

## **Distribution Agreement**

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Eden Schapera

April 10, 2024

# Untangling RR Lyrae in Gaia DR3

by

Eden Schapera

Dr. Alissa Bans

Adviser

Department of Physics

Dr. Alissa Bans

Adviser

Dr. Merida Batiste

Committee Member

Dr. Erin Bonning

Committee Member

2024

Untangling RR Lyrae in Gaia DR3

By

Eden Schapera

Dr. Alissa Bans

Adviser

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Department of Physics

2024

## Abstract

### Untangling RR Lyrae in Gaia DR3

By Eden Schapera

Although the GAIA database is an incredibly useful tool to assess baseline variability, the low and uneven cadence across different Gaia sectors makes it very difficult to distinguish between objects with similar lightcurve features. Using previously verified W. Ursa Majoris and RRc Lyra stars, we develop a classification pipeline using minima depth variation in the lightcurve to distinguish between genuine RR Lyrae and eclipsing binaries. We assess a set of 72 variables with poor or incomplete lightcurves in Gaia using TESS archival data, and identify 12 objects as genuine RR Lyrae based on their minima variation and secondary lightcurve features. We further verify the lightcurves of three variables using observational data from the Emory observatory, reinforcing the efficacy of our classification pipeline.

Untangling RR Lyrae in Gaia DR3

By

Eden Schapera

Dr. Alissa Bans

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences  
of Emory University in partial fulfillment  
of the requirements of the degree of  
Bachelor of Science with Honors

Department of Physics

2024

## Acknowledgements

This research made use of Lightkurve, a Python package for Kepler and TESS data analysis (Lightkurve Collaboration, 2018).

## Table of Contents

Chapter 1: Background	-----1
I: RRc Lyrae	-----2
II: W. Ursa Majoris Eclipsing Binaries	-----3
III: Astronomical Databases and Previous Research	-----5
Gaia	
TESS	
Combining Gaia and TESS for Time Domain Photometry	
Chapter 2: Initial Classification and TESS Analysis	-----8
I: Identification of Candidates	-----8
II: The Lightkurve Pipeline	-----8
Data Reduction and Cleaning	
Minima Classification	
Chapter 3: Observational Studies	-----24
I: Identification of Candidates	-----24
II: Data Collection	-----25
III: Calibration	-----26
IV: Processing	-----28
VI: Visualization and Analysis	-----29
LAMOST J073756.25+311646.5	
ASAS J074316+1705	
LAMOST/ATO J106.1517+08.6631	
Chapter 4: Discussion and Future Research	-----38
References	-----41
Tables	-----43
Figures	-----50
Appendix 1: Smoothing Lightcurve Data	-----52
Appendix 2: The Lightkurve RRc Pipeline	-----56



## Analysis of RR Lyrae and Eclipsing Binary Candidates

### Chapter 1: Background

RR Lyrae are a category of periodic variable star identified by their characteristic period-luminosity relationship. The prototype of this category was first identified by Williamina Fleming in 1893<sup>[1]</sup>, and since then has become core components of distance estimation within our galaxy and the galactic halo.

As a star between 0.5 and 0.7 solar masses evolves through the red giant phase, thermal instability causes opacity changes in ionized helium through the kappa mechanism. Decreased opacity increases thermal pressure and drives expansion, cooling the star. As the star cools, opacity drops again, and the star contracts<sup>[2]</sup>. The period of an RR Lyrae's cyclical expansion and contraction is highly correlated with its luminosity. Due to their age, these stars are typically found within globular clusters. By measuring the period of these objects, we determine their absolute magnitude, and can thus determine distances to associated structures within our galaxy.

With the advent of large-scale satellite telescope surveys, one of the most common methods of identifying RR Lyrae is through automated machine learning searches of archival data. While an incredibly useful tool to conduct preliminary classification, machine learning pipelines may struggle to differentiate between variable objects which share similar features to genuine RR Lyrae<sup>[20]</sup>. Of particular interest is the overlap between shared envelope eclipsing binaries and a subset of RR Lyrae which exhibit particularly sinusoidal lightcurves.

In this research, we study objects initially identified by machine learning searches of the GAIA DR3 database to be low-confidence RR Lyrae or eclipsing binaries. We analyze the periodicity and structure of these objects' lightcurves using data from the Transiting Exoplanet

Survey Satellite (TESS) database, and we conduct follow-up observational studies of these objects to generate independent lightcurves and confirm our TESS analysis. Using these tools, we establish a robust classification method for these variable objects.

## I: RRc Lyra Characteristics

RR Lyrae are distinguished into several categories depending on the characteristics of their period-luminosity relationship and shape of their lightcurves.

Type A/B RR Lyrae (RRab) are the most common category of RR Lyrae, comprising approximately 91% of all known variables within this family. RRab variables typically exhibit a period between 12 and 24 hours and can be distinguished by their highly asymmetric lightcurve. By contrast, type C RR Lyrae (RRc) exhibit a much more sinusoidal variation in luminosity across their period. In addition, RRc stars typically have periods on the order of 6-12 hours, significantly shorter than their RRab counterparts. [3]

The lightcurves V2438 Sgr and OGLE BLG-RRLYR-6970, representative of RRab and RRc variables respectively, are presented below for comparison. Both stars were identified and verified through the OGLE gravitational lensing survey. [4]

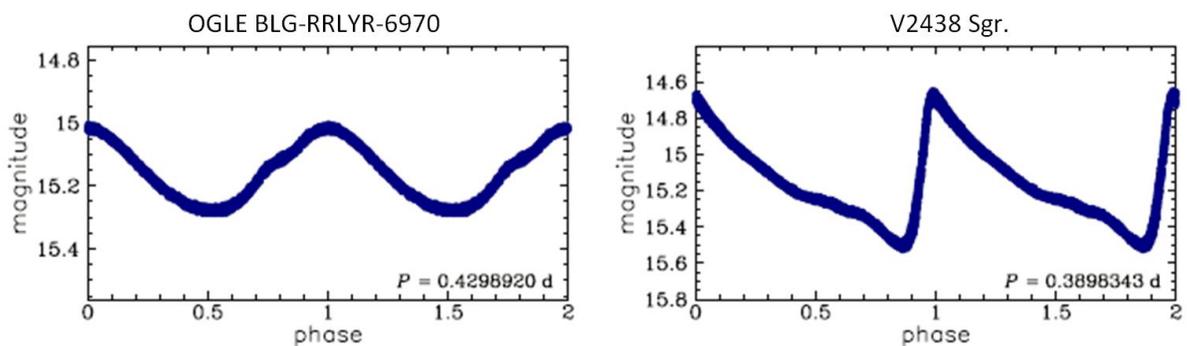


Figure 1: Lightcurve of OGLE BLG-RRLYR-6970, a RRc type variable, and V2438, a RRab type variable. The RRc lightcurve is approximately sinusoidal, with a small asymmetry on the

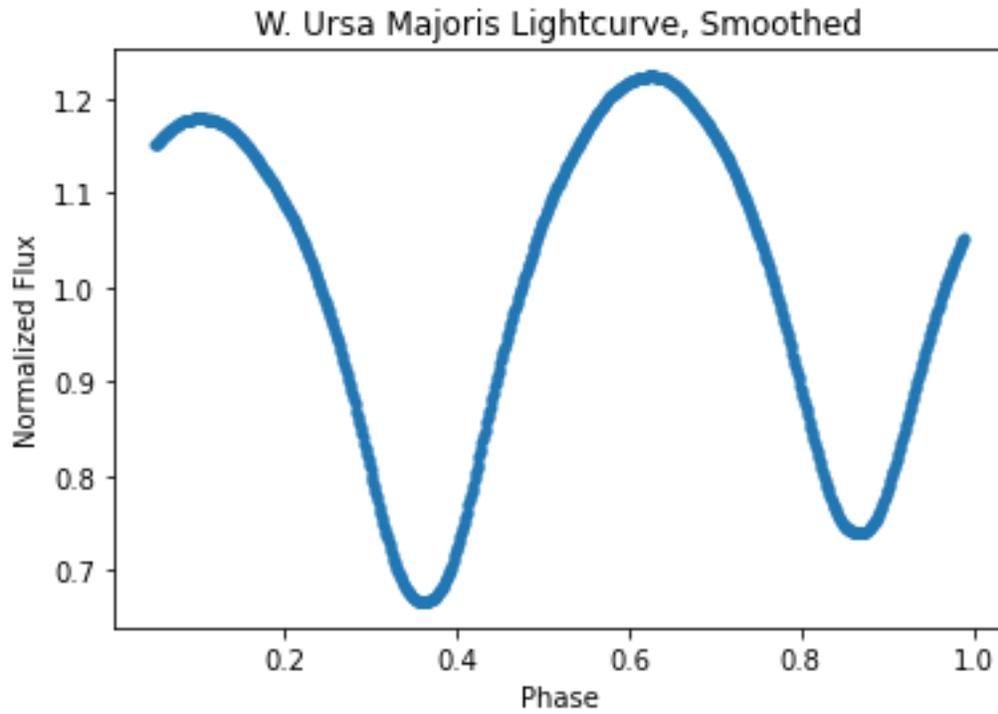
*ascending edge. By contrast, the RRab lightcurve is highly asymmetrical. These features are typical for stars of these categories.* <sup>[4]</sup>

## **II: W Ursa Majoris Eclipsing Binaries**

Eclipsing binaries are a subcategory of binary systems where the orbital plane is parallel to the observer. Maximum luminosity occurs when both stars are visible and decreases when one star in the pair eclipses the other. The maximum decrease in luminosity, known as the primary eclipse, occurs when the hotter star of the pair is partially or completely blocked by the cooler star. A secondary eclipse occurs when the cooler star is blocked by the hotter star.

Although verification of RR Lyrae is the primary motivation for this research, eclipsing binaries also offer a wealth of information on stellar properties and the distribution of stars throughout the galaxy. Using the variation and timing of transits between the primary and secondary eclipses, it is possible to derive both the spectral class, mass, and orbital separation for both stars in the pair. <sup>[5]</sup>

As the orbital separation of a binary pair begins to approach their Roche limit, material begins to transfer between the two stars and forms a shared envelope or contact binary. During this transfer, temperature begins to equilibrate, and the shared envelope forms a more homogeneous mass distribution. As a result, the decrease in luminosity during the primary and secondary eclipse becomes close to equal, and the shape of the lightcurve becomes approximately sinusoidal. <sup>[6]</sup> The lightcurve of W. Ursa Majoris, a well-known variable which the subtype of overcontact binaries is now named after is shown below, generated using our Lightkurve python pipeline (see Chapter 2).



*Figure 2: Lightcurve of W. Ursa Majoris. Data from approximately 20 days of TESS observation was folded by a 0.33-day period to generate the above dataset (see Chapter 2). Note, the flux from this binary is approximately sinusoidal with two distinct magnitude minima.*

### **III: Astronomical Databases and Previous Research**

#### ***Gaia***

The Gaia database is an all-sky catalogue of over 3 billion objects observed by the Gaia space telescope. The satellite telescope gathers parallax distance, astrometric, and photometric data R, B, and G band filters.

Since the launch of the Gaia mission in 2013, there have been three public data releases, Gaia DR1 in 2016, DR2 in 2018, and DR3 in 2020. A fourth release is planned for the end of 2025, with the fifth and final release by the end of 2030.<sup>[7]</sup>

Numerous automated classification pipelines have been performed on the Gaia database to identify variable objects including RR Lyrae. In Clementini et. al. (2023)<sup>[9]</sup> the Specific Object Study (SOS) pipeline is used to assess 2.1 million stars of the Gaia DR3 catalogue using a combination of radial velocity and photometric data. Of the published dataset, this study identifies 72 objects with poor Gaia lightcurves or uncertain classifications, and another 282 objects which have SOS classifications that differ between this study and other automated classification studies.

#### ***TESS***

TESS, or the Transiting Exoplanet Survey Satellite, is an all-sky survey designed to discover and catalogue exoplanetary systems using time-domain photometry.<sup>[10]</sup> Exoplanetary transits are typically rapid, so TESS data is captured at a cadence rate of 2 minutes for a small stars and every 30 minutes for other fields of view. This significantly faster cadence allows the creation of highly detailed light-curves and more precise measurements of stellar variability than similar all-sky surveys.<sup>[11]</sup> Unlike Gaia data, TESS data has a uniform cadence across all targets, but some sectors of the sky currently lack coverage.

In addition, a number of tools have been created to facilitate TESS data acquisition and processing. In particular, the Lightkurve python package<sup>[12]</sup> offers a user-friendly toolkit to generate and analyze lightcurves from TESS archival data. It combines a number of core features for standard data analysis and visualization libraries, including matplotlib, astropy, and scipy.

### ***Combining TESS and Gaia for Time Domain Photometry***

Although Gaia provides time-domain photometric data, its orbital and rotational motion results in an uneven and object-dependent cadence which is often lower than required for high precision periodicity measurements.<sup>[8]</sup> This relatively low cadence may result in features of short-period variables being missed due to low sampling. Some candidates initially classified as RR Lyrae using Gaia data may actually belong to another class of similarly variable object - in particular, objects tagged as RRc variables have been misclassified as W. Ursa Majoris eclipsing binaries.

Similarly, while TESS has a faster and more consistent cadence than Gaia, its pixel size is much larger (approximately 21 by 21 arcseconds).<sup>[19]</sup> This can cause classification uncertainty when several stars are in close proximity to one another and necessitates follow up ground-based observational studies. Additionally, many TESS sectors suffer from scattered light.

By utilizing Gaia data to conduct initial variability assessment, and then conducting more in-depth analysis using TESS photometry, we mitigate the drawbacks of both instruments. In chapter 2, we explore this approach by examining set of uncertain variable objects in the Gaia database using TESS lightcurves to establish a more robust classification. Independent verification of target variability provides yet another method of confirming target classification and eliminating error in both the TESS and Gaia datasets. In Chapter 3, we conduct

observational studies of three uncertain Gaia variables and compare them with our TESS lightcurves to evaluate the validity of our classifications.

## Chapter 2: Initial classification and TESS Analysis.

### I: Identification of Candidates

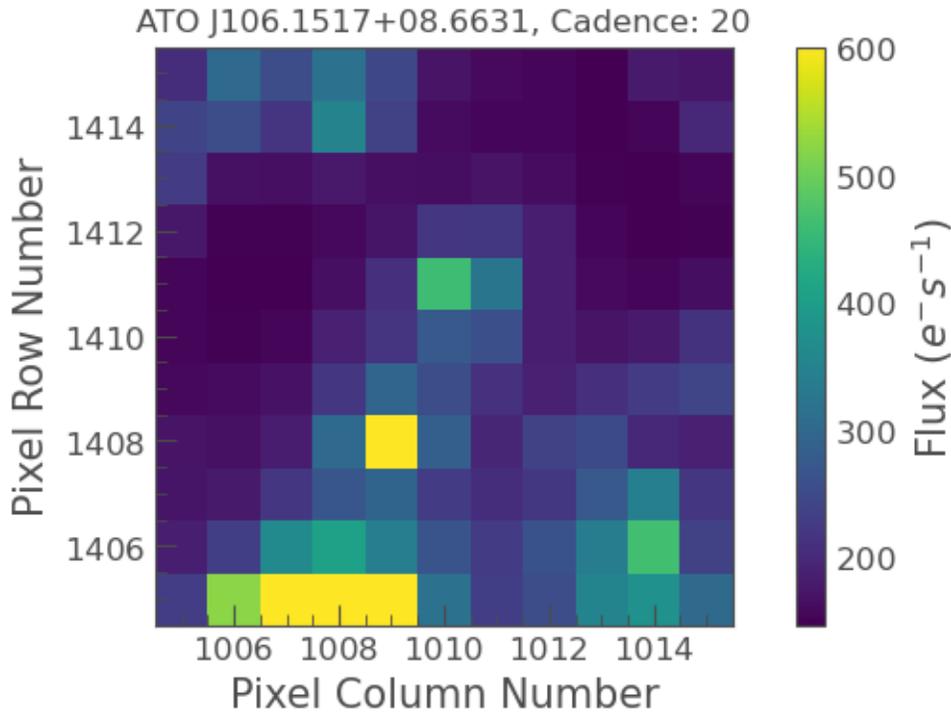
Research candidates were selected from the pool of objects marked by Clementini et. al (2023) as those with poor lightcurves or uncertain measurements. This resulted in a total list of 72 potential research candidates.

Of the 72 candidates, no TESS sector coverage was available for 8 objects along the ecliptic. In addition, in 5 cases the target was in too close proximity to a bright neighbor for data collection, while in an additional 7 cases the target object was too dim for accurate measurement. This resulted in a final target set of 52 objects whose Gaia photometric classification were uncertain.

### II: The Lightkurve Pipeline

TESS data for each target is downloaded by querying the Lightkurve service with the target coordinates. This results in a number of *targetpixelfile* objects, an 11 by 11 (231 by 231 arcsecond) array of pixels and their measured flux for each cadence of the 29 day observation.

In order to isolate our target star from the surrounding area, we define an aperture mask using either automatic threshold masking or manual pixel selection. Threshold masking automatically selects any pixels which have a flux value above a certain flux threshold, while manual pixel selection allows the user to identify a range of pixels in both axis which should be included as target data.



*Figure 3: The targetpixelfile cutout used for target four. ATO J106.1517+08.6631 is located at the center of the frame.*

When we expect our star to be the brightest object within about 21 arcseconds, threshold offers the most streamlined method of isolating the target from the background. However, this method fails when there is another bright object in frame, or when there are several stars in close proximity to the target object. In these cases, we use manual pixel selection to identify our target. This often requires a conservative estimate target area and may decrease the signal to noise ratio.

In order to select an appropriate method for our target, we identify a frame where the background and target star have good contrast and apply both methods. In Figure 6 are forced to use manual pixel selection as threshold masking defaults to the brighter star located in the bottom left corner of the frame.

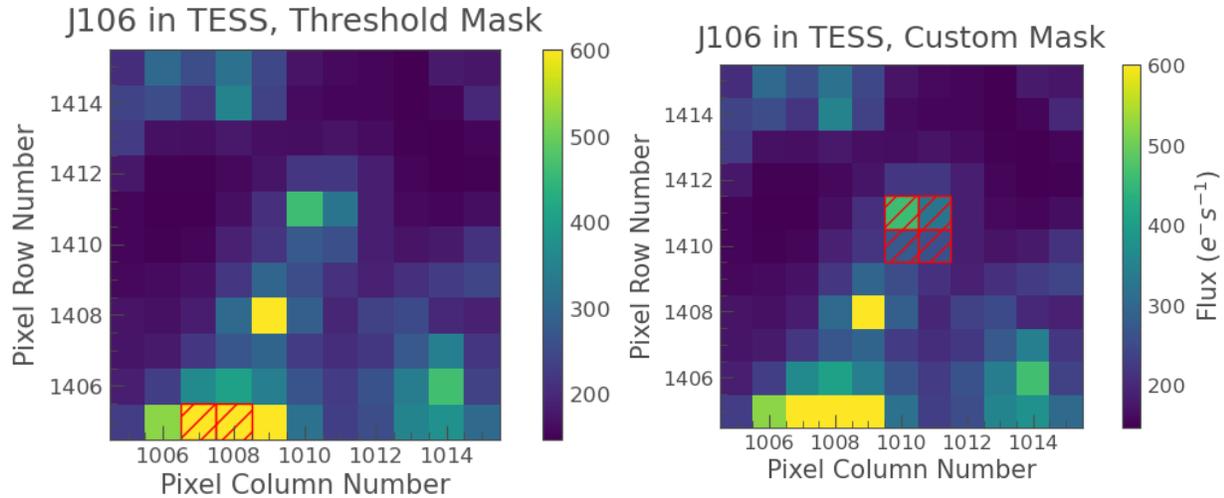


Figure 4: Threshold masking and aperture masking used to select target pixels. Threshold masking incorrectly selects the bright neighbor rather than the target star in the center.

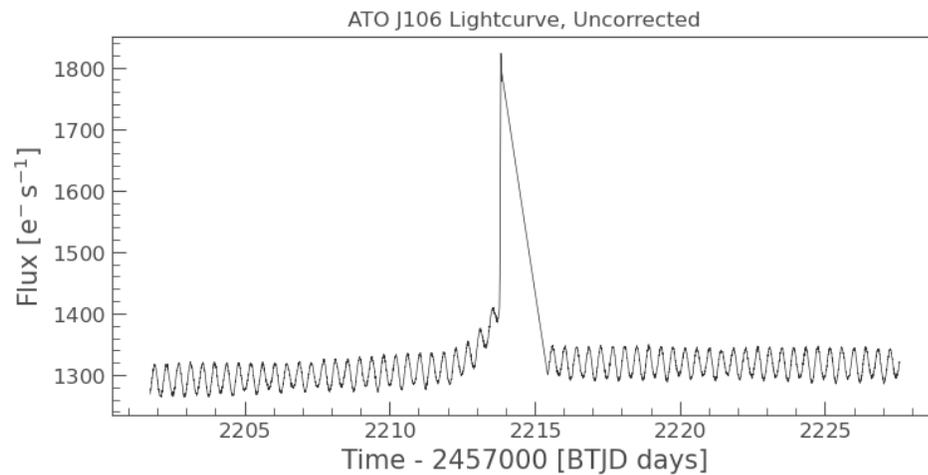


Figure 5: Lightcurve of LAMOST/ATO J106.1517+08.6631, generated using the manual aperture selection in Figure 6. The lightcurve exhibits short-term variation on the order of 9 hours due to genuine stellar variability, while a spike in luminosity observed at  $BTJD=457000+2214$  is due to scattered light from the earth. This light pollution is corrected using pixel level decorrelation.

After identifying an appropriate aperture mask, we plot the total flux of our selected region as a function of time. This results in an uncorrected lightcurve for our target. An example of this first lightcurve approximation is provided for target LAMOST/ATO J106.1517+08.6631 (Figure 7).

In addition to short period variability on the order of approximately 9 hours, we also observe long term trends which dominate the data near BTDJ=2457000+2214. This is primarily due to scattered light from the earth and moon as TESS orbits every 13.7 days. In order to eliminate these trends, we must subtract this effect from our data using regression techniques.

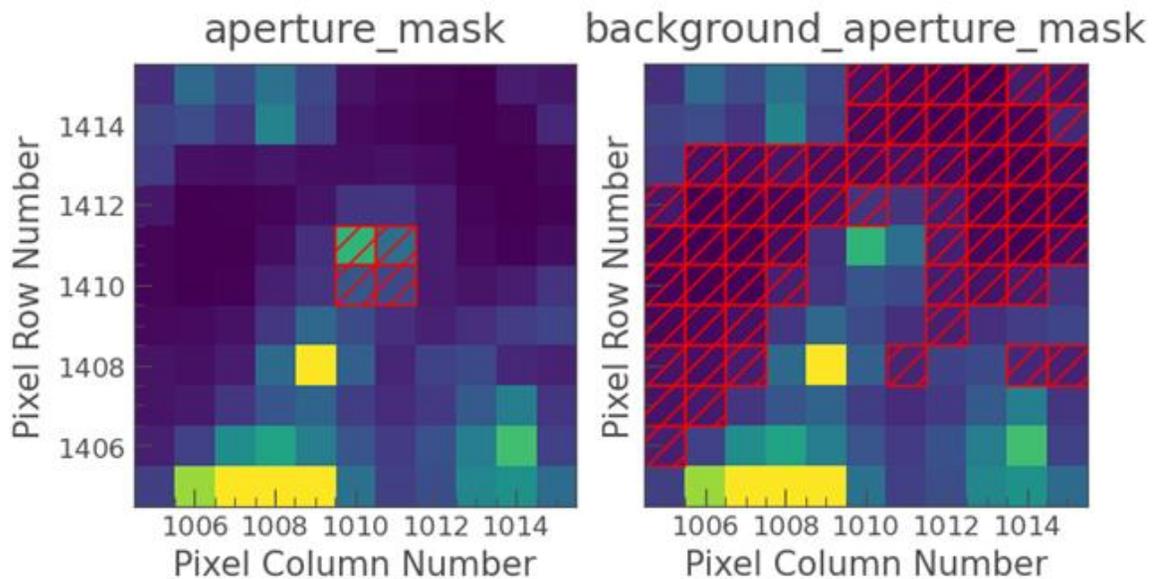
### ***Data Reduction and Cleaning***

Our primary method of data reduction in this analysis is pixel-level decorrelation (PLD). This algorithm was originally developed for analysis of Spitzer telescope data. However, this technique is generalizable to other datasets and has been successfully used in several other studies to eliminate systematic error. <sup>[13][14]</sup>

The PLD algorithm measures the flux of background pixels for each frame and uses linear regression to create a model of the systematic error over – including the scattered light signature present in many TESS datasets (see Appendix 3 for a detailed discussion of the PLD algorithm).

The error model is subtracted from the target data, and outputs the cleaned target lightcurve. As with aperture masking, PLD correction requires the user to select pixels which correspond to background flux. A simple approach is to select any pixels where the flux is below one standard deviation of the average flux per frame. <sup>[15]</sup> This is generally sufficient to ensure that no signal from background stars is included in the error model.

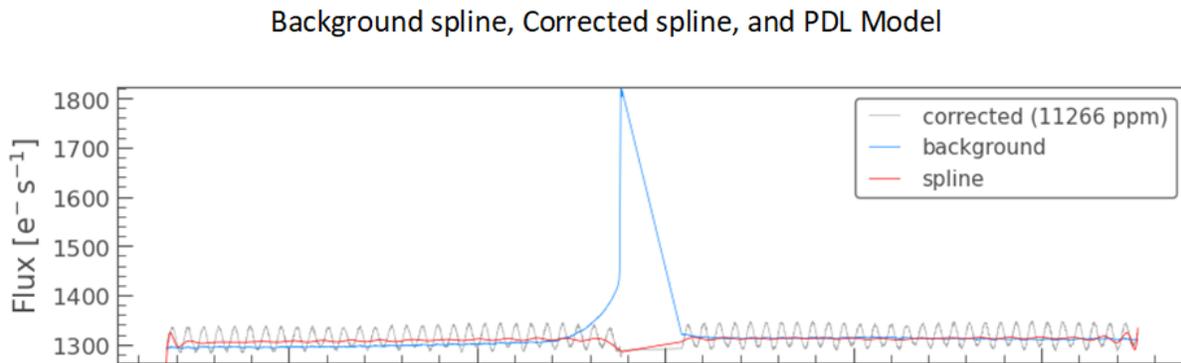
We can explore the efficacy of this approach using the `pld.diagnose()` tool provided in the Lightkurve package. Shown below is the aperture mask for LAMOST/ATO J106.1517+08.6631, and the  $1\sigma$  background aperture. Note, no visible stars are included in the background aperture, and we can be relatively confident that the background flux estimate will be accurate.



*Figure 6: Aperture masks and background masks used in the PLD algorithm. A design matrix containing the background light trends for each pixel in the background mask is used to create an overall background/noise spline.*

The Lightkurve library offers built-in implementation of the PLD algorithm. Using our `targetpixelfile` object, target aperture, and background mask, we generate a reduced dataset. Initial correction parameters, including the number of PCA components and spline properties, are specific to TESS but may be adjusted to fine tune the correction.

We explore the efficacy of the PLD model and determine if any overfitting occurred by comparing the background splines and target splines. Data during portions of the PLD model that are overcorrected may be manually removed. Through testing on a set of known variables (see *minima classification*) we establish that default TESS PLD parameters were sufficient to eliminate background flux without overfitting the model.



*Figure 7: Splines generated by the PLD algorithm. The background spline (blue) effectively corrects the scattered light observed in the original data. Areas where the target spline (red) is not constant – indicating potential overfitting - are manually removed from the dataset*

The background spline (blue) represents noise from scattered light and background flux and is generated by a linear combination of flux from pixels corresponding to the background mask. This spline is scaled and subtracted from the target data, resulting in the corrected lightcurve. The red target spline represents the average continuum level of the corrected lightcurve.

### ***Minima Classification***

Classification of targets was performed by measuring the average variation between two consecutive minima of the target lightcurve. In genuine RRc Lyrae variables, we expect a very

low variation in flux between two periods, while for W Ursa Majoris variables, we expect a much higher variation in luminosity between half phases due to slight differences in stellar temperature.

To determine an appropriate threshold for classification, we measure the minima variation on a verified sample of 12 RRc Lyrae by Sneden et. al. (2017)<sup>[16]</sup> Of this group, one object was excluded from the dataset due to a bright neighbor preventing accurate data collection. Furthermore, two objects did not have TESS data available, and were removed from the sample.

For the remaining 9 verified RRc Lyrae, a reduced lightcurve was obtained using the Lightkurve pipeline previously outlined. Each objects' period was identified using the built-in `to_periodogram()` function, and the power-frequency relationship was plotted for verification.

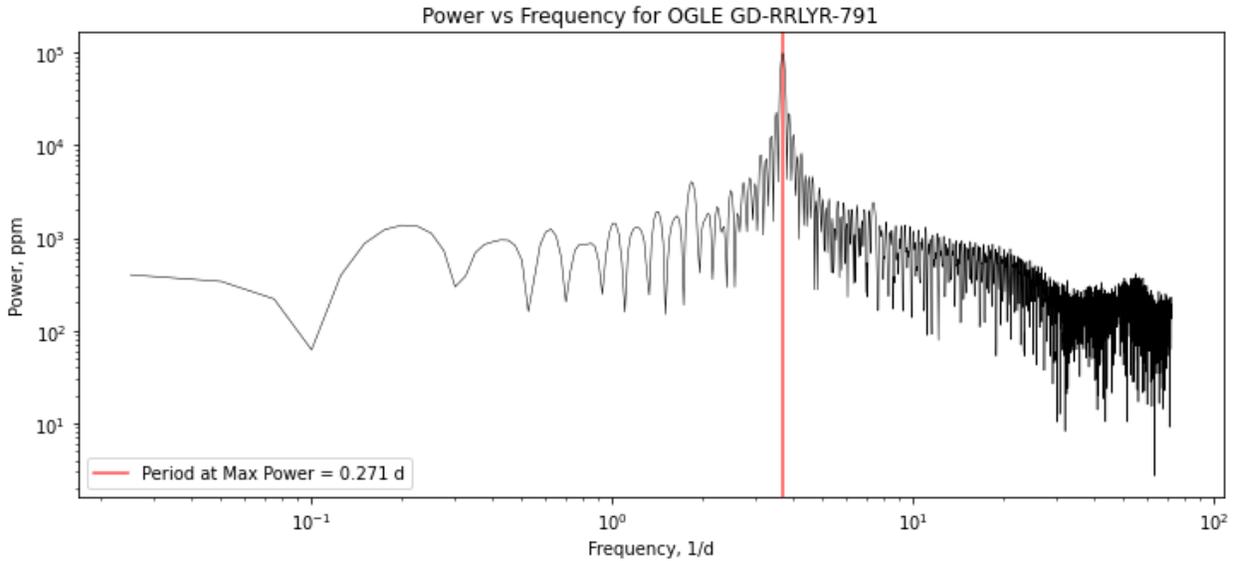


Figure 8: Periodogram of OGLE GD-RRLYR-791. The frequency which corresponds to the maximum power is highlighted in red. This period is used for to fold the lightcurve in Fig. 11.

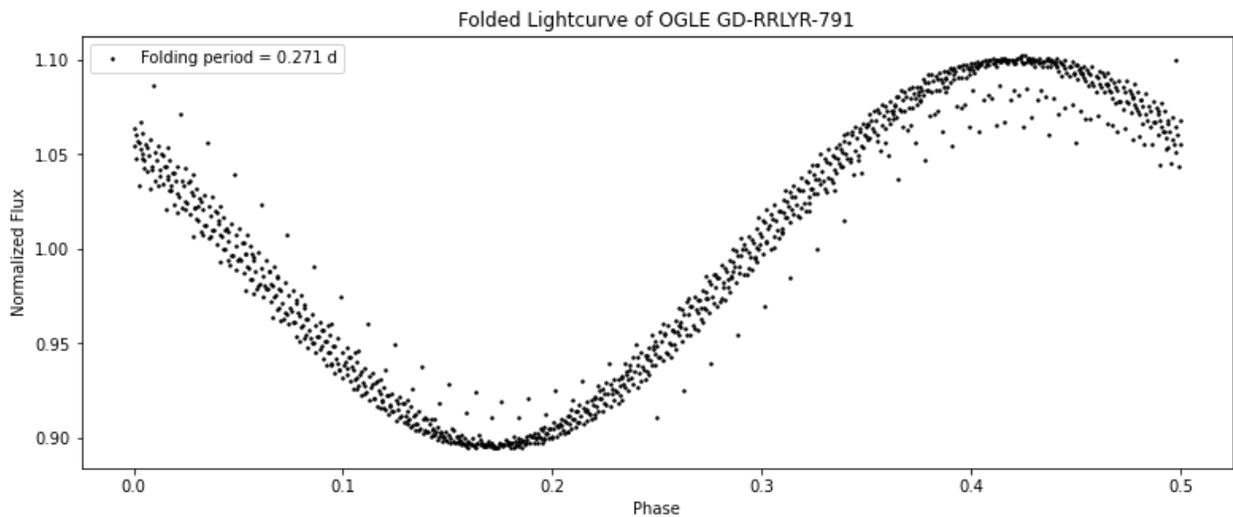
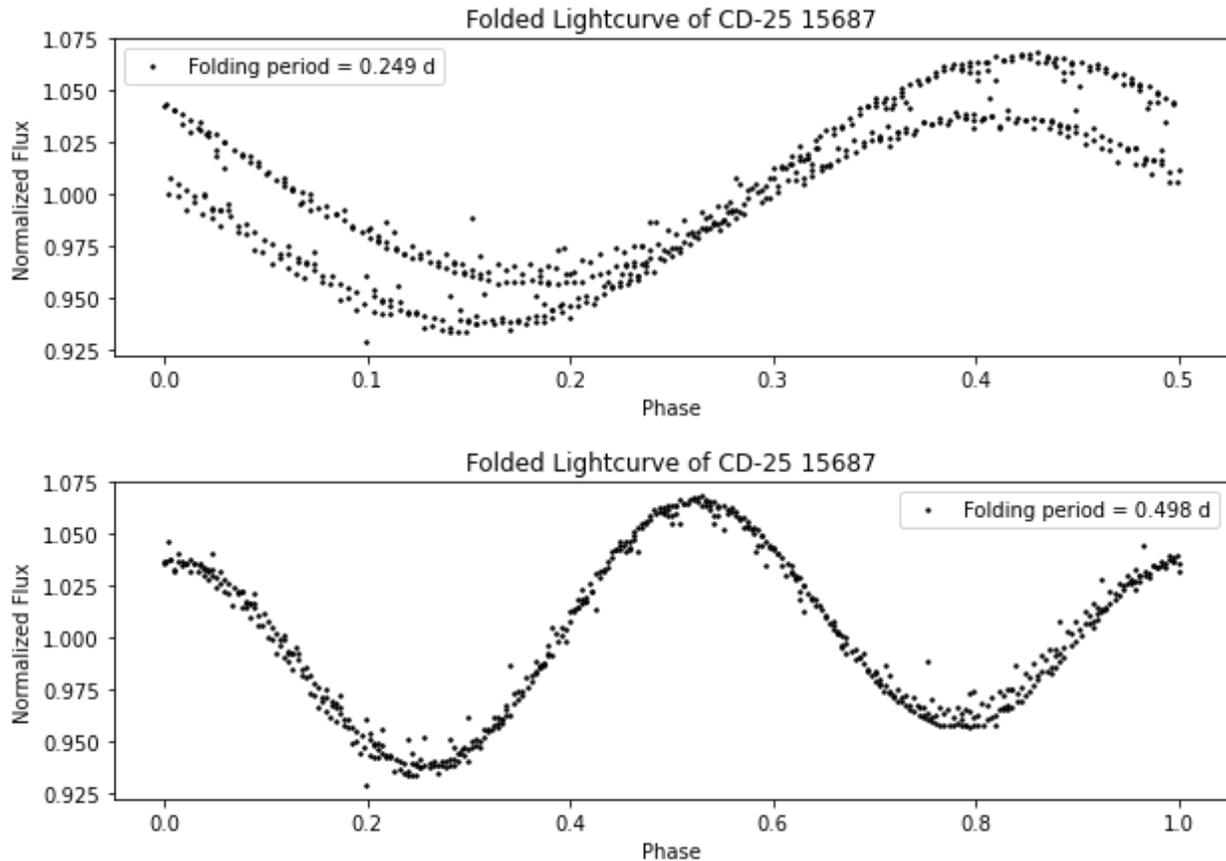


Figure 9: Folded lightcurve of OGLE GD-RRLYR-791. The folding period of 0.271 effectively combines the 28 day TESS dataset into a single lightcurve.

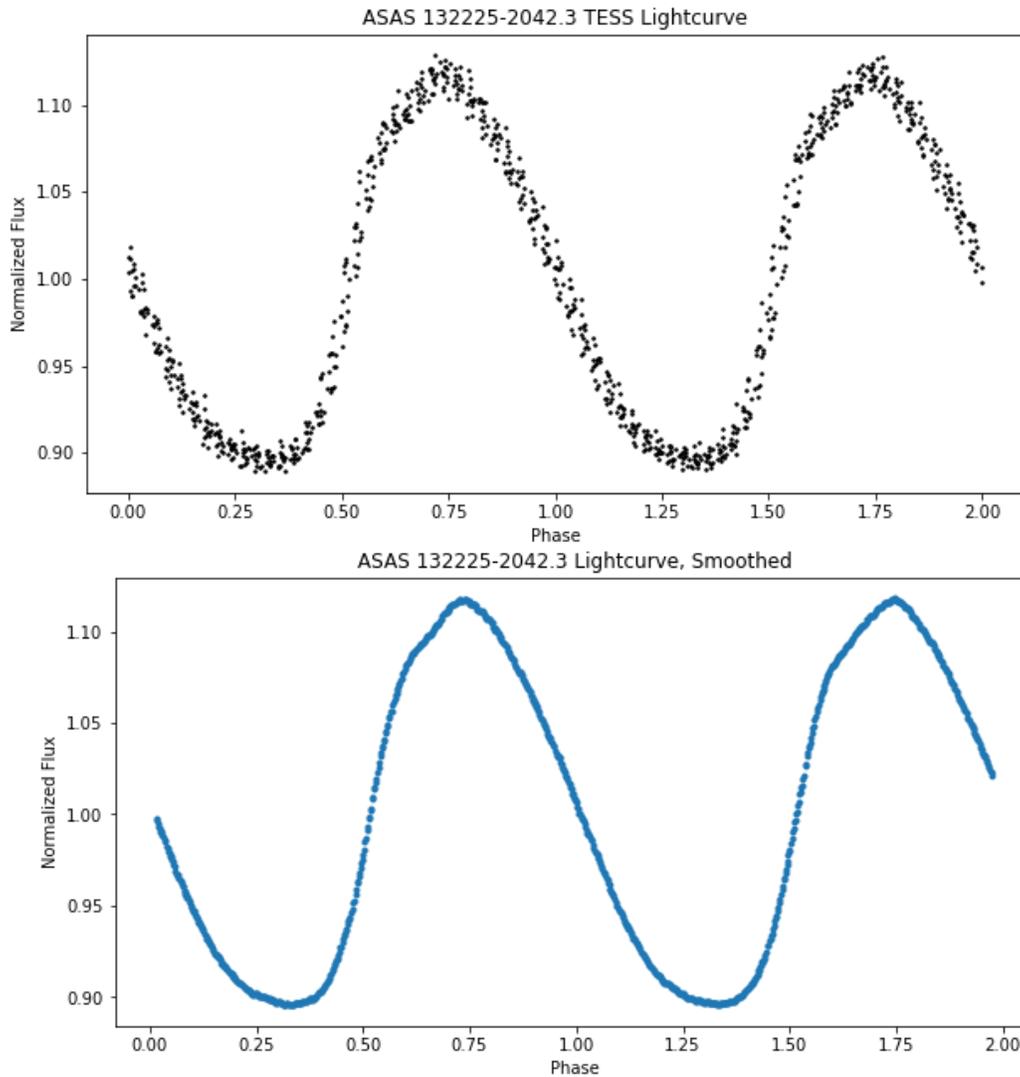
In the above example of OGLE GD-RRLYR-791 (object 19 of the Clementini et. al 2023 dataset), we identify that the period at max power is approximately 0.271 days. Folding the TESS dataset by this value, we observe this value appropriately combines multiple periods.



*Figure 10: Folded lightcurve of CD-25 1568 - an eclipsing binary. Folding TESS data by the measured period at max power (0.249 days) results in half-periods overlapping. Doubling this folding period results in a clear differentiation between the two minima.*

Based on testing of the objects in Sneden et. al (2017), the Lightcurve folding algorithm accurately identifies the period RR Lyra objects. However, through testing of objects in Soszynski, et. al (2014), we identify that the folding algorithm often erroneously identifies the period of eclipsing binaries as half of its real value, causing the half periods to stack on-top of each other and obscure variability. Therefore, the 28 day TESS dataset is folded by two times the measured period at max power, resolving two periods of RR Lyra and one full period of eclipsing binaries (Fig. 12). This has no effect on the shape or minima variability of RR Lyra objects, but effectively separates the minima of eclipsing binaries for accurate classification.

To reduce noise present in the folded lightcurve, we apply a rolling average filter with a window size of twenty (Fig. 13, bottom). For an in-depth discussion on the efficacy of this smoothing algorithm and its potential effects on classification, see Appendix 1. The local minima of this dataset are then identified by comparing the value of each point to its neighbors within a 20-point window. We subtract the values of both minima identified through this process to determine the average difference between minima on successive periods.



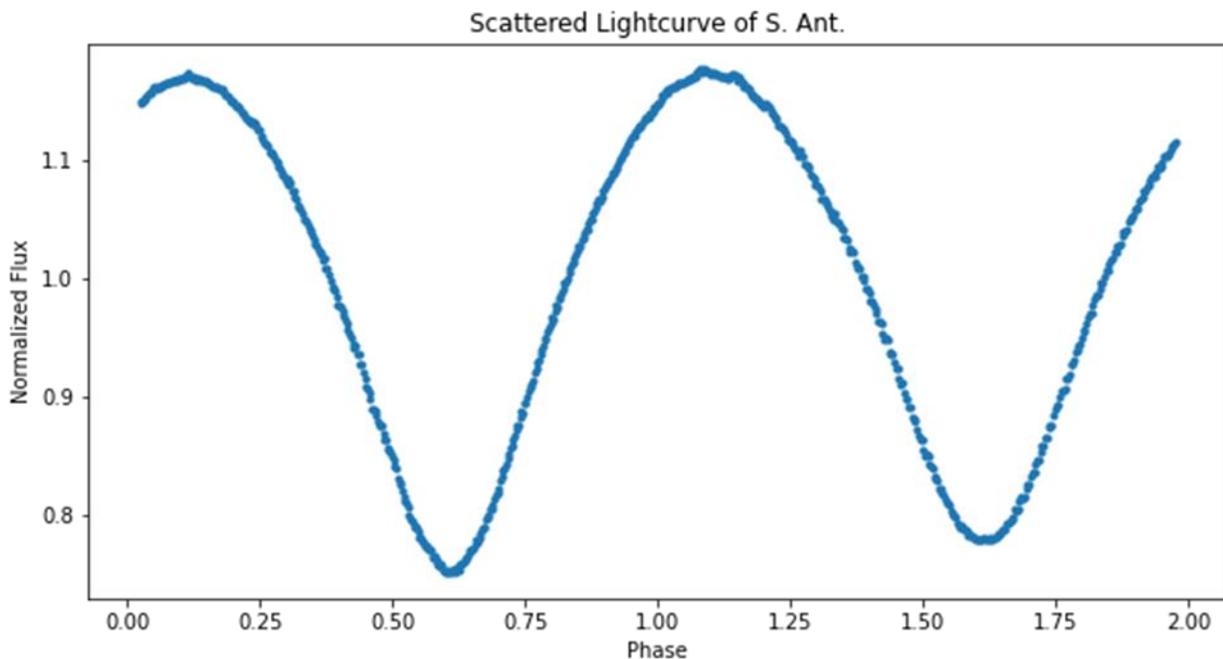
*Figure 11: Scattered (top) and smoothed (below) lightcurve of ASAS 132225-2042.3 in TESS data. Distinguishing features, including a 'notch' on the rising edge of each slope and an asymmetric increase and decrease in luminosity are still clearly visible. This object belongs to the set of genuine RR Lyrae verified by Sneden et. al. (2017)*

This process is repeated for the remaining 9 objects in the Sneden et. al. (2017) sample.

We identify that on average, minima between two periods of a genuine RR Lyrae vary by approximately 0.11%, with a standard deviation of 0.5%.

Assuming an approximately normal distribution, we establish a minima variability threshold for future classification of RRc Lyrae of 0.16%. Retroactively applying this threshold to the dataset of confirmed RRc candidates, this metric would correctly identify 8 out of 9 as genuine RRc variables. The isolated RR Lyra which fell outside this threshold, ASAS 14332-041.2, has a minima variation of 0.21% which is still well outside the average variation observed with genuine eclipsing binaries. The low variation threshold selected for this study was intentionally conservative to provide the cleanest possible sample of RR Lyrae. A list of each RRc object used in this analysis, along with their minima variation, is provided in Table 1.

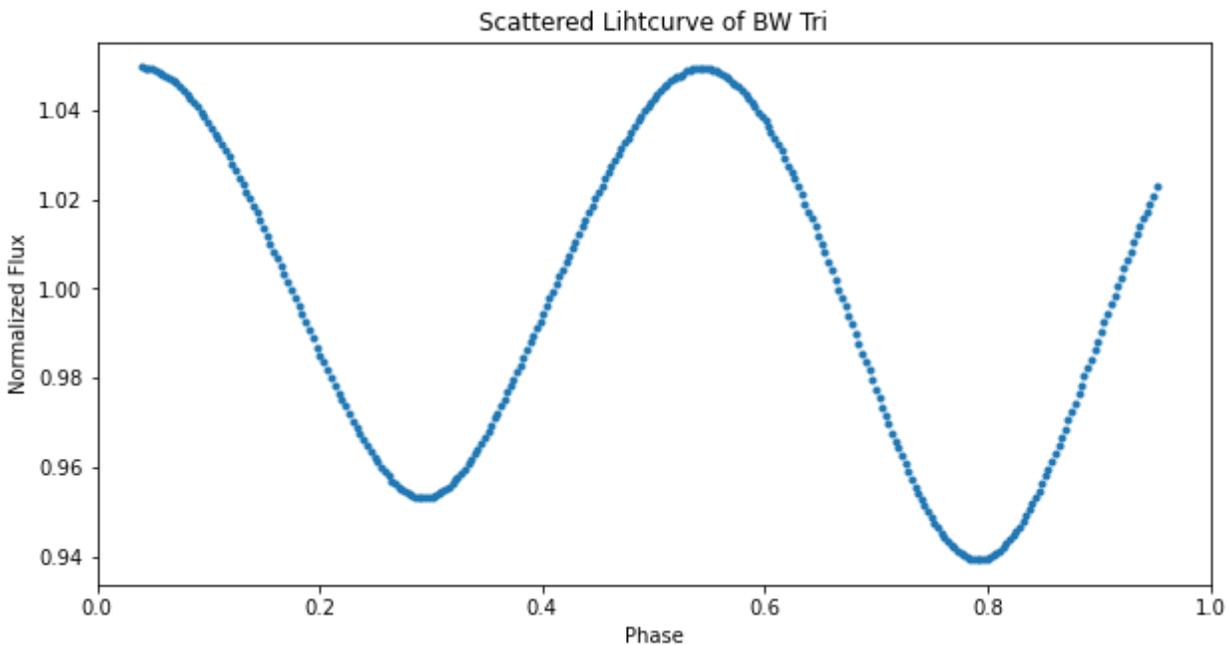
We further validate this classification metric by studying a sample of confirmed W Ursa Majoris (W. UMa) variables compiled and individually studied by Soszynski, et. al (2014). Applying the same lightcurve processing procedure as before, we obtain 8 folded and smoothed W. UMa lightcurves.



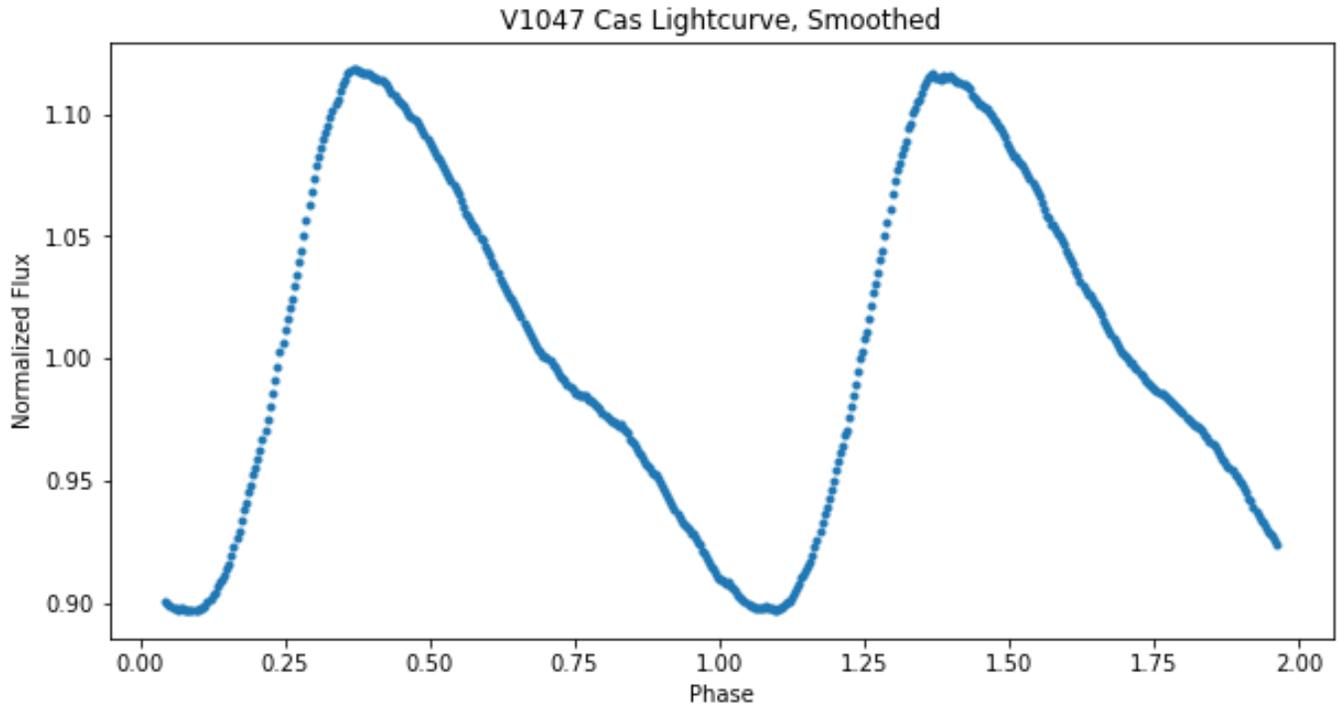
*Figure 12: Smoothed and folded lightcurve of the W. UMa variable S. Ant (Sneden et. al, 2017). The average difference between minima is approximately 3.7% - significantly higher than genuine RRc variables.*

Within this group, primary and secondary minima varied significantly more than in the RRc sample, on average by 4.1% ( $\sigma = 2.1\%$ ). The minima threshold correctly classified all W UMa variables in this sample as non-RRc objects, indicating that this metric is an appropriate tool to identify likely RRc candidates. A table of all W. UMa reference objects and their minima depth variation is provided in Table 2.

This procedure was repeated for all 52 remaining objects from Clementini et. al (2023). Of this group, 18 objects were identified as W Ursa Majoris variables, while 12 were identified as genuine RR Lyrae. In addition to automated threshold classification, each of the positively identified RR Lyrae was manually inspected and confirmed to exhibit secondary features that aid in classification including an asymmetric lightcurve or notch near maximum luminosity.



*Figure 13: Lightcurve of BW Tri (ID 3) The minima variation is 1.3% -- approximately 10 times larger than the RRc threshold. This object was classified through the Lightkurve pipeline as an eclipsing binary.*



*Figure 14: Lightcurve of V1047 Cas (ID 17). The average minima variation is less than 0.1%, and the lightcurve exhibits highly asymmetric features consistent with other RR Lyrae. This variable was classified as an RR Lyra by our Lightkurve pipeline, and visually appears to be a type A/B variable.*

In addition to these positively confirmed variables, we additionally identified 10 RRC candidates with minima variation below the 0.16% threshold, but without any secondary characteristics – notably a notch on the ascending edge - to confirm their classification.

Finally, we identified several variables in the dataset which do not clearly align with our standards for RR Lyrae or W UMa variables. These variables exhibit highly consistent minima across several periods, with a ‘double hump’ on the falling edge of each period. Based on literature searches of similar variables, we hypothesize that these objects may be RV Tauri stars, a variant of Type II Cepheids that are often mistaken for W Ursa Majoris variables.<sup>[18]</sup> A total of 9 of these objects were identified.

Figure 15: Lightcurve of OGLE-GD-RRLYR-791 (ID 19). Minima variation is less than 0.07%, within the range of our RR Lyra criteria, but the lightcurve lacks secondary characteristics to confirm this classification.

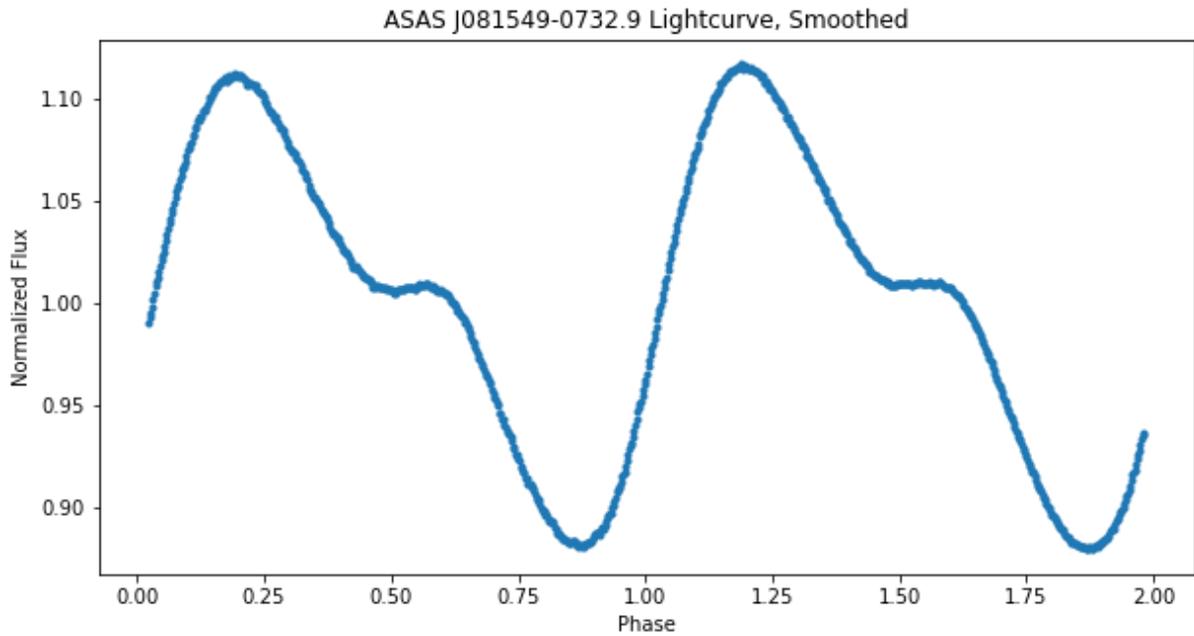
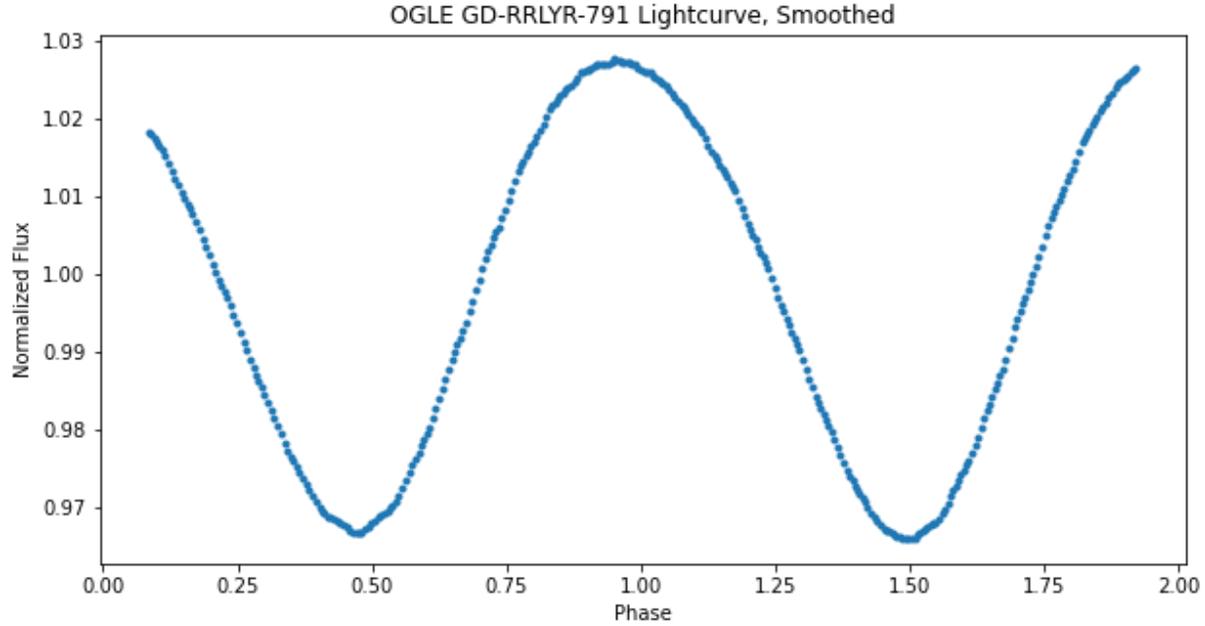


Figure 16: Lightcurve of ASAS J081549-0732.9 (ID 58), a potential RV Tauri star. A 'double hump' is visible on the falling edge of each period. Although this variable was classified as an RR Lyra through our Lightcurve pipeline, it was manually reclassified as a potential RV Tauri.

A full list of all objects studied through the TESS Lightkurve pipeline is available in Table 3, along with their minima variability and classification.

## Chapter 3: Observational Studies

### I: Identification of Candidates

After determining an initial classification for each object, follow-up observations were conducted for a selection of these objects to confirm the TESS-based classification. The Emory observatory utilizes a 24-inch f/8 DFM Cassegrain telescope and an Apogee Alta U47 camera, with 1024x1024 pixels at 13x13 microns each. For each observation, an R-band filter was used with approximately the same characteristics as the Johnson/Cousins Rc filter.

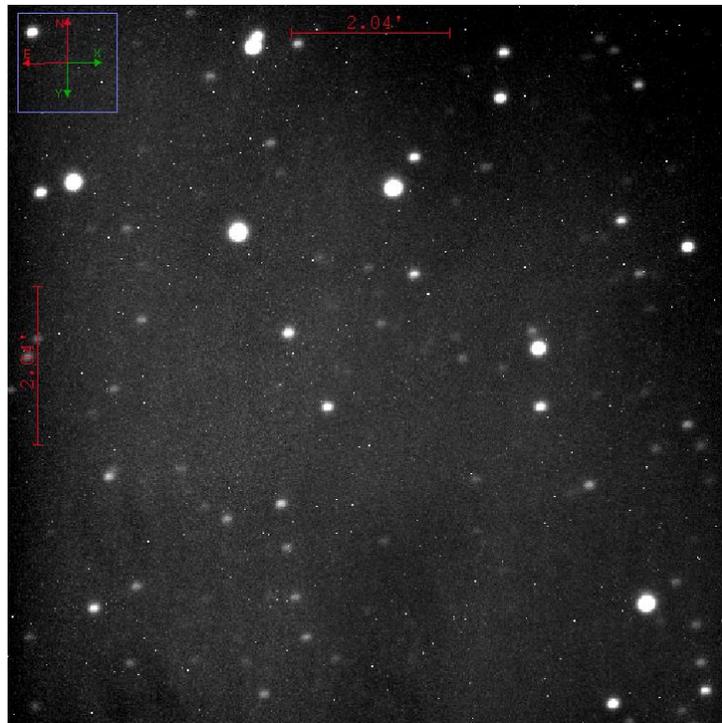
A number of cuts were performed on 72 object datasets from Clementini et. al (2023) to determine each target's suitability for observation. Based on the location of the observatory, only a subset of the objects was observable from Atlanta se during late winter and early spring. We imposed right ascension (RA) and declination (DEC) cuts to the total target list to isolate candidates between 90- and 210-degrees RA, and greater than 0 degrees DEC. This significantly reduced the subject pool to 5 candidates.

Due to adverse seeing and light pollution conditions in Atlanta, as well as the specifications of the Emory observatory, targets were limited to those with a magnitude below +16. This criterion reduced the total subject pool further from 5 to 4.

Due to limited observation time and adverse weather conditions, only 3 of 4 suitable targets were successfully observed. Additional analysis of these candidates, and future observations of additional objects would aid in validating Lightkurve classifications. A list of the three observed targets, observation date, coordinates, and basic photometric information is available in Table 4.

## II: Data Collection

Each target was observed for approximately 4 to 8 total hours depending on visibility and weather conditions. This duration ultimately prevented us from capturing a complete lightcurve of any one target, but allowed us to observe multiple objects for comparative analysis. Exposure time varied depending on the magnitude of each target but was adjusted so that in the initial frame, flux from the target reached approximately half of the saturation limit of the CCD. This selection ensured that any magnitude changes due to inherent variability would be entirely captured within the linearity range. Typical exposure times for targets were between 120 and 150 seconds. Each exposure was saved as a .fit file, with header data indicating target coordinates, target name, observer, date, and exposure time.



*Figure 17: Uncalibrated exposure of  
LAMOST/ATO J106.1517+08.6631*

Adverse weather conditions and light pollution resulted in several sources of error throughout the data collection process. In addition, significant light pollution of approximately 15 frames was observed in 3 of 5 total observation runs. This source of error has been isolated to infrared emission from a camera located in the observatory dome, which has since been removed. In addition, high winds and poor telescope tracking resulted in some frames being out of alignment or blurred. These errors were corrected by plate solving, normalizing target flux to background stars, and removing unusable frames.

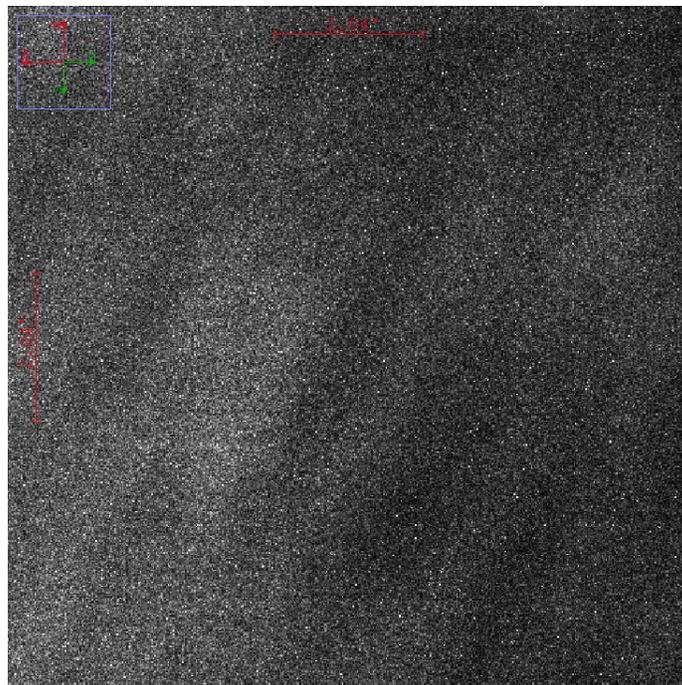
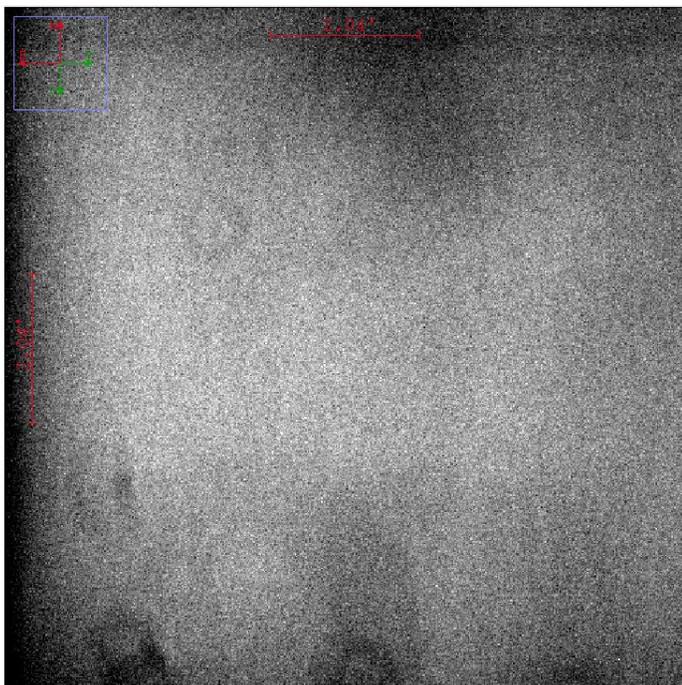
### **III: Calibration**

After each night of observation, a series of flat, dark, and bias calibration frames were collected. These calibration images were used to correct flat field response, thermal noise, and readout noise, respectively.

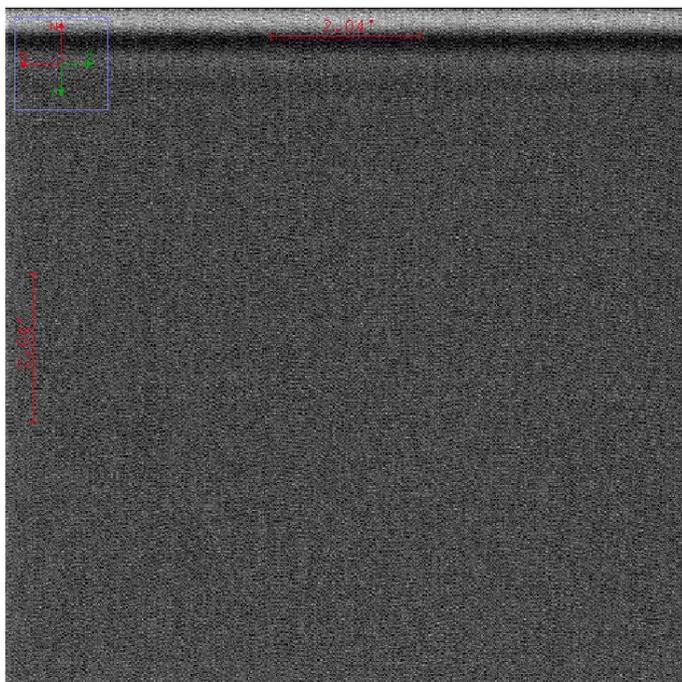
Flat frames were collected by exposing the telescope to against a uniformly illuminated white background with an exposure time between 1 to 2 seconds. Exposure time and illumination were adjusted to minimize the standard deviation of flux across the CCD pixels. These frames measure the flat-field response of the telescope and correct for the effects of dust and other imperfections in the optical system. 20 flat calibration images were obtained for each observation run.

Dark frames were collected by exposing the CCD for the same duration as target data while the CCD shutter was closed and the telescope lens cap in place. Thus, any flux observed by the CCD was due to thermal and readout noise. Between 10 and 20 such images were obtained per observation night.

Finally, bias frames were generated by measuring the flux from the CCD with zero exposure time, and the shutter closed. Again, 20 bias frames were collected per night. This



isolates any flux recorded to the readout noise of the CCD.



*Figure 18: Calibration data: a dark calibration frame (upper right), flat field calibration frame (upper left), and bias calibration frame (lower left).*

#### **IV: Processing**

All target files were opened as a virtual stack and analyzed using AstroImageJ.<sup>[17]</sup> Master bias, dark, and flat frames were generated by averaging the flux of individual calibration frames, and each target frame was reduced by subtracting and dividing master calibration frames, as appropriate.

Target images were then aligned using plate solving, provided by astronomy.net. This process automatically identifies the sources present in each image by cross-referencing astronomical databases. Each identified source in one frame is then stacked on-top of its counterpart in another frame, creating a fully aligned set of images. After alignment, reference stars with approximately the same luminosity as the target star were identified in order to measure target variability against a fixed reference point. Each reference star selected was researched in literature to ensure no past identification as a variable and independently verified as non-variable by comparing its flux relative to other stars in the field during data analysis.

Multi aperture photometry was performed on aligned, reduced images. During this process, aperture size was selected by measuring the point spread function of the target star. The size of the inner aperture was selected to capture 95% of the target flux according to its point-spread function, with the middle aperture and outer aperture at 1.5 and 2.5 times the diameter of the inner aperture, respectively.

The inner aperture isolates flux from the target and background, while the annulus between the middle and outer aperture isolates flux from the background. Using the total flux of the background annulus, and the number of pixels in this region, we identify the amount of flux per pixel from background noise. This value is then subtracted from each pixel of the target aperture to isolate flux from the target star.

Flux from all apertures was saved as a CSV file for future analysis.

## **V: Visualization and Analysis**

The relative flux for all apertures was imported into Python for visualization using matplotlib. Flux from the target was divided by the average flux from all selected comparison stars, normalized, and then averaged to derive a reduced lightcurve for the target.

For targets with multiple nights of observation, identical comparison stars were used across each night. The reduced flux data for each observation was combined into a single lightcurve and folded according to the period derived during the Lightkurve pipeline analysis.

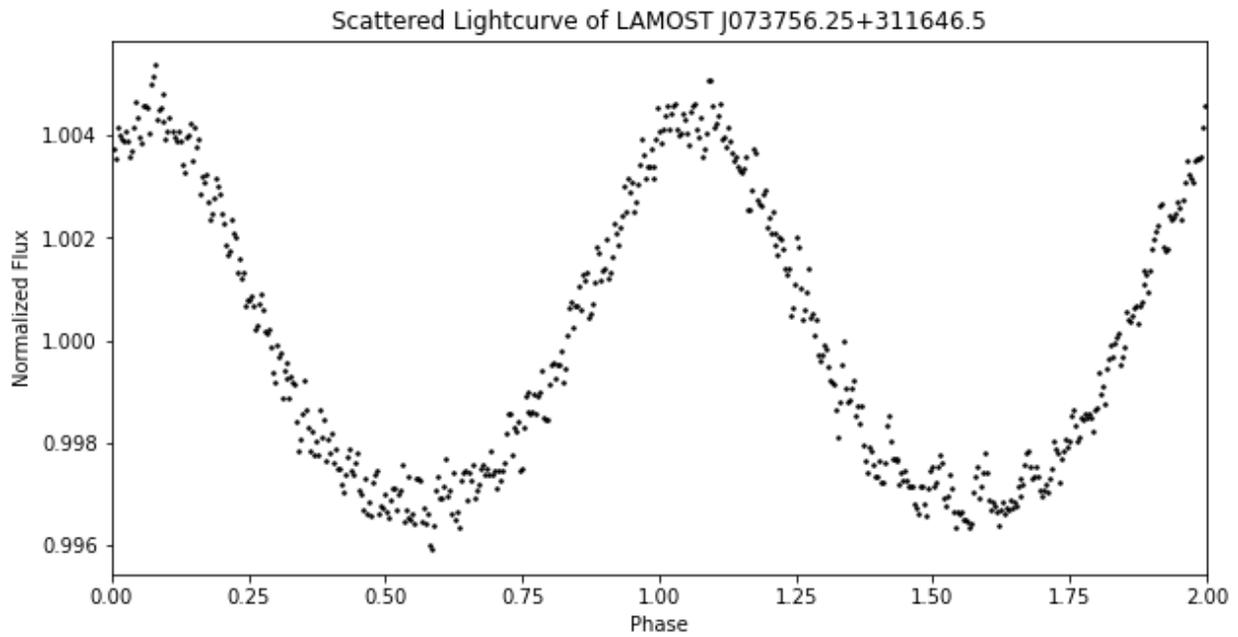
After reducing the data, a rolling average smoothing filter with a window of 3 was applied to each lightcurve to reduce noise from light pollution across each cadence (see Appendix 1). Each target lightcurve was visually inspected to confirm the features identified during the pipeline analysis.

## **VI: Observational Follow-up Results**

### ***LAMOST J073756.25+311646.5***

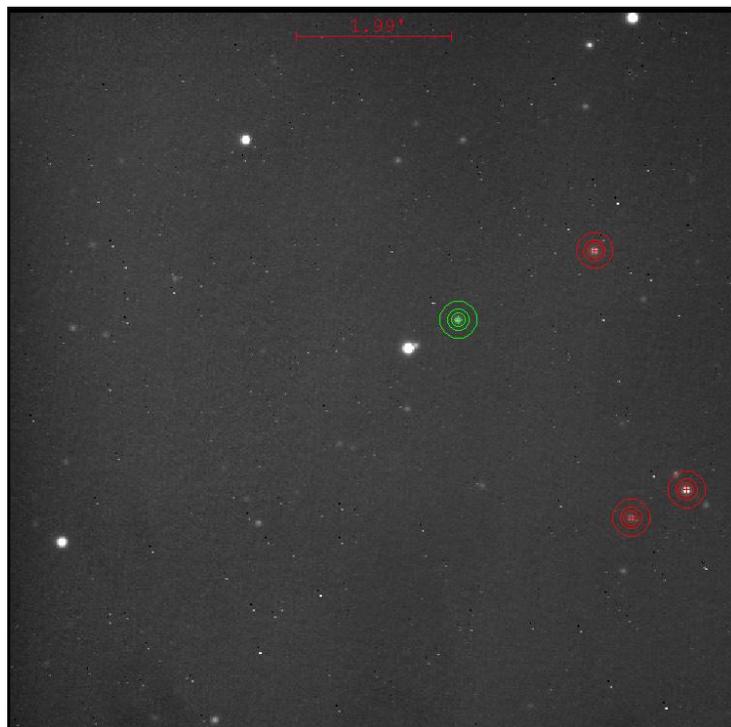
LAMOST J073756.25+311646.5 (LAMOST J07, ID 36) was initially categorized by Clementini et. al. (2023) as an eclipsing binary. TESS analysis of the target indicated a highly symmetric lightcurve, but a transit depth variation of only 0.025%. The low variation in transit depth nominally suggested that this object could be an RRc Lyra, but no secondary lightcurve features were present to confirm this classification.

This target was observed for 4.7 hours on April 25-26, under clear conditions. Wind speeds were initially high and interfering slightly with tracking, but resolved within the first hour of data collection.



*Figure 19: TESS Lightcurve of LAMOST J07073756.25+311646.5. Note that the lightcurve is highly symmetric, there is very little minima variation between periods.*

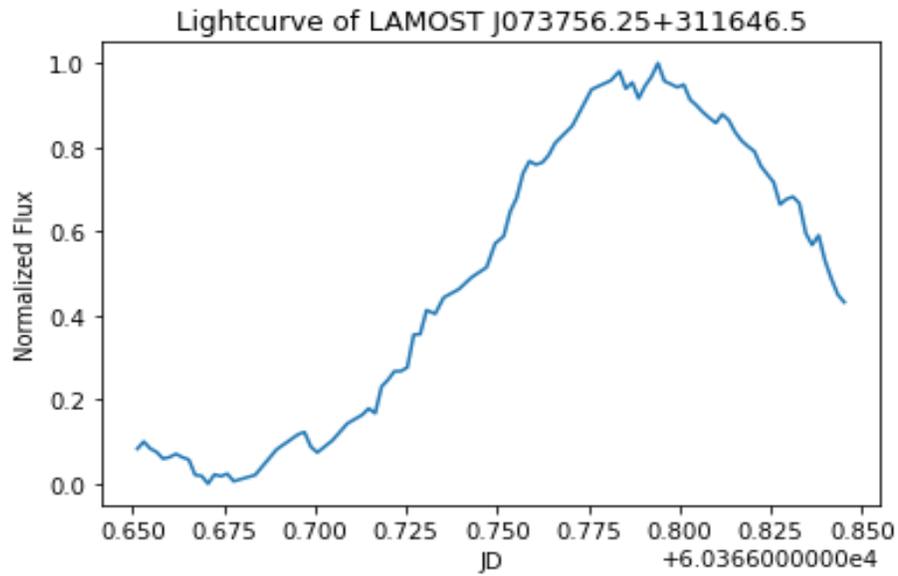
Following the observation, multi-aperture photometry was performed on the target and three reference stars in-frame and exported to Python.



*Figure 20: Reduced frame of LAMOST J07073756.25+311646.5 with apertures. The target star is indicated with a green aperture, while red apertures indicate a reference star.*

Approximately 75% of a complete period was obtained during the observation, and Python analysis of the observational lightcurve confirmed the variability initially observed in TESS data. The lightcurve appears relatively symmetric around the maximum at  $JD = 60366 + 0.8$ , with a period that aligns with the value determined through our TESS pipeline of 0.257 days. Comparing our observational data with the folded TESS lightcurve, we observe that our observations fit the TESS model exceptionally well between  $JD = 0.025$  and  $0.20$ , with some minor deviation near the beginning of data collection. It is possible that this could be an artifact of ambient light changes during twilight, but this may also be a genuine feature of the target's lightcurve. Additional observational data is crucial to confirm the source of this deviation.

Figure 21a: Observational lightcurve of LAMOST J07



LAMOST J073756.25+311646.5 TESS Lightcurve and Observational Data

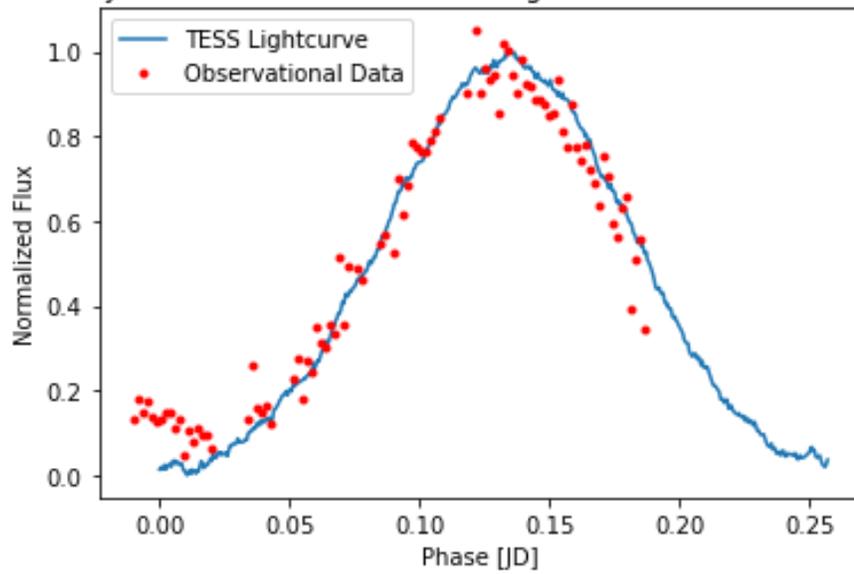
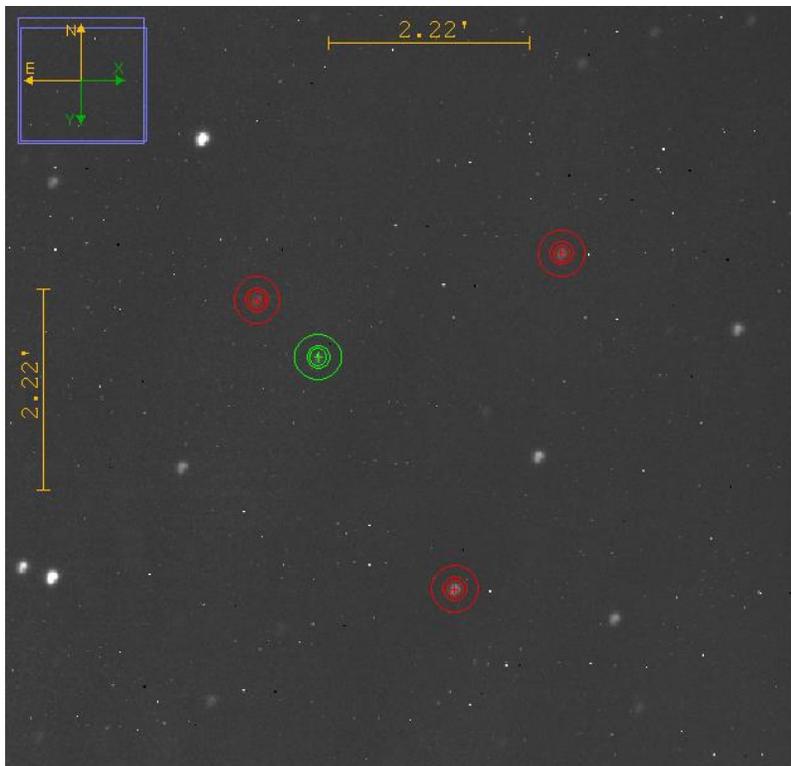


Figure 21b: Smoothed TESS lightcurve of LAMOST J07 with observational data overplotted in red.

### *ASAS J074316+1705*

As with LAMOST J07, ASAS J074316+1705 (ASAS J07, ID: 9) was initially classified by Clementiti et. al. (2023) as an eclipsing binary but was reclassified through the Lightcurve pipeline due to low minima variation of only 0.07%.

This target was observed for 3.9 hours on March 6-7, during clear weather. No significant light pollution was observed during this period, and tracking remained stable. Due to the short period of observation, we expected to only observe approximately 60% of the total lightcurve.



*Figure 22: Reduced frame of ASAS J07 with apertures used for multi-aperture photometry. The target is selected in green while reference stars are selected in red.*

Multi aperture photometry was performed with three reference stars, and the data was processed using Python. This observation primarily captured the rising-edge of the lightcurve, and approximately  $\frac{1}{4}$  of the falling edge.

Observational data for ASAS J07 is exceptionally well modelled by the TESS lightcurve between JD=0.10 and 0.225. This reinforces our classification of this object, and the efficacy of our Lightkurve pipeline for variable classification.

ASAS J074316+1705 TESS Lightcurve and Observational Data

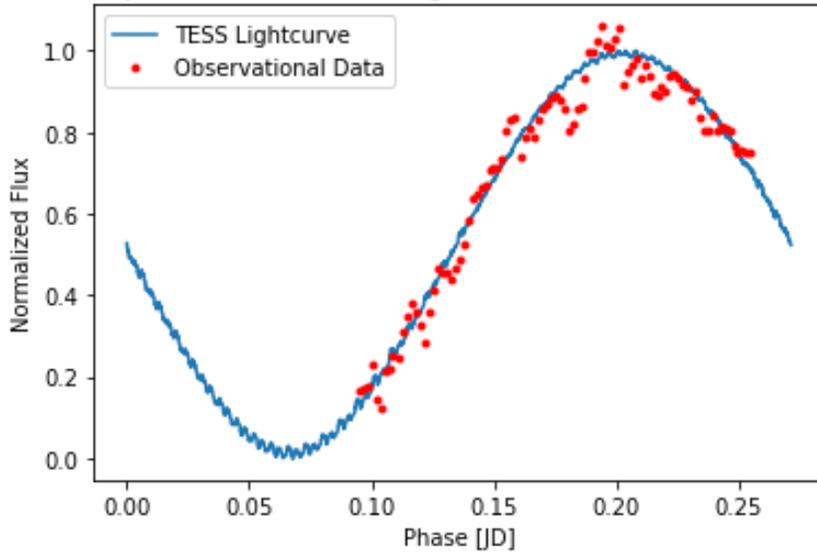
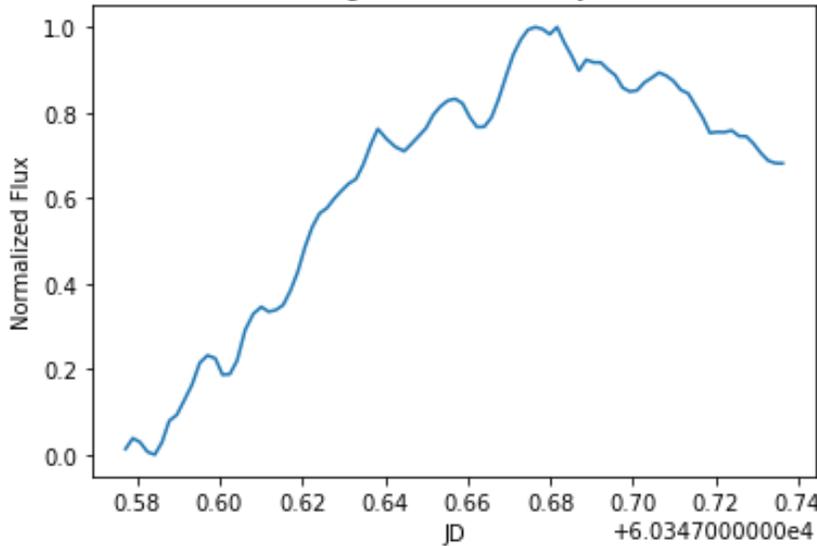


Figure 23: Top: *Smoothed TESS lightcurve of ASAS J07 with observational data overplotted in red. The observational data has been shifted by shifted +0.095 days to align with the TESS model*

Observational Lightcurve of ASAS J074316+1705

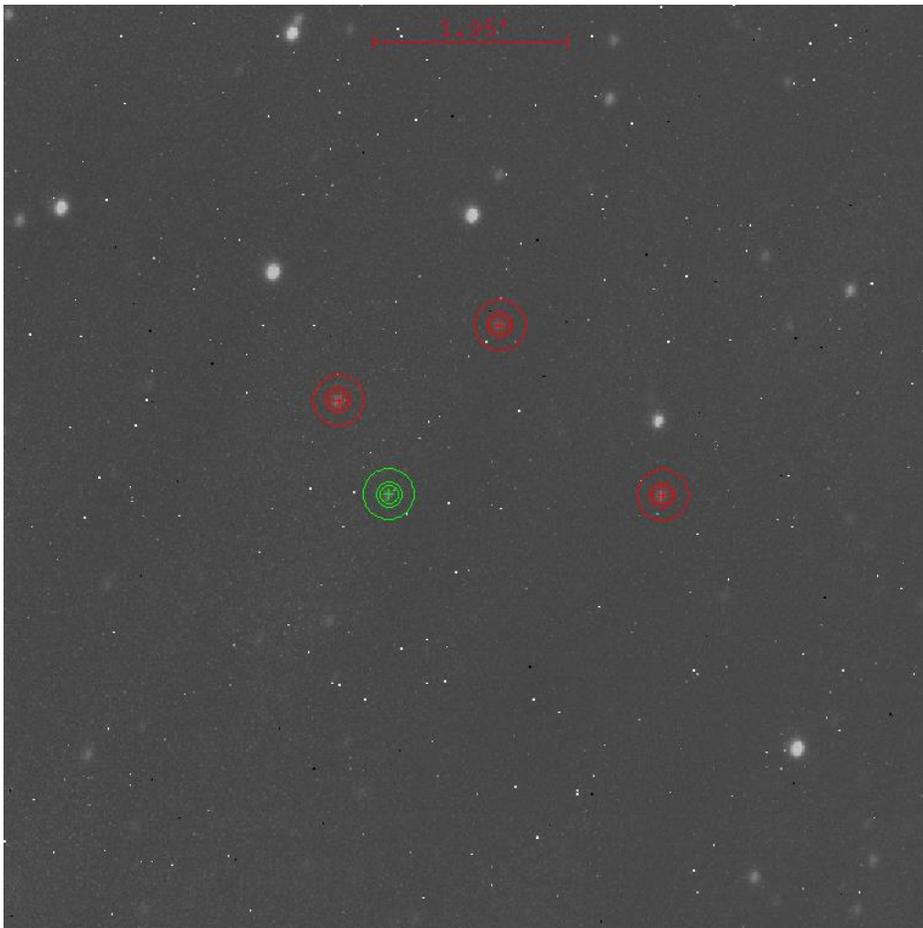


Bottom: *Isolated observational lightcurve of ASAS J07*

**LAMOST/ATO J106.1517+08.6631**

LAMOST/ATO J106.1517+08.6631 (ATO J106, ID 23) was initially categorized by Clementini et. al. (2023) as a low confidence eclipsing binary. While the minima depth variation was high in TESS data, this object exhibited a slight asymmetry near its peak luminosity reminiscent of the notch found on some RR Lyrae. Thus, a primary motivation for observational studies of this variable was to ascertain the validity of these features.

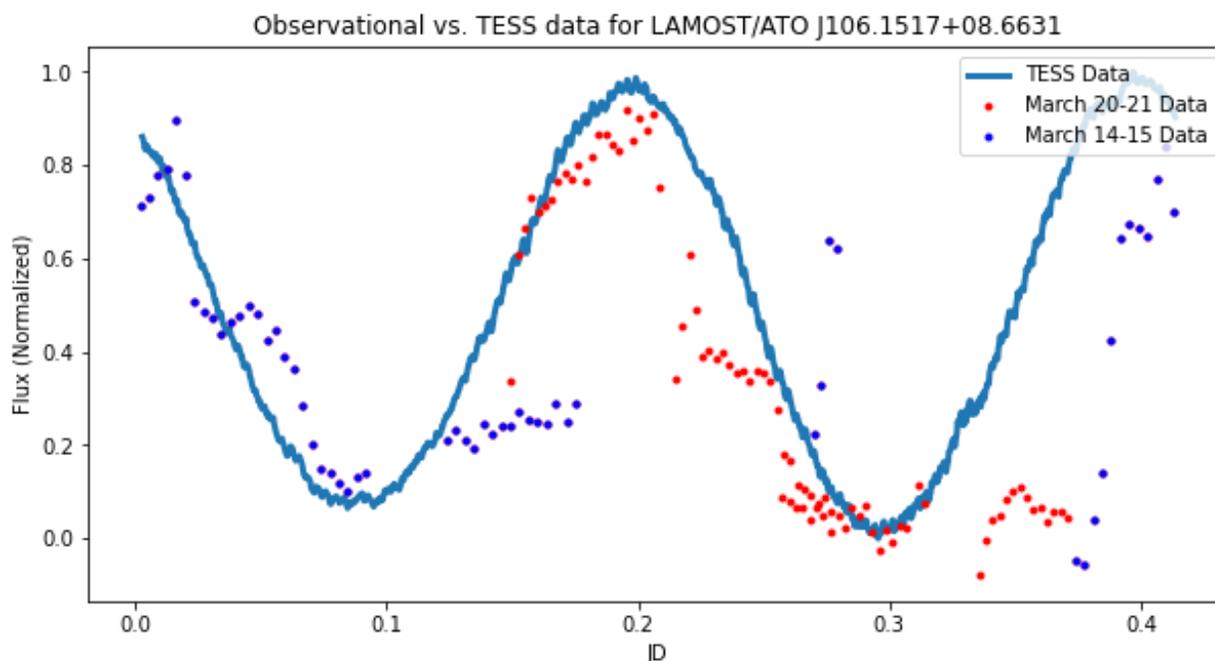
ATO J106 has the longest expected period of all objects considered for follow-up observational study, so data was collected over two separate nights. 6 hours of data was collected on March 14-15 and an additional 4.5 hours were collected on March 20-21.



*Figure 24: A calibrated image of ATO J106 (green) with reference stars highlighted in red.*

Each night of observation data was individually processed and combined into a single dataset. The combined set of observational data was then folded by a period of 0.415 days.

Unfortunately, a significant amount of scatter is present in the observational data for ATO J106 (Fig. 25). During the last quarter of the object's period, there is almost no correlation between the observational data and the expected lightcurve. Between  $JD=0.0$  and  $0.6$ , the observational data aligns reasonably well with our TESS model, and the observed minima fits our expected value. The asymmetry at the peak of the lightcurve near  $JD=0.4$  is well captured by our observational data, although slightly lower in magnitude than expected. However, due to the significant noise and lack of complete coverage at both minima we cannot positively classify this object. Additional data for the first quarter of this objects lightcurve would significantly aid future classification attempts.



*Figure 25: Observational data of ATO J106 overplotted with the expected TESS lightcurve. Data for each night of observation are color coded separately. Note that the minima variation observed in TESS is captured by the observational data, reinforcing this object's potential classification as a *W. UMa* variable.*

## Chapter 4: Discussion and Future Research

The full-sky coverage of Gaia offers an unprecedented opportunity to uncover many previously hidden stellar objects, and initial machine learning searches of the Gaia DR3 release has aided in the classification of over 2.1 million stars. Many of these objects are novel RR Lyrae and eclipsing binaries, which may open new opportunities for distance estimation throughout the galaxy. However, due to the low sampling rate of Gaia, crucial distinguishing features of these objects lightcurve may be missed. As a result, photometric classifications of variables in Gaia have a potentially high error rate.

Throughout this research, we have explored the categorization of low-confidence variable objects in Gaia DR3 using a number of different methodologies which aim to reduce error in these classifications. Cross-checking uncertain variables using TESS archival data effectively compensates for the low sampling rate of Gaia. However, due to the large pixel size and limited coverage, this approach is only applicable to a subset of Gaia targets.

The minima variation classification pipeline developed for this research offers a robust method of excluding some W. Ursa Majoris variables from RR Lyrae, but alone is not able to positively identify all RR Lyrae from other types of variable object. For instance, while we can confidently ascertain that BW Tri (Fig. 12) is *not* an RR Lyra based on its high minima variation, we cannot *exclude* the possibility that OGLE-GD-RRLYR-791 (Fig. 11) is an eclipsing binary purely based on its low minima variation.

Although the PLD algorithm used to process TESS data is highly effective in many cases, some instances were noted in which it overcorrected or failed to subtract noise from the lightcurve of bright objects (Appendix 3). In addition, it is possible that this algorithm could have influenced the minima depth variation measured through the TESS pipeline, however, additional testing is required to ascertain the significance of these effects.

Still, this pipeline provides a highly useful tool for folding, processing, and visualizing TESS lightcurves of variable stars. Combining minima variation with visual inspection for secondary lightcurve features, we are able to classify with high confidence multiple genuine RR Lyrae that were otherwise misclassified in the Gaia DR3 dataset.

Follow-up observational studies of these objects provides yet another tool to verify the variability identified both in TESS and Gaia. While Gaia has a high angular resolution, its cadence is not high enough to effectively capture the details of a short-period variable. TESS has a much higher cadence than GAIA, but often fails to resolve target stars from other objects in close proximity. When available, observational follow-up studies potentially mitigate both of these issues, offering the best independent verification method for stellar classifications.

Unfortunately, due to limited observation time, this study only partially utilized the full capabilities of this technique. Observational follow-up data of both LAMOST J073756.25+311646.5 and ASAS J074316+1705 visually align well throughout the period of observation and reinforces the efficacy of our Lightkurve processing pipeline. However, neither object was observed for their full period, and so we are unable to empirically confirm their minima variation. While noisy, observational data of LAMOST/ATO J106.1517+08.6631 captures the minima variation initially determined using TESS data and supports the W. Ursa Majoris classification of our pipeline. Future ground-based observations are crucial to verify both the minima variation and lightcurve features of W. UMa and RR Lyra variables identified in this paper.



## References

- [1] [3] Smith, H. A. (1995). *RR Lyrae Stars*. Cambridge University Press.
- [2] Dederick, E., & Jackiewicz, J. (2017). A possible mechanism for driving oscillations in hot giant planets. *The Astrophysical Journal*, 837(2), 148. <https://doi.org/10.3847/1538-4357/aa61a8>
- [4] Soszynski, I. & Udalski, A. & Szymanski, M. & Pietrukowicz, P. & Mroz, P. & Skowron, J. & Kozłowski, Szymon & Poleski, R. & Skowron, D. & Pietrzynski, G. & Wyrzykowski, Lukasz & Ulaczyk, K. & Kubiak, M.. (2014). Over 38 000 RR Lyrae stars in the OGLE galactic bulge fields. *Acta Astronomica*. 64.
- [7] Prusti, T., de Bruijne, J. H., Brown, A. G., Vallenari, A., Babusiaux, C., Bailer-Jones, C. A., Bastian, U., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Milligan, D. J., Panem, C., ... Zschocke, S. (2016). The *gaia* mission. *Astronomy & Astrophysics*, 595. <https://doi.org/10.1051/0004-6361/201629272>
- [9] Clementini, G., Ripepi, V., Garofalo, A., Molinaro, R., Muraveva, T., Leccia, S., Rimoldini, L., Holl, B., Jevardat de Fombelle, G., Sartoretti, P., Marchal, O., Audard, M., Nienartowicz, K., Andrae, R., Marconi, M., Szabados, L., Evans, D. W., Lecoœur-Taïbi, I., Mowlavi, N., ... Eyer, L. (2023). Gaia data release 3. *Astronomy & Astrophysics*, 674. <https://doi.org/10.1051/0004-6361/202243964>
- [10][11] Massachusetts Institute of Technology. (n.d.). *Observations*. TESS Transiting Exoplanet Survey Satellite. <https://tess.mit.edu/science/observations/>
- [12] Lightkurve Collaboration et. al. 2018: Lightkurve: Kepler and TESS time series analysis in Python, *Astrophysics Source Code Library* ascl:1812.013
- [13] Deming, D., Knutson, H., Kammer, J., Fulton, B. J., Ingalls, J., Carey, S., Burrows, A., Fortney, J. J., Todorov, K., Agol, E., Cowan, N., Desert, J.-M., Fraine, J., Langton, J., Morley, C., & Showman, A. P. (2015). *spitzer* secondary eclipses of the dense, modestly-irradiated, giant exoplanet  $\text{HD 189733 B}$  using pixel-level decorrelation. *The Astrophysical Journal*, 805(2), 132. <https://doi.org/10.1088/0004-637x/805/2/132>
- [14] Luger, R., Agol, E., Kruse, E., Barnes, R., Becker, A., Foreman-Mackey, D., & Deming, D. (2016). Everest: Pixel level decorrelation of K2 light curves. *The Astronomical Journal*, 152(4), 100. <https://doi.org/10.3847/0004-6256/152/4/100>

- [16] Sneden, C., Preston, G. W., Chadid, M., & Adamów, M. (2017). The RRC stars: Chemical Abundances and envelope kinematics. *The Astrophysical Journal*, 848(1), 68. <https://doi.org/10.3847/1538-4357/aa8b10>
- [6] Lucy, L. B. (1968). The light curves of W Ursae Majoris stars. *The Astrophysical Journal*, 153, 877. <https://doi.org/10.1086/149712>
- [8] Boubert, D., Everall, A., & Holl, B. (2020). Completeness of the gaia-verse – I. when and where were Gaia’s eyes on the sky during DR2? *Monthly Notices of the Royal Astronomical Society*, 497(2), 1826–1841. <https://doi.org/10.1093/mnras/staa2050>
- [15] *Lightkurve.keplertargetpixelfile.create\_threshold\_mask Documentation*. Lightkurve. (n.d.). [https://docs.lightkurve.org/reference/api/lightkurve.KeplerTargetPixelFile.create\\_threshold\\_mask.html?highlight=create\\_threshold\\_mask](https://docs.lightkurve.org/reference/api/lightkurve.KeplerTargetPixelFile.create_threshold_mask.html?highlight=create_threshold_mask)
- [17] Collins, K. A., Kielkopf, J. F., Stassun, K. G., & Hessman, F. V. (2017). Astroimagej: Image processing and photometric extraction for ultra-precise astronomical light curves. *The Astronomical Journal*, 153(2), 77. <https://doi.org/10.3847/1538-3881/153/2/77>
- [18] Soszyński, I. (2017). The OGLE Collection of Variable Stars. Classical, Type II, and Anomalous Cepheids toward the Galactic Center. *Acta Astronomica*, 67(297). <https://doi.org/10.48550/arXiv.1712.01307>
- [19] NASA. (n.d.). *Observing technical details*. NASA. <https://heasarc.gsfc.nasa.gov/docs/tess/observing-technical.html>
- [20] Cabral, J. B., Ramos, F., Gurovich, S., & Granitto, P. M. (2020). Automatic catalog of RR Lyrae from ~14 million VVV light curves: How far can we go with traditional machine-learning? *Astronomy & Astrophysics*, 641, A83. DOI: 10.1051/0004-6361/202038314

## Tables

**Table 1: Minima variation for confirmed RR Lyrae**

<b>ASAS Identifier</b>	<b>Minima Difference</b>	<b>Notes</b>
023706-4257.8	0.001536856	
094541-0644.0	0.00113081	
095328+0203.5	0.001029559	
101332-0702.3	0.001321967	
123811-1500.0	0.000175428	
132225-2042.3	0.000723183	Figure 10
143322-0418.2	0.002134002	
190212-4639.2	0.000551584	

**Table 2: Minima variation for confirmed W. Ursa Majoris variables**

<b>Identifier</b>	<b>Minima Difference</b>	<b>Pipeline Classification</b>
AB And	0.068843333	ECL
S Ant	0.037739624	ECL
TU Boo	0.027569353	ECL
eps CrA	0.024405277	ECL
SX Crv	0.026107652	RR
V1191 Cyg	0.014564398	ECL
XY Leo	0.054092986	ECL
CE Leo	0.048218453	ECL

**Table 3: Minima variation and classification for Gaia DR3 Variables**

Target ID	Gaia Classification	RA	DEC	Minima Variability	Lightkurve Classification	Secondary Features?
1	ECL	35.52084	20.0213	0.043951	ECL	
2	ECL	58.8649	81.3534	0.000885	RR	Yes
3	ECL	34.02603	34.63023	0.016097	ECL	
4	ECL	329.7278	-24.5261	0.023571	ECL	
5	ECL	114.2033	-60.1085	0.019258	ECL	
6	ECL	57.68542	-72.4979	0.009485	ECL	
7	ECL	102.2204	-13.2856	0.00422	ECL	
8	ECL	125.4739	-30.6887	0.010308	ECL	
9	ECL	115.8151	17.09085	0.000783	RR	Yes
10	ECL	250.649	-46.5693	0.006996	ECL	
11	RRc/ECL?	32.91502	10.88761	0.000893	RR	Yes
12	ECL	123.8855	-55.9209	0.003204	ECL	
13	RRL?/ECL?	135.7902	-38.5343	0.000182	ECL	Yes
14	ECL	233.336	-49.2179	Faint		
15	ECL	354.7571	57.76953	0.002764	ECL	
16	ECL	133.7134	-13.6517	0.003426	ECL	
17	ECL?	10.50622	54.25137	0.000929	RR	Yes
18	ECL	356.0142	38.80609	0.002554	ECL	
19	ECL	126.4544	-49.696	0.000147	RR	
20	ECL	339.1527	49.90346	0.001223	ECL	
21	ECL	111.3283	-9.03204	0.002867	ECL	
22	ECL	284.6458	-9.39006	No data		
23	LikelyECL	106.1518	8.66311	0.002938	ECL	
24	ECL	125.1369	-59.9995	0.00021	RR	

25	ECL	338.188	-11.4205	0.000493	RR	Yes
26	ECL	8.621	73.55835	4.32E-05	RR	
27	ECL	268.5393	-60.8327	0.00019	RR	
28	ECL	126.4544	-49.696	0.000693	RR	
29	ECL	278.084	14.76311	Faint		
30	ECL	23.05287	-49.5613	0.000216	RR	
31	ECL	260.0214	-33.3536	No data		
32	ECL	134.7057	-44.4554	Faint		
33	ECL	5.48508	37.94372	Faint		
34	ECL	92.74734	32.73736	Faint		
35	ECL/RRc?	126.2956	-37.4907	Faint		
36	ECL?	114.4844	31.27955	0.000252	RR	
37	ECL?	349.0122	-32.7783	0.000161	RR	
38	ECL?	42.44324	-69.1841	0.000227	RV Tauri	
39	ECL?	317.5578	-38.8975	0.001566	RV Tauri	
40	ECL?	217.2137	-6.95105	0.000536	RR	Yes
41	ECL?	111.3833	-32.0116	0.000158	RV Tauri	
42	ECL?	161.9256	-55.1707	0.000324	RR	
43	ECL?	350.9276	-41.9568	0.000597	RV Tauri	
44	ECL?	274.4136	10.30744	No data		
45	ECL?	329.3296	58.72098	Faint		
46	ECL?	326.4337	16.49886	0.005078	ECL	
47	ECL?	222.0758	-49.2072	0.001364	RV Tauri	
48	ECL?	241.056	-28.1772	0.001532	RR	Yes
49	ECL?	268.2061	-40.6731	0.000171	RR	Yes
50	ECL?	227.8425	-51.6455	0.002026	ECL	
51	ECL?	74.62177	2.03769	0.00028	RR	Yes
52	ECL?	100.1365	10.66434	0.000293	RR	Yes

53	ECL?	73.22596	-53.2401	0.000564	RV Tauri	
54	ECL?	77.61917	-46.2853	0.001231	RV Tauri	
55	ECL?notRRL	160.2138	-64.7132	0.000318	RR	Yes
56	uncertain-classification	268.9849	-22.9983	No data		
57	uncertain-classification-NotRRL	184.6051	-43.052	0.010851	ECL	
58	uncertain-classificationR Rab?/ECL?	123.9549	-7.5481	0.002576	RV Tauri	
59	uncertain-classification	250.7021	13.79764	No data		
60	uncertain-classification	322.4973	12.16906	Bright neighbor		
61	uncertain-classification	255.3072	-30.1072	Bright neighbor		
62	uncertain-classificationR Rab?	323.3584	-0.82164	bright neighbor		
63	uncertain-classificationR Rc/ECL?	249.4204	-30.5277	0.000127	RR	
64	uncertain-classification	305.642	47.04218	No data		
65	uncertain-classification-RRc/ECL?	268.4189	-29.3085	No data		
66	uncertain-classification-RRab?	258.5605	-37.6352	Faint		
67	uncertain-classification-RRc/ECL?	271.4211	-34.5228	Faint		

68	uncertain-classification-RRab/ECL?	251.8863	-20.3443	No data		
69	uncertain-classification	201.6674	-47.4434	Bright neighbor		
70	uncertain-classification-RRab??	286.3201	16.3251	0.001535	ECL	
71	uncertain-classification	322.5036	12.16567	Bright neighbor		
72	NotRRL/ECL?	356.2311	44.40214	0.003462	RV Tauri	

**Table 4: Information for Observation Targets**

<b>Target ID</b>	<b>RA (Deg.)</b>	<b>DEC (Deg.)</b>	<b>G Mag. (Gaia)</b>	<b>Period (Days)</b>	<b>Date Observed</b>
9	115.8151	17.09085	11.697312	0.27089	2/06-07/24
23	106.1518	8.66311	13.205950	0.41528	2/14-15/24 2/20-21/24
36	114.4844	31.27955	13.553999	0.25711	2/25-26/24

## Figures

- 1: Comparison of RRab Lyra and RRc Lyra Lightcurves
- 2: W. Ursa Majoris Lightcurve
- 3: *Targetpixelfile* Cutout of ATO J106.1517+08.6631
- 4: Comparison of Threshold Task and Custom Mask for ATO J106.1517+08.6631
- 5: Uncorrected TESS Lightcurve of LAMOST/ATO J106.1517+08.6631
- 6: LAMOST/ATO J106.1517+08.6631 Cutout with Aperture and Background Masks
- 7: PLD-Corrected Lightcurve of LAMOST/ATO J106.1517+08.6631
- 8: Periodogram of OGLE GD-RRYLR-791
- 9: Period Folded Lightcurve of OGLE GD-RRYLR-791
- 10: CD-25 15687 Folded by Half and Full Periods
- 11: Scattered and Smoothed Lightcurve of ASAS 132225-2042.3
- 12: Lightcurve of S. Ant
- 13: Lightcurve of BW Tri
- 14: Lightcurve of V1047 Cas
- 15: Lightcurve of OGLE-GD-RRLYR-791
- 16: Lightcurve of ASAS J081549-0732.9
- 17: Uncalibrated exposure of ATO J106
- 18: Calibration Data for Observation
- 19: TESS Lightcurve of LAMOST J07073756.25+311646.5
- 20: Reduced Frame of LAMOST J07073756.25+311646.5
- 21: Lightcurve of LAMOST J07073756.25+311646.5 with Observational Data
- 22: Reduced Frame of ASAS J074316+1705
- 23: Lightcurve of ASAS J074316+1705 with Observational Data

- 24: Reduced Frame of LAMOST/ATO J106.1517+08.6631
- 25: Lightcurve of LAMOST/ATO J106.1517+08.6631 with Observational Data
- 26: Pure and Noisy Signal
- 27: Recovered Signal through Smoothing
- 28: Comparison of Smoothing Window Effects on CD-25 15687
- 29: Comparison of Uncorrected and PLD Corrected W. Ursa Majoris Lightcurve

## Appendix 1: Smoothing Lightcurve Data

Given an array of flux data, a moving average is calculated by computing the average of all  $N$  points both before and after a test point  $I$  (including  $I$ ), where  $N$  is the window size of the rolling average. This process effectively reduces the high frequency noise of a dataset. To demonstrate this efficacy of moving averages to reduce noise, we apply uniform random noise with a maximum and minimum value between  $-0.3$  and  $0.3$  to a perfectly sinusoidal signal with an amplitude of 2.

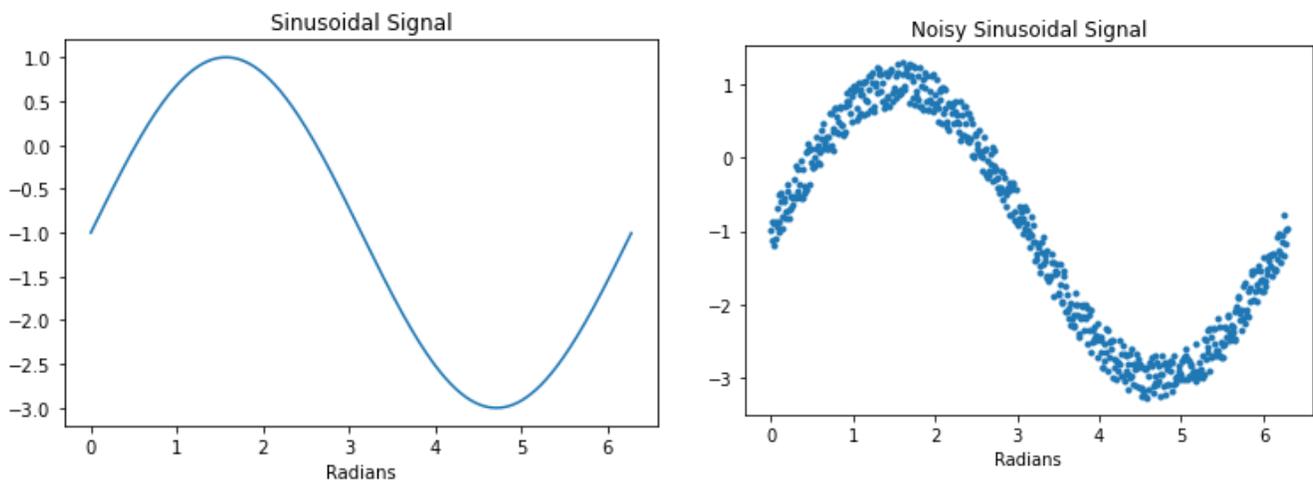
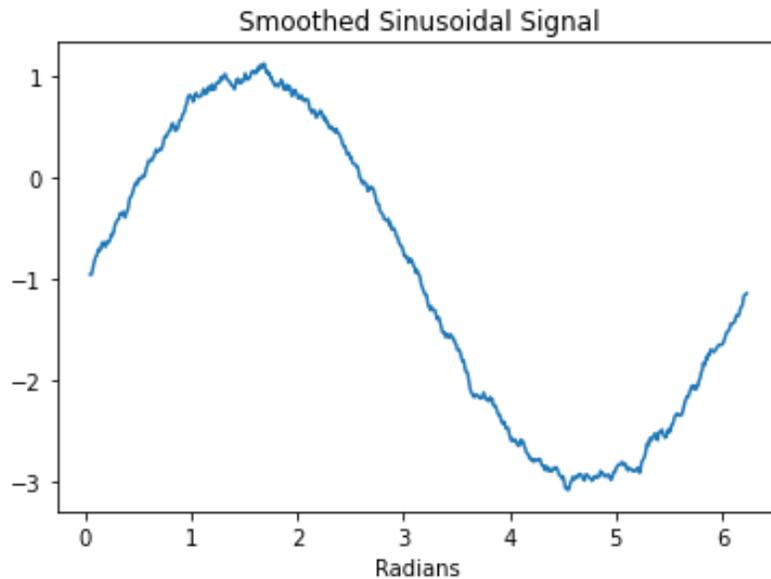


Figure 26: Pure sinusoidal signal and noisy sinusoidal signal. Noise was applied through the addition of a uniformly distributed random value between  $-0.3$  and  $0.3$ .

The addition of noise to this signal significantly impairs analysis. We now apply a moving average with a smoothing window size of 5 to this dataset. While there is still some noise present, the addition of smoothing significantly improves the visibility of the underlying trend.



*Figure 27: The noisy data (Fig. 26) after smoothing. The averaging process aids in recovery of the original signal.*

In the analysis of both TESS and observational data, we similarly use moving averages extensively to reduce noise. However, it is valuable to investigate the effects this process may have on the shape and amplitude of our target lightcurves.

We explore this question by testing rolling averages with varying window sizes on a target with extremely clear raw observational data. The following lightcurve corresponds to CD-25 15687 (ID 4), classified as an eclipsing binary due to its extremely large minima variation.

We apply a rolling average with window sizes ranging between 10 and 80 and observe its effects. As the window size increases from 0 to 20, short period noise present in the lightcurve is reduced without significantly altering the overall shape of the lightcurve. Furthermore, the measured minima variation remains approximately constant.

As the window size increases beyond 20, we observe significant deviations from the unsmoothed lightcurve. Both minima and maxima are reduced in magnitude. We also observe large changes in the measured minima difference.

Based on this analysis, we conclude that a smoothing window size of 20 is sufficient to eliminate high frequency noise without compromising the overall integrity of the lightcurve. This window size is used throughout our TESS data analysis to process folded lightcurves.

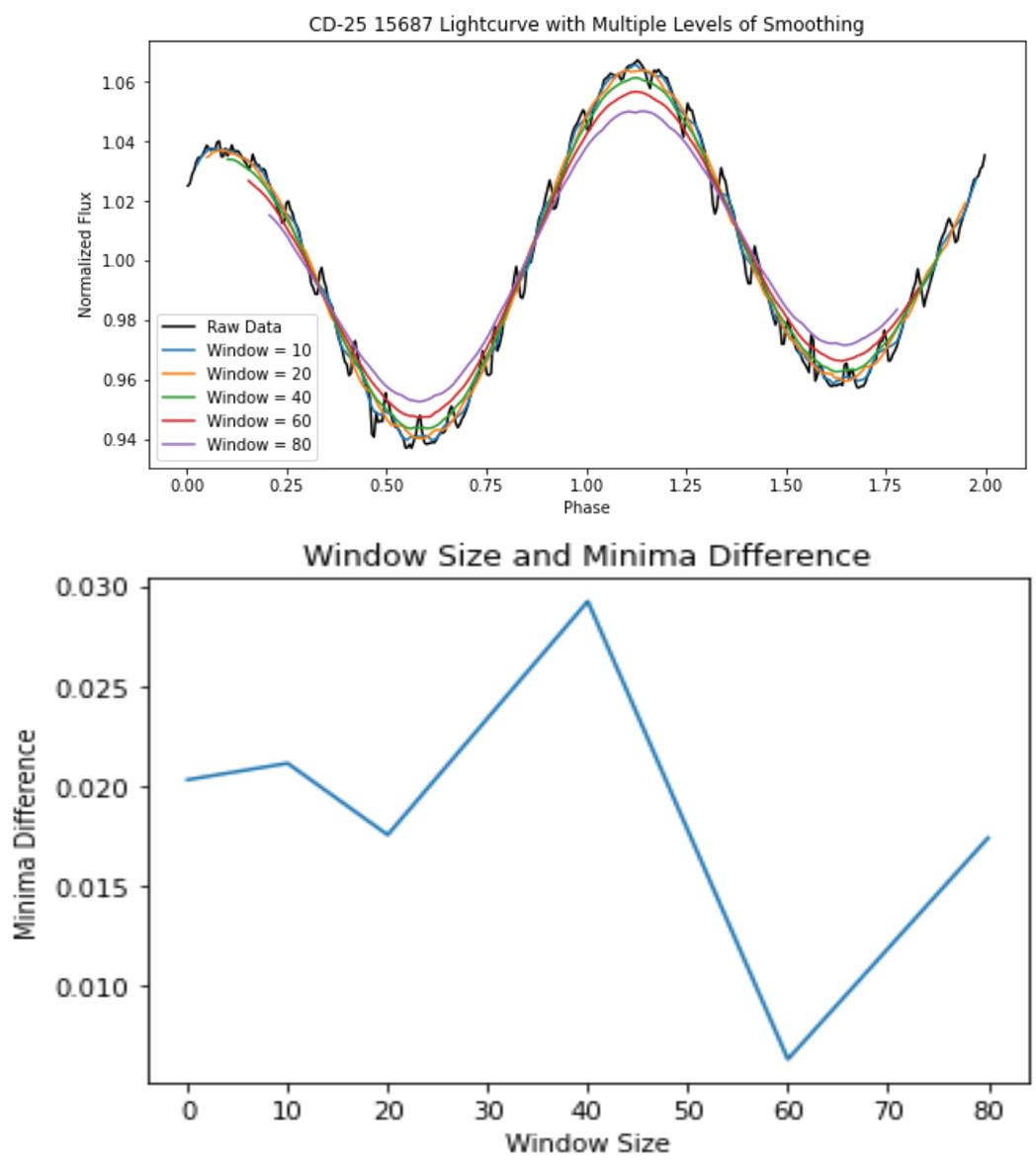


Figure 28: A comparison of various smoothing windows on a noisy dataset. At a window size of 20, we obtain the greatest reduction in noise while preserving important characteristics of the target lightcurve.

## **Appendix 2: The Lightkurve RRc Pipeline**

The full source code for the Lightkurve RRc Pipeline can be found on GitHub at <https://github.com/EdenSchapera/RRcLightkurvePipeline>

### Appendix 3: PLD Correction

Pixel level decorrelation (PLD) generates a clean lightcurve by subtracting a model of background and systematic noise from a dataset. This model is generated by examining the relationship between flux in N background pixels (see *Data Reduction and Cleaning*).

Mathematically, this model takes the form:

$$m_i = \sum_l^N a_l \frac{f_{i_l}}{\sum_k f_{i_k}} + \sum_l^N \sum_m^N b_{lm} \frac{f_{i_l} f_{i_m}}{(\sum_k f_{i_k})^2} + \dots$$

where  $\mathbf{m}_i$  is a vector of noise models at time  $\mathbf{t}_i$ ,  $\mathbf{f}_i$  is the flux in the  $\mathbf{I}^{\text{th}}$  pixel at time  $\mathbf{t}_i$ ,  $\mathbf{a}_l$  is the first order PLD coefficient, and  $\mathbf{b}_{lm}$  is the second order PLD coefficient.

The vector  $\mathbf{m}_i$  may be quite large depending on the number of pixels in the background mask, so principal component analysis (PCA) is performed to reduce the number of basis vectors and generalize the model.

After PCA, the model is further optimized using chi-squared minimization:

$$\chi^2 = \sum_i \frac{(y_i - m_i)^2}{\sigma_i^2}$$

and solving for

$$\frac{\partial^2 \chi}{\partial a_i^2} = 0$$

Finally, an overall background spline is generated using the reduced, optimized models at each cadence. By default, the spline is fit using a 5<sup>th</sup> degree polynomial, but this order may be manually adjusted by the user.

Although the PLD data regression technique has proven very effective in the majority of targets analyzed in this research, several anomalies have been noted which are worth exploring in greater detail. In particular, when analyzing particularly bright objects, the PLD process tends to overcorrect data. In the following example, we assess W. Ursa Majoris (G-band magnitude 7.7), which is exceptionally bright compared to the majority of targets we explore from Clementini et. al. (2023).

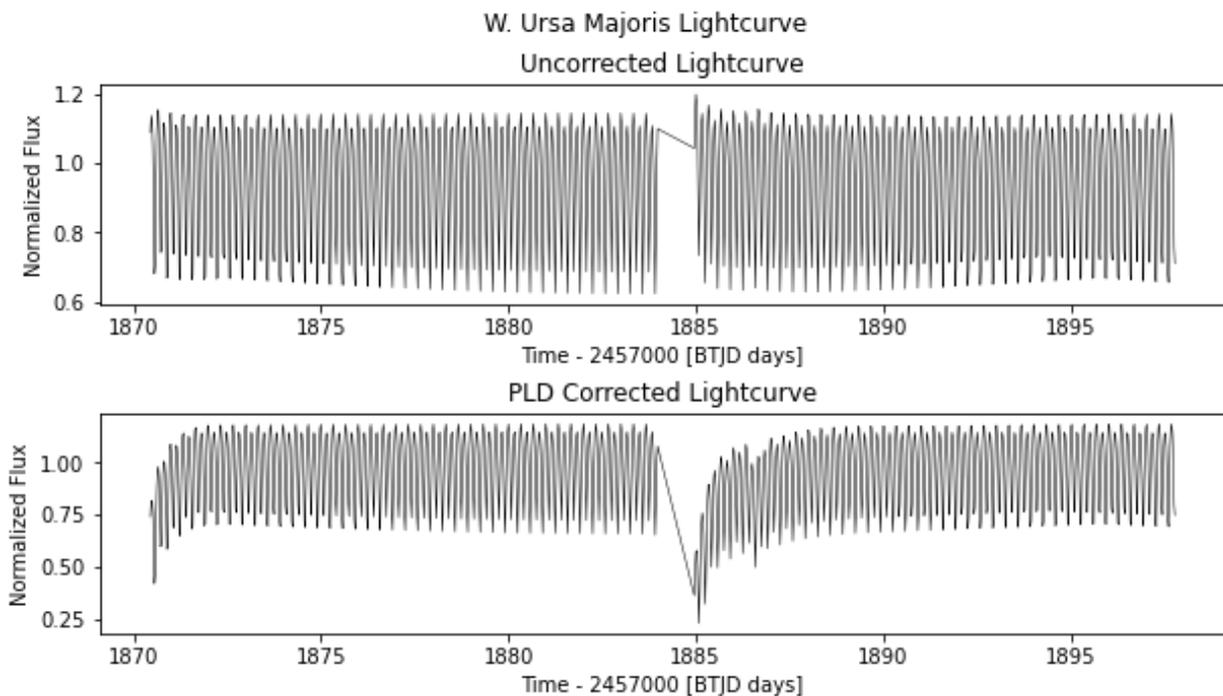


Figure 29: An uncorrected (top) and PLD corrected (bottom) lightcurve of W. Ursa Majoris. The PLD algorithm overfits the target data, resulting in a decrease in flux at BTJD 245700+1870 and BTJD 245700+1885.

We hypothesize that this overcorrection may be due to light from the target star being present in the background mask. When a single bright target dominates a *targetpixelfile* cutout, the mean brightness is significantly skewed in favor of that object. Our standard procedure of selecting pixels for the background mask with a flux greater than one standard deviation below the mean may not be sufficient to entirely isolate purely background regions.

While this effect has significant implications for analyzing bright targets, almost all targets considered in this research had magnitudes greater than 11. Therefore, we are relatively confident that the results of our Lightkurve analysis pipeline remain uncontaminated.