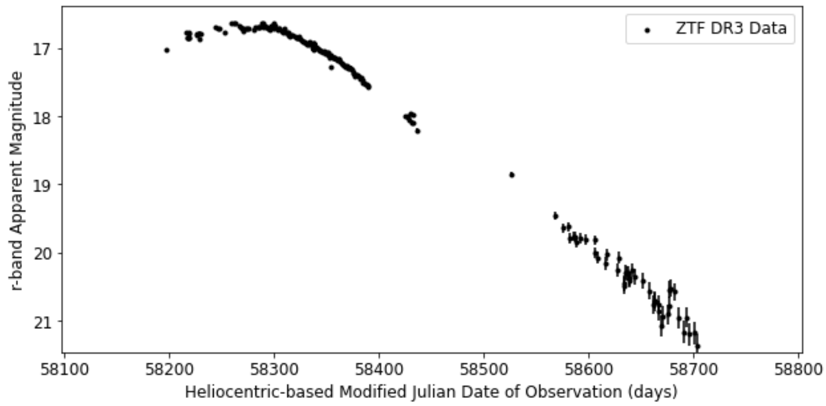
are the results of applying and extracting periods of variability from Lomb Scargle Periodograms applied to Gaia DR2, ZTF DR3, and normalized light curves. We also publish periods of variability and M, A, and t_0 fitting parameters found by applying our fitting algorithm to Gaia DR2 and ZTF DR3 light curves separately without normalizing and combining the data. To avoid accessing the Gaia Archives, we publish Gaia fit frequencies and frequency errors for each of our ZTF-Gaia Crossmatched sample sources when they are available, from which we extract Gaia fit periods for a select sample of sources (see Section 1). Finally, we publish the Gaia DR2 and ZTF DR3 light curve filter codes identifying the band of each fit light curve, and the number of points in each fit light curve.

An important result was discovered when comparing our fit periods to Gaia periods for sources where such periods were available. We were able to make this comparison for 20% of the ZTF-Gaia Crossmatched sample and 34% of the BAaDE-Gaia sample. We identified majority samples for the ZTF-Gaia Crossmatched and BAaDE-Gaia samples separately. Our majority samples are formed by selecting sources along the 1:1 ratio line formed in "one to one" diagrams, as described in Section 3.5. in Figures 6 and 7a, Gaia periods are plotted against normalized LC fit periods for each sample. The outstanding 34% of sources are included in outlier samples after this selection is made. The intercepts used to select these sources in the MajorityCut() function described in Section 3.5 are presented in Table 2. We selected the majority samples for the ZTF-Gaia Crossmatched and BAaDE-Gaia samples separately due to an internal Python issue that prevented the extraction of BAaDE-Gaia sources from the ZTF-Gaia Crossmatched sample.

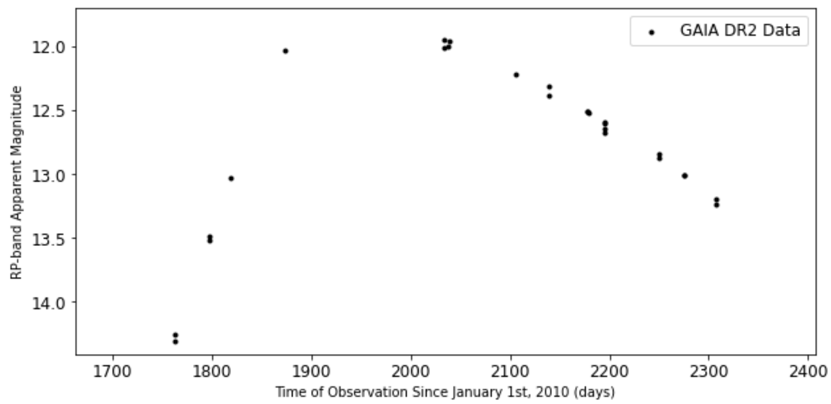
Our period fitting algorithm is designed to minimize the errors on fit periods through the use of the long time baseline produced during our light curve normalization process. Additionally, our fit periods are fit to more data than Gaia periods, which statistically produces a better fit period. Sources with Gaia periods have Gaia light curve data that is 'good' enough for Gaia to fit and published a frequency of variability with sufficiently small frequency errors. Thus, for majority sample sources where the Gaia fit period and our fit period are consistent, we expect our fit periods of variability to be accurate with period errors smaller than those extracted from Gaia.

Figure 6 reveals two notable characteristics of our outlier and majority samples. As discussed in Section 3.3, aliasing is a known problem possible in both our LSR sinusoidal fits and Lomb Scargle Periodograms used in our period fitting algorithm. We find further evidence of aliasing in our period fitting algorithm when examining Figure 6. Similarly to the linear appearance of sources around the one to one ratio line with consistent periods fit by our algorithm and Gaia, we notice other lines present in the scatter of our outlier sample. Though this phenomenon requires further investigation, these linear patterns are a sign of aliasing in our outlier sample, where our fit periods are some multiple of the periods published by Gaia for sources found along these lines.

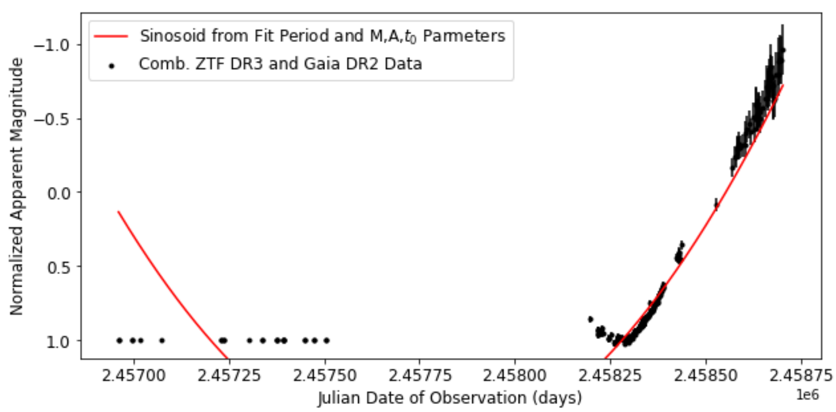
Figures 7a and 7b, one by one diagrams for the BAaDE-Gaia sample, reveal an important characteristic of the sample of sources used in our period - k-band



(a)



(b)



(c)

Figure 3. Misidentification error light curves: source 2021849536746341376. Gaia did not publish a period for this source. Our period fitting algorithm fit a nonsensical period of 247206 ± 17526132 days to this source's normalized light curve. The ZTF DR3 (3a) and Gaia DR2 (3b) light curves are atypical for LPVs, and may indicate poor data quality and misclassification by Gaia. This is corroborated by the normalized and combined light curve (3c) for this source, where poor data quality has caused the Gaia DR2 light curve to appear flat when normalized.

.

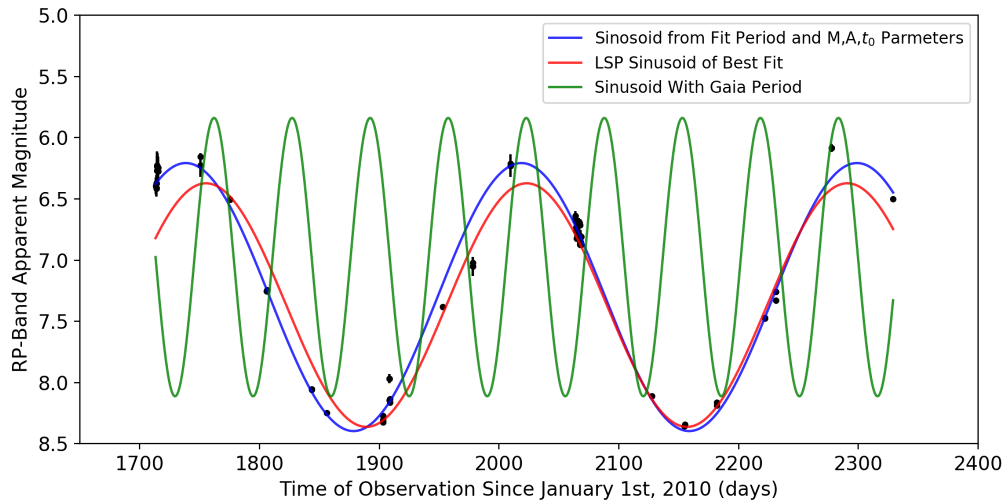


Figure 4. Gaia aliased light curve: Gaia source 4979075051323208832. Our fitting algorithm produces a period that is inconsistent with that provided by Gaia, but consistent with the period fit by the LSP.

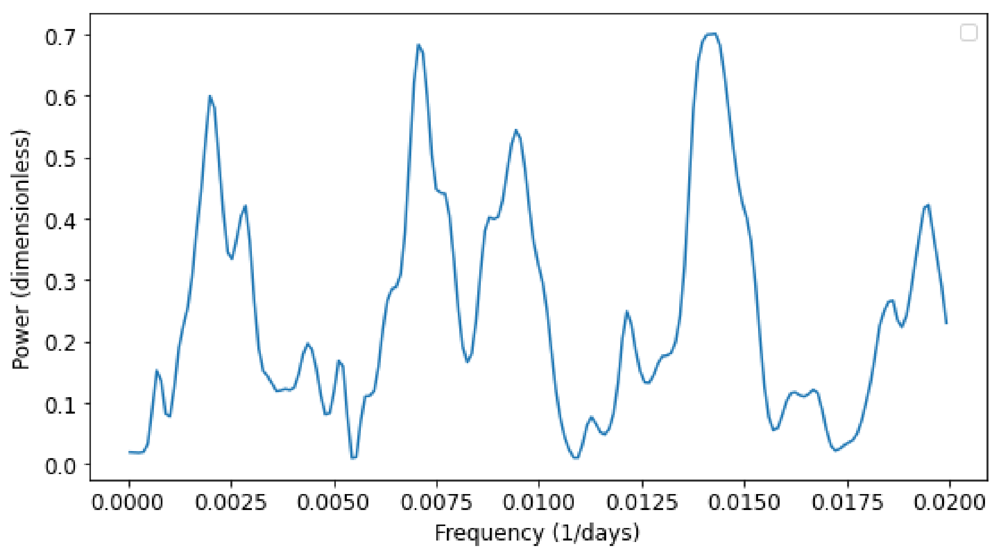


Figure 5. LSP for the combined and normalized light curve of Source 4138630692585295744

Sample	Num. Sources Majority Sample	Num. Sources Outlier Sample	b
ZTF-Gaia Crossmatched	4639	2386	45
BAaDE-Gaia	31	16	40

Table 2. MajorityCut() intercepts and majority sample/outlier sample statistics

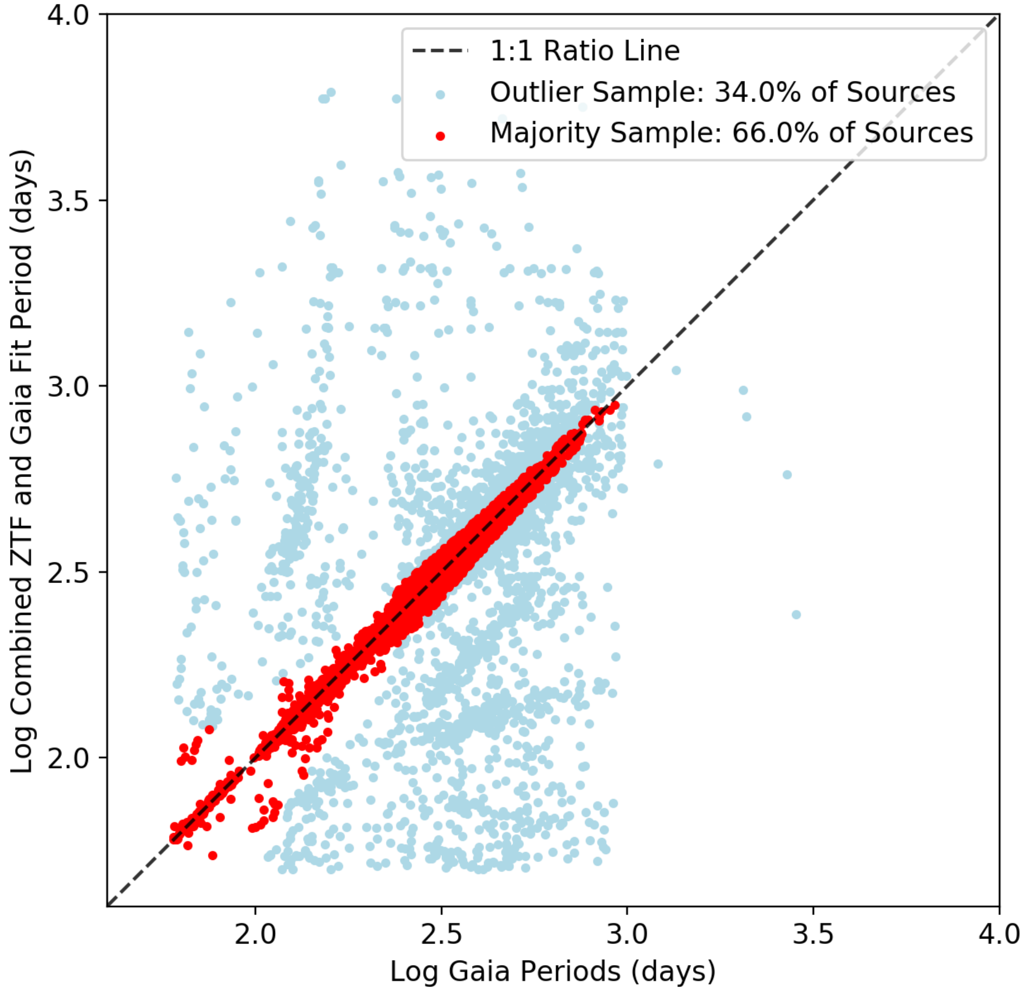
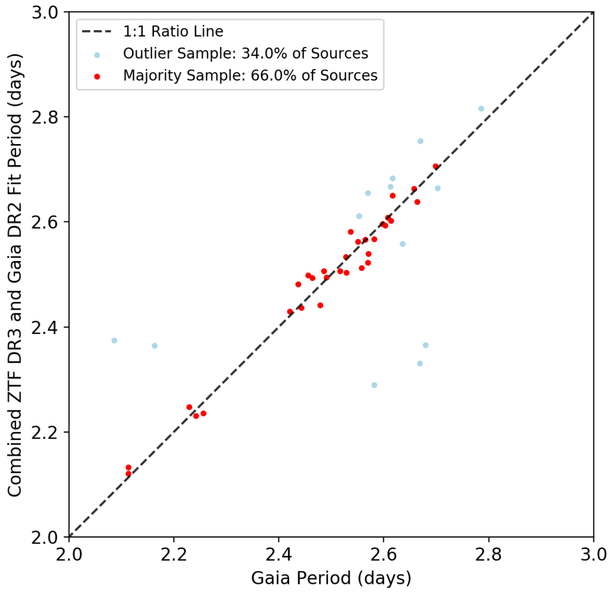


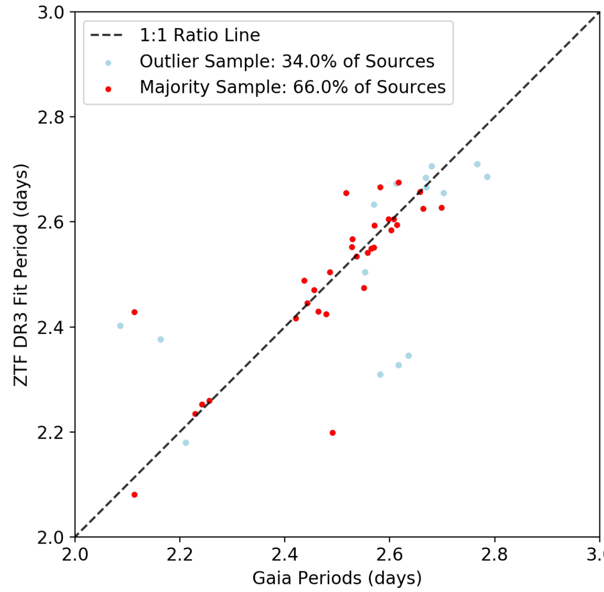
Figure 6. One by one plot. This figure compares our normalized LC fit period on the y-axis to Gaia periods on the x-axis for the ZTF-Gaia Crossmatched Sample.

absolute magnitude relationship. We selected our majority sample by using normalized light curve fit periods, but we also highlighted our majority sample sources in Figure 7b where ZTF DR3 light curve fit periods are compared to Gaia periods. Though there are a few outlier sources that fall outside of the region around the 1:1 ratio line in this plot and more scatter, most of the BAaDE-Gaia majority sample sources fall around the 1:1 ratio line, indicating that our fit ZTF DR3 LC periods are consistent with those published by Gaia, which are fit to completely different data. This suggests that we have selected a sample of LPVs for a log period - k-band absolute magnitude relationship with accurate periods, as the periods fit to Gaia DR2 and ZTF DR3 light curves separately are consistent with each other.

Finally, we present our log period - k-band absolute magnitude relationship for BAaDE-Gaia majority sample LPVs in the solar neighborhood. In Figure 8b, we overlay known log period - k-band absolute magnitude relationships for LPVs in the LMC, as done by Quiroga-Nuñez et al. (2020). We do not identify a distinct functional form relating points plotted in Figure 8a. There appears to be a slight linear relationship between points in figure 8b between the C' and D pulsation mode lines. We believe that this relationship is not present in our data, and rather suggested by the overlaid established relationships, as this linear trend is not evident in figure 8a in the absence of these lines.

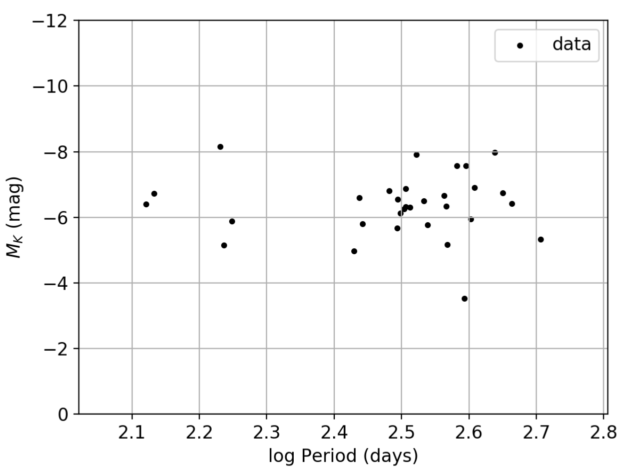


(a)

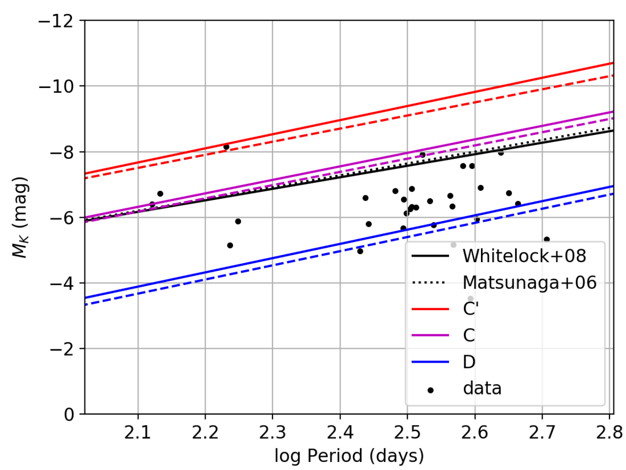


(b)

Figure 7. One by one plots. Normalized LC fit periods and ZTF LC fit periods are compared with Gaia Periods in 7a and 7b for the BAaDE-Gaia sample sources.



(a)



(b)

Figure 8. Period-luminosity relations. 8b overlays PLR for LPVs in the LMC published by Matsunaga & IRSF/SIRIUS Team (2007), and Whitelock et al. (2008), and the pulsation mode relationships overlaid by Quiroga-Nuñez et al. (2020)

.

5. DISCUSSION AND FUTURE WORK

We discussed two known problems of our period fitting algorithm in Section 3.3 by presenting one or two examples of problems that occurred for obvious fit period outliers. Problems that have been demonstrated to exist within our period fitting algorithm and fit light curve data are aliasing and incorrect classification errors. Sources with fit periods that were consistent to Gaia periods were not examined for the presence of these problems, which may exist more widely in our fit period sample. Figure 6 suggests that aliasing is present in fit periods for sources with Gaia periods to an extent beyond one or two anomalous sources. Future work should explore the prevalence of identified problems in our samples of fit periods and uncover other problems that may impact the accuracy of our fit periods.

Future work should design a method for testing the 'goodness' of sinusoidal fits produced by our period fitting algorithm. There are established measures and parameters, such as the chi parameter, that can accurately measure the quality of a LSR fit. However, since aliasing does occur for LSR fit sinusoids with small errors on fit parameters, it may be useful to develop an automated method that can ascertain the quality of our fits by examining how well our fitted sinusoid fits the light curve as a whole, rather than individual points in the fitted light curve. It will be useful to automate a process similar to manually 'looking' at a light curve in its entirety and seeing if the sinusoid fit by our algorithm is convincing. Regardless, as evidenced by Figures 7a and 7b and our discussion of them in Section 4, we expect the periods of sources used in our log period - k-band absolute magnitude to be accurate.

We have several reasons to believe that the log period - k-band absolute magnitude for LPVs in the solar neighborhood to be flat, which we will discuss further in this section. However, the functional form of our log period - k-band absolute magnitude relationship could be impacted by large errors in the k-band absolute magnitudes of the sources used for our PLR. We aim to keep our sample of sources used to examine the PLR as large and representative as possible, to avoid mistakenly identifying a false relationship suggested by a small sample size with little scatter. However, we should check the errors on our BAaDE-Gaia majority sample k-band absolute magnitudes and eliminate any sources with sufficiently large k-band absolute magnitude errors from our log period - k-band absolute magnitude relationship.

Assuming our average k-band absolute magnitudes and log periods are accurate for BAaDE-Gaia majority sample sources, we may expect the flat PLR that we report. Note that most sources plotted in the PLR are found between the C' and D pulsation modes in both Figure 8b and Figure 13 of Quiroga-Nuñez et al. (2020). The BAaDE-Gaia majority sample PLR, which is a subsample of the sources used in the PLR published by Quiroga-Nuñez et al. (2020), features an overdensity of sources in the same region as Figure 13 of Quiroga-Nuñez et al. (2020).

The BAaDE-Gaia majority sample consists of oxygen-rich LPVs in the solar neighborhood with approximate metallicities of Z $\approx$ 1 solar. We know the distances to each of these stars and their dust extinction values, though these values are different for each star. This is not the environment of the LMC, where most established PLRs have been published. Whitelock et al. (2008) and Whitelock et al. (1991) publish linear relationships for oxygen rich Miras and Miras in the LMC respectively. But, these LPVs are distinctly different from those found in the solar neighborhood. LPVs in the LMC are low metallicity stars with the same approximate distances from the Earth and the same dust extinction, three characteristics that are not true of the BAaDE-Gaia sample sources. Whitelock et al. (2008) suggests that the PLR for Miras in the Milky Way may only exist for low metallicity stars, such as those found in Globular Clusters. Future studies of the PLR for LPVs in the Milky Way should take the potential dependence of the PLR on LPV circumstellar shell thickness and metallicity into account.

6. CONCLUSION

In this paper, we present a period fitting algorithm that combines two well tested algorithms, the Lomb Scargle Periodogram and a least squares regression fit of a sinusoidal function, to fit and extract periods of variability from apparent magnitude light curve data. We apply this fitting algorithm to normalized and combined Gaia DR2 and ZTF DR3 light curves to provide our fitting algorithm with more points and generally produce more accurately fit periods. We also take advantage of long time baselines produced through our light curve normalization process to minimize the errors on our fitted periods. This error minimization constrains the errors on periods extracted from the Gaia Achieves when they are consistent with our fit periods. Our fitting algorithm is applied to a sample of 36,000 ZTF-DR3 and Gaia DR2 crossmatched long period variables. We are able to fit periods to 33,492 of these sources. Originally, only 7,362 of these sources had periods that we could extract from Gaia. Additionally, we identify a majority sample of 66% of our crossmatched sources with fit periods that are consistent with those fit by Gaia.

We extract a sample of 31 BAaDE - Gaia - ZTF cross-matched sources with accurate periods and small period errors, good Gaia parallaxes, extinction maps, k-band infrared photometry, and periods consistent with those extracted from Gaia fit frequencies. These long period variables are used to reproduce the log period - k-band apparent magnitude in the solar neighborhood (r<2 kpc) studied previously by Quiroga-Nuñez et al. (2020), Whitelock et al. (2008), and Ita et al. (2004) groups. We find that the log period - k-band apparent magnitude relationship for oxygen-rich LPVs in the solar neighborhood is flat. We suggest that the log period - k-band apparent magnitude relationship for long period variables does not exist in the solar neighborhood in the same form as the relationships found in the LMC. We suggest that log period - k-band apparent magnitude in the solar neighborhood may depend on the metallicity or circumstellar shell thickness of the LPV stars being related.

7. ACKNOWLEDGEMENTS

This project was funded by the generous support of the NRAO Summer student program and the National Science Foundation.

We would also like to thank our collaborators at the BAaDE project and BAaDE Collaboration, Ylva Pihlstrom, Mark Claussen, Megan Lewis, Brandon Medina, and Michael Stroh, for their feedback, guidance, and support throughout the course of this project.

This research has made use of the SVO Filter Profile Service (http://svo2.cab.inta-csic.es/theory/fps/) supported from the Spanish MINECO through grant AYA2017-84089.

APPENDIX

A. PYTHON FUNCTIONS

Relevant project functions can be found in the following Google Drive folder: https://drive.google.com/drive/folders/1Goe7qKAzncbM8OqlJ3ATdQr-djvURVBi?usp=sharing. Below, we summarize the required packages and inputs and outputs of each function. We recommend using the newest version of each package.

PeriodFit(): https://docs.google.com/document/d/1NHwPpiwHZQFgomxh1yBK7M2b8gFwvlbwXa_SGGEYix8/edit?usp=sharing

Packages: astropy.timeseries, pandas, numpy, scipy.optimize, statistics, astroquery.gaia, requests, os

Inputs: Gaia DR2 Source ID of LPV (source_id),ASQL Query to retrieve Gaia DR2 light curve (LC) (jobString), Name of imported Gaia LC- predefined, no path information (GaiaLCfilename), Name of ZTF LC with path information (ZTFLCfilename), folder where extracted Gaia LC are saved (LCfolder), Predefined url required for ADQL query (URL), band of Gaia DR2 LC - b'RP' or b'BP' (GaiaFilter), band of ZTF DR3 LC - 'zr' or 'zg' (ZTFfilter)

Outputs: normalized LC fit period and error (FinalPeriod, FinalPeriodError), normalized LC LSP fit period (FitFirstGuessFinalPeriod) , ZTF LC fit period and error (ZTFOnlyFitPeriod, ZTFOnlyPeriodError), ZTF LC LSP fit period (FirstGuessZTFOnly), Gaia LC fit period and error (GaiaOnlyFitPeriod, GaiaOnlyPeriodError), Gaia LC LSP fit period (FirstGuessGaiaOnly), extracted Gaia period and error (GaiaPeriod, GaiaPeriodError), Gaia published frequency and error (GaiaFrequency, GaiaFrequencyError), band of Gaia LC- RP or BP (GaiaFilterString), band of ZTF - zr or zg LC (ZTFfilter), normalized LC M, A, t_0 parameters (MCOMB, ACOMB, shiftCOMB), ZTF LC M, A, t_0 parameters (MZTF, AZTF, shiftZTF), Gaia LC M, A, t_0 parameters (MGAIA, AGAIA, shiftGAIA), number of points in Gaia LC (GaiaNumPoints), number of points in ZTF LC (ZTFNumPoints). These outputs are esentially the columns of the Fosdick_2021_LPV_Periods table linked to Appendix B.

PeriodFancyFit(): https://docs.google.com/document/d/1aaTnsLptt3tCOtpRnTnn3tg2v9H2qGCX2-jnD3a-_3E/edit?usp=sharing

Packages: astropy.timeseries, matplotlib.pyplot, pandas, numpy, scipy.optimize, math, astropy.table, statistics, astroquery.gaia, requests, os

Inputs:Gaia DR2 Source ID of LPV (source_id),ASQL Query to retrieve Gaia DR2 light curve (LC) (jobString), Predefined url required for ADQL query (URL), Name of ZTF LC with path information (ZTFLCfilename), ZTF band - 'zg' or 'zr'(ZTFfiltercode), Gaia band - b'RP' or b'BP' (Gaiafiltercode), Gaia band for figure labels - 'RP' or 'BP' (GaiafiltercodeString), ZTF band for figure labels - 'g' or 'r' (ZTFfiltercodeString), folder where extracted Gaia

LC are saved (LCfolder). User must also indicate internally where produced figures are saved by providing path and file names for each figure inside the function.

Outputs: normalized LC fit period and error (FinalPeriod, FinalPeriodError), extracted Gaia period and error (GaiaPeriod, GaiaPeriodError). Figures are saved automatically when code is run.

MajorityCut(): https://docs.google.com/document/d/1lNuGXdpfhcjmxuuxEp8u4fSk9QbuBzSsmNgH3WW9V1g/edit?usp=sharing

Packages: None

Inputs: b-intercept (b), slope (m), x-axis and y-axis data and errors (x, y, xErr, yErr), Gaia DR2 source ID (SourceID).

Outputs: majority sample x-axis and y-axis data and errors (majx, majxErr, majy, majyErr), outlier sample x-axis and y-axis data and errors (outlierx, outlierxErr, outliery, outlieryErr), Num. Sources in majority sample (Majsize), Num. Sources in outlier sample (Outsize), Gaia DR2 Source IDs of sources in outlier and majority samples (OutSourceID,MajSourceID).

B. PERIODFIT FITTING RESULTS TABLE

Fosdick_2021_LPV_Periods:
https://docs.google.com/spreadsheets/d/1PwYWUseuZGyXHWkSaoaojb1KWaqq5H7RWkx7cVH3F-M/edit?usp=sharing

The table linked above consists of the outputs of the PeriodFit() function when applied to our ZTF-Gaia Crossmatched sample of 36,000 LPVs. The columns of this table are Gaia_DR2_Source_ID, Fit_Period(days), Fit_Period_Error(days), period fit from LSP of normalized LC (Fit_Period_First_Guess(days)), Fit_Period_From_ZTF_Data(days) Period_Error_From_ZTF_Data(days), period fit from LSP of ZTF LC (First_Guess_Period_ZTF(days)), Fit_Period_From_Gaia_Data(days), Period_Error_From_Gaia_Data(days), period fit from LSP of Gaia LC (First_Guess_Period_Gaia(days)), Gaia_Published_Frequency(1/days) Gaia_Published_Frequency_Error(1/days), Gaia LC filter code: 'RP' or 'BP' (Gaia_LC_filter), ZTF LC filter code: 'r' or 'g' (ZTF_LC_filter), Fosdick 2021 LSR fitting parameters: normalized LC (M_Parameter_Comb, A_Parameter_Comb, t0_Parameter_Comb), Fosdick 2021 LSR fitting parameters: ZTF LC (M_Parameter_ZTF, A_Parameter_ZTF, t0_Parameter_ZTF), Fosdick 2021 LSR fitting parameters: Gaia LC (M_Parameter_Gaia, A_Parameter_Gaia, t0_Parameter_Gaia), number of points in Gaia LC (Num_Points_Gaia_LC), number of points in ZTF LC (Num_Points_ZTF_LC), flag.

We use several flags to indicate different outcomes of our period fitting algorithm and characteristics of our light curve data. The flag "Gaia did not publish a period" is used when Gaia did not publish a frequency of variability for the desired source. A similar flag, "Gaia frequency error large. Calculated period and error from Gaia published frequency may not be accurate," is used when the error on the frequency published by Gaia is larger than 20% of the published frequency value, and a Gaia period can not be extracted for the source. When Gaia does not publish a frequency of variability for a source, we publish the frequency as 'masked,' retaining the value extracted from the Gaia Achieves.

Our other flags relate to our period-fitting algorithm and it's outcomes. The "too few points fit not attempted" flag occurs when the g and r band ZTF DR3 light curves of a source have less than 5 data points and we do not attempt to fit the source's period. Conversely, when our fitting algorithm fails and we can not fit a period of variability for a source, the flag "fit failed" is used. When a source is flagged with either of these flags, we publish zeros (0) in all other columns of the Fosdick_2021_LPV_Periods table. When a source is not flagged, we publish a zero (0) in the flag column.

C. ZTF DR3 LIGHT CURVES

ZTF DR3 light curves: https://drive.google.com/drive/folders/11I94F4g9du07F6TIk6VRQwT5zG7soGQE?usp=sharing

Here we provide a link to the ZTF DR3 light curves crossmatched with Gaia DR2 LPVs in our ZTF-Gaia Crossmatched sample. These files are named schematically as Gaia_DR2_()_ztf_dr3_lc.csv, where () is the source's Gaia DR2 Source ID.

REFERENCES

Capitanio, L., Lallement, R., Vergely, J. L., Elyajouri, M., & Monreal-Ibero, A. 2017, A&A, 606, A65, doi: 10.1051/0004-6361/201730831

Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4, doi: 10.1051/0004-6361/201832756

Höfner, S., & Olofsson, H. 2018, The Astronomy and Astrophysics Review, 26, doi: 10.1007/s00159-017-0106-5

Ita, Y., Tanabé, T., Matsunaga, N., et al. 2004, MNRAS, 347, 720, doi: 10.1111/j.1365-2966.2004.07257.x

Jiménez-Esteban, F. M., Engels, D., Aguado, D. S., González, J. B., & García-Lario, P. 2021, Monthly Notices of the Royal Astronomical Society, 505, 6051, doi: 10.1093/mnras/stab1596

Jura, M., & Kleinmann, S. G. 1989, ApJ, 341, 359, doi: 10.1086/167499

Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, A&A, 625, A135, doi: 10.1051/0004-6361/201834695

Matsunaga, N., & IRSF/SIRIUS Team. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 378, Why Galaxies Care About AGB Stars: Their Importance as Actors and Probes, ed. F. Kerschbaum, C. Charbonnel, & R. F. Wing, 86. https://arxiv.org/abs/astro-ph/0610816

Pols, O. 2012, Stellar Structure and Evolution Lecture Notes, Chapter 11, https://www.astro.ru.nl/~onnop/education/stev_utrecht_notes/chapter9-11.pdf

Quiroga-Nuñez, L. H., van Langevelde, H. J., Sjouwerman, L. O., et al. 2020, The Astrophysical Journal, 904, 82, doi: 10.3847/1538-4357/abb5a6

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

Sjouwerman, L. O., Pihlström, Y. M., Rich, R. M., et al. 2018, in Astrophysical Masers: Unlocking the Mysteries of the Universe, ed. A. Tarchi, M. J. Reid, & P. Castangia, Vol. 336, 180–183, doi: 10.1017/S1743921317009292

VanderPlas, J. T. 2018, ApJS, 236, 16, doi: 10.3847/1538-4365/aab766

Whitelock, P., Feast, M., & Catchpole, R. 1991, Monthly Notices of the Royal Astronomical Society, 248, 276

Whitelock, P. A., Feast, M. W., & Van Leeuwen, F. 2008, Monthly Notices of the Royal Astronomical Society, 386, 313, doi: 10.1111/j.1365-2966.2008.13032.x

Wood, P. R. 2000, Publications of the Astronomical Society of Australia, 17, 18–21, doi: 10.1071/AS00018

Wood, P. R., Olivier, E. A., & Kawaler, S. D. 2004, The Astrophysical Journal, 604, 800, doi: 10.1086/382123