SuperWASP Variable Stars: Cataloging and Identifying Variable Stars by Light Curve Analysis

A Bachelor's Dissertation Thesis

by

Aarushi Rawat

under the supervision of

Dr. Anunay Kr. Choudhary

Submitted to the University of Delhi for the degree of
BACHELOR OF SCIENCE
in Physics

Department of Physics
Sri Venkateswara College
University of Delhi
Dhaula Kuan, New Delhi, Delhi 110021

Certificate

This is to certify that the dissertation project titled "**SuperWASP Variable Stars: Cataloging and Identifying Variable Stars by Light Curve Analysis**" submitted to the **University of Delhi** for the degree of **Bachelor of Science in Physics**, is the record of bona fide research work done by Aarushi Rawat under my supervision and guidance.

Dr. Anunay Kr. Choudhary
(Supervisor)
Assistant Professor
Sri Venkateswara College
University of Delhi
New Delhi-110021

Contents

Acknowledgement

I would like to express my heartfelt gratitude towards my mentor Dr. Anunay Kr. Choudhary. It is through his guidance and patience that has enabled me to produce this work. Needless to say, his unwavering support and belief encouraged me throughout this project.

I would also like to thank my institution, Department of Physics, Sri Venkateswara College for giving me the opportunity to work on this project

I would also like to extend my sincere thanks to my junior, Ms Kamalpreet Kaur who has played an important role in this project by devoting a significant portion of her time and efforts and acting as a constant support. I would also like to thank my peer Ms Sanu Soumya for guiding we with her invaluable computational skills and my senior Mr Mrunmoy Jena,whose prior experience and knowledge proved invaluable in the completion of this project.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support.

This project uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

This project makes use of data from the DR1 of the WASP data (Butters et al. 2010) as provided by the WASP consortium, and the computing and storage facilities at the CERIT Scientific Cloud, reg. no. CZ.1.05/3.2.00/08.0144 which is operated by Masaryk University, Czech Republic.

This project has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (DOI: 10.26093/cds/vizier). The original description of the VizieR service was published in A&AS 143, 23

List of Tables

List of Figures

Abstract

Stars hold the most integral part in teaching us about the universe. They are the primary engines of the cosmic evolution and the birthplace of all the elements found in the nature. The elements that make up all of us. The study of variable stars, specially, reveals immensely invaluable insights in the better understanding of our universe. Their study can give us information of the various stellar properties like its evolution, structure, mass, luminosity, radius, temperature and composition. This is because it the very nature of the star being variable that provides clues to answer these questions. For example, Cepheid variables help us determine galactic and inter-galactic distances and the age of the universe. Mira Variables give us insights on the evolution of Sun. Cataclysmic variables help us understand large scale disc behaviour which can be extended to study activity inside active galaxies and super massive black holes. Therefore,

"To know our place in the cosmos you have to understand stars. To know the stars, you must understand their variability."

And hence the study of variable stars in extremely important. Variable stars data collected over decades, needs to be systematically studied to avail answers. However most professional astronomers don't have the time to analyse millions of data available, and this is where the contribution of amateur astronomers comes into picture. Through resources like Citizen Science Platforms, amateur astronomers are being able to make highly useful contributions to science by utilizing visual, photographic, spectroscopic data.

Hence, as an effort, an attempt has been made to study 500 light curves of Variable Stars using data from the SuperWASP archive and the Zooniverse Citizen Science Platform. The light curves have been classified according to their variability type and cross-matched with existing surveys to extract unique stars and collect more information about the already classified ones. Their periods, magnitudes and accuracy of classification has been assessed to draw further conclusions.

Chapter 1

Theory

1.1 WASP Project and SuperWASP Cameras

Wide Angle Search for Planets (WASP) is an international consortium established in 2000 by a group of astronomers based in UK. It comprises scientists from various academic institutions like the University of Cambridge (Wide Field Astronomy Unit), the Instituto de Astrofisica de Canarias, the Isaac Newton Group of Telescopes, the University of Keele, the University of Leicester, the Open University, Queen's University of Belfast, and the University of St. Andrews.

After the success of their first WASP0 camera, they developed a much ambitious project of the multidetector SuperWASP camera. There are two identical robotic observatories under the SuperWASP facility; SuperWASP-N, at Observatorio del Roque de los Muchachos on the island of La Palma in the Canary Islands and SuperWASP-S at the Sutherland Station of the South African Astronomical Observatory.

1.1.1 The Hardware System

Each SuperWASP instrument contains an equatorial mound on which eight wide-field Canon cameras are installed and together they cover northern and southern hemisphere of the sky, excluding the galactic plane. Each instrument has a focal length of 200mm, an aperture of 14cm (f1.8), and a field of view of 7.8x7.8 deg^2, with total allowed sky coverage of 482 deg^2 squared and an angular scale of 13".7 $pixel^{-1}$ per exposure. These lenses also have excellent apochromatic properties. These are backed by Charged Coupled Devices (CCDs) that consist of 2048x2048 Pixels, each 13.5µm in size.[1]

The main scientific aim of this project is to perform ultra-wide search for exoplanets and to search for bright transiting exoplanet systems suitable for spectroscopic follow-up observations. These observatories allow monitoring of millions of stars simultaneously, taking images as fast as one per minute and collecting up to 100GB data per night.

The first 6-month season of SuperWASP-North observations produced light curves of $\approx$ 6.7 million objects with 12.9 billion data points. [1]

Once the pictures are clicked, the huge quantities of data are stored using a Distributed Data Acquisition Cluster, with each detector controlled by a dedicated Data Acquisition System PC with local storage disks. The Data Acquisition System software is a high-level software which controls the entire SuperWASP system, that is, its robotic mound, the CCD Cameras and other additional hardware systems. This software is provided by a modified version of the open-source Linux software Talon. It also has extensions to include support for the multiple CCD camera and some in-house modifications to add a command-line interface to supplement the standard graphical interface.

This storing of data is in turn controlled by a central machine TCS i.e., Telescope Control System. Overnight the images get stored on the DAS but once one night of observations are concluded, the data gets compresses and moved to a RAID system and gets copied to a tape to be transported.

Once the data is obtained, it needs to be reduced, because it helps in accelerating computations and also cost of storage and further transmission.

1.1.2 Data Reduction

A robust and largely automatic reduction pipeline [1, 2] has been developed to reduce the data from SuperWASP Cameras.

- **Point Spread Function Profiles:** As the images acquired are slightly defocused and the PSF has a FWHM of a few pixels; the spatially independent PSF model is subtracted hoping that this sacrifice of statistical accuracy will result in reduced systematic errors that could arise from under sampling of stellar images.
- **Frame Classification and Calibration:** In the pipeline, an automated frame classifier was developed to identify the nature of the frames which automatically classifies frames by statistical measurements. The frames are then classified as bias, flat, dark, image or unknown.The pipeline carries out a number of statistical validity tests on each type of calibration frame, rejecting suspect frames before constructing master bias, dark, and flat-field frames. These calibration frames are then used to increase the signal to noise ratio of the images and reduce thermal noise, read noise, vignetting etc. [3]
- **Astrometry:** Astrometry is the study of precise movements and positions of various celestial objects. And hence it is important to obtain an astrometric solution corresponding each image. The celestial coordinates of the image center are obtained from the equatorial mount coordinates recorded in the data headers, and from the known offset positions of individual eight cameras. Subsets of the Tycho-2 [4] and USNO-B1.0 [5] catalogs are made for every camera according to the pre-programmed direction of the mount. The recognized star patterns formed by the brightest 100 stars in both catalogs are used to establish a preliminary plate solution. Now the pipeline software creates a photometric input catalog from the list of all USNO-B1.0 objects brighter than magnitude R $>>$15 (in the USNO system), whose positions fall within the boundaries of the CCD image. This output catalog is now ready for the photometry stage.

- **Aperture Photometry:** Photometry is used to measure the flux or intensity of light that is given out by celestial objects. Aperture photometry is done on all the positions of all the stars in the images from the catalog which is prepared to obtain these flux values. This has two major rewards for consequent data analysis and recovery: every object gets associated with its photometric measurement from the beginning, and the aperture for every object is always centered at a precise position on the CCD. The resulting photometry still contains time- and position-dependent trends, which is removed by post-photometry calibration.

- **Post-Pipeline Calibration—PPWASP:** Post-photometry calibration of the data is achieved through the use of code that constructs a theoretical model and is subtracted from the data leaving residual light curves. The input catalog contains extra data on the sky background level such as aperture radius, raw instrumental aperture fluxes etc. These results are the output in FITS binary tables. So, to reduce the raw instrumental magnitudes to calibrated standard magnitudes and remove the four main trends in the photometry: the effects of primary and secondary extinction, the instrumental color response, and the system zero point, the FITS binary tables are inputted into the PPWASP. Once post-processing is complete, the calibrated fluxes are added to the binary FITS tables ready to be ingested into the archive.

1.1.3 Archive

The archive server is built according to the internal architecture of a storage cluster i.e., Beowulf-style cluster. [2] It comprises three major classes of data:

- **The Raw Images:** The raw images are stored in the form of FITS files on the archive folder and are then indexed using RDBMS. The image files as processed by the pipeline are stored in their original form and are located within a directory on a hierarchical storage management system. An entry into the RDBMS allows images matching a user query to be found easily and efficiently.

- **The Bulk Processed Photometry:** The photometric data is also stored in form of FITS files on the archive server. However, its processing is more complicated. This is because the end user making the request requires only a single file containing all the photometric data of a single celestial object collated from a large number of images taken over a large time duration. But the actual output of the pipeline would be a collection of files which would each contain the photometry of all celestial objects in a given image and end user would end up receiving an overwhelming amount of data. For this reason, unique identifier is made, which makes it easy to build up a complete light curve from various categories of information. And this process gets immensely assisted by the catalog-driven nature of the WASP pipeline.

- **An Extensible Catalog:** It provides data for all objects observed by SuperWASP. It includes the equatorial coordinates and their relation to the USNO-B1.0 catalog. It is an extensible catalog that can be amplified with the data of further analyses of the light curves. Access to all this data is easily available to users via simple command-line tools.

1.2 Light Curves

Light curves are graphical representation of the variation in apparent brightness (magnitudes) of celestial objects with respect to time, usually in Julian Date. [6] They can be obtained after photometric analysis of images which gives us the values of fluxes of each celestial object in an image. The light curve is the single most important graph in variable star astronomy. It is used to analyze short -term or long-term changes in the magnitude of a celestial object. And this analysis can reveal a lot of secrets about the stars, like information about their periodic behavior, the orbital period of eclipsing binaries, or the regularity (or lack) of stellar eruptions. An even more in-depth analysis of them can help us calculate the mass and sizes of stars. Moreover, a large amount of data spanning over several years can tell one about the periodicity of a star which in turn can reveal information about its structure.

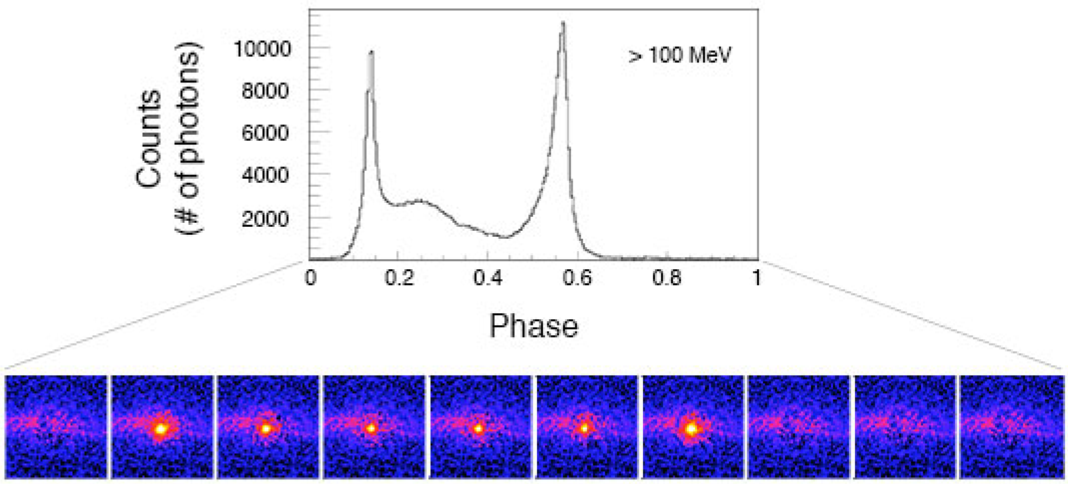


Figure 1.1: Time-lapse images of the Vela pulsar (bottom) and how they translate into a light curve (top) [7]

Now let us say we have 20 images consisting a Star X over a period of 4 months, clicked from the SuperWASP cameras. As explained previously, using aperture photometry we can obtain the flux values for this star from each image. Now once the apparent magnitude of Star X is measured from every object, these measurements are strung to make a light curve. A typical SuperWASP light curve may contain over twenty thousand measurements and the SuperWASP archive contains light curves of more than 30 million stars and over 11TB of data.

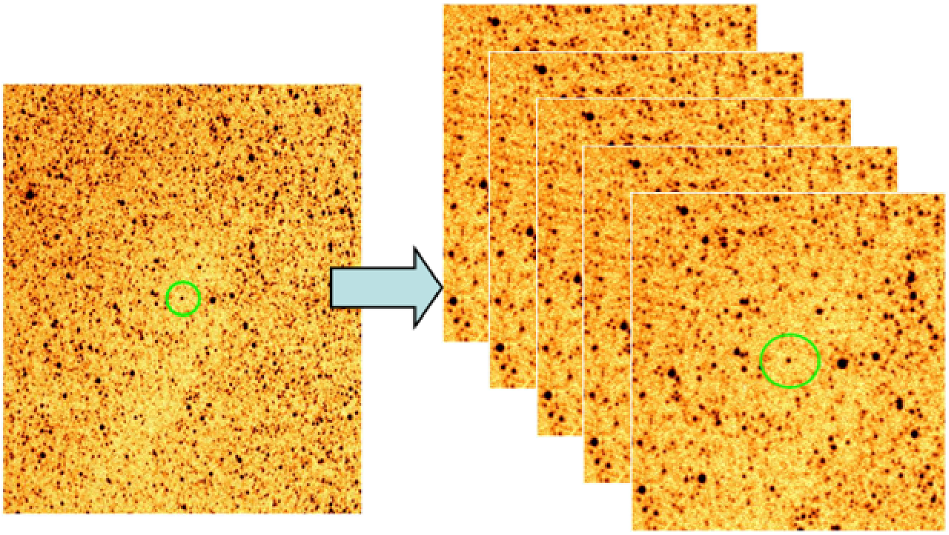


Figure 1.2: Wide-Angle Images obtained from SuperWASP cameras for one observation (left) and multiple observations (right) which undergo the SuperWASP data reduction pipeline to form Light Curves. [8]

Now let's say the values we obtained for the apparent magnitude of Star X are as follows:

Table 1.1: Apparent Magnitude Values for Star X [9]

Date	Magnitude	Date	Magnitude
April 21	9.2	June 20	8.7
April 27	9.3	June 26	8.3
May 3	9.7	July 2	8.6
May 9	9.9	July 8	9.1
May 15	9.6	July 14	9.1
May 21	9.8	July 20	9.2
May 27	9.9	July 26	9.5
June 2	9.7	Aug 1	9.9
June 8	9.1	Aug 7	9.7
June 14	8.8	Aug 13	9.7

Then a light curve for it can be plotted as:

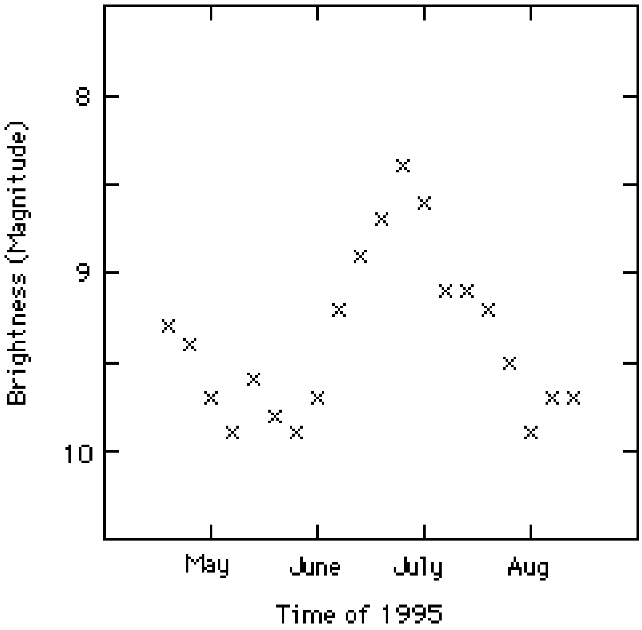


Figure 1.3: Simple plot of the Apparent Magnitudes of Star X with respect to time. [9]

The plot shows the brightness of a celestial object viewed through a telescope every 6 days over the course of a few months. Thus, we have obtained a light curve for the measured object.

Now once such a light curve is obtained, we can use it to compare with the standard light curves to possibly identify the type of object we're observing or even discover new objects.

As we can already deduce by now that light curves are basically a time series data of celestial objects. Astronomers can now use a number of methods to study these time series data which allows them to determine periodic components of data, which can further be related to the physical properties of the celestial objects, like rotation, binary period etc. This study of light curves is also known as timing analysis.

Time-series analysis is an extremely filed of mathematical and statistical analysis. Such analysis of time-series measurements helps in the physical understanding of the behavior of a time varying system by quantifying the variation itself. It helps in answering several important questions like: what makes the system time variable? What makes the system similar to or different from other systems? Is the system predictable? And can some limits be placed on the behavior of the system? There are several different techniques of time-series analysis that can be applied to variable star data.

1.3 Methods of Timing Analysis

1.3.1 Epoch Folding

We know, anything that repeats itself over regular intervals is called a periodic or a cyclic behavior. Hence when we want to obtain information about what is happening at a particular moment, we would not need to know which cycle we are in, as each cycle is identical. We would instead be interested in knowing, what part of the cycle we are observing. Therefore, say, we are observing a perfectly periodic star, then its variation would depend only on where the stare lies in its cycle, a quantity called phase. [10]

To illustrate this with an example, let's assume we have a 24-hour clock. Now we know a clock depicts a periodic behavior of period of 1 day (24 hours). However, to know the time at any moment, we need not know what day it is (i.e., which cycle it is), we are only required to know the portion of the day (i.e., how far we are into the cycle).

Now for a clock, we measure "how far we are into the cycle" in units of hours or minutes. However, a cycle simply starts at 0 and ends at 1. Hence for this situation, the phase for the clock is the fraction of the whole cycle that was completed till now. But for this, it is extremely important to know the period of the cycle. For a clock the period is 24 hours, therefore at 12 noon, its phase will be 12/24 i.e., 0.5.

Another important point to notice here is that the phase we have calculated depends on the initial conditions for the clock. We assumed that the day starts at 00:00hrs and hence calculated 12 noon to be phase 0.5. however, if we choose our day to begin at 06:00 hours, then 12 noon will be only 6 hours into its cycle, therefore its phase will be 6/24 i.e., 0.25.

Therefore, to compute a phase, we also need to know the starting time of a cycle, also known as the initial phase of the cycle or *epoch*. For the clock, the epoch is usually midnight.

Then to find out the phase of a cycle we can use the formula,

$\phi = \frac{t-t_0}{P}$

Where ϕ is the phase we need to fine at time t, for a cycle having period P and epoch is t_0.

But, let's say we need to find the phase for a moment 36 hours since the epoch, therefore the phase will be ϕ = 36-0/24= 1.5

But the cycle always begins at 0 and ends at 1, therefore something is not right. We know now that phase is "how far we are into the cycle," and that we are no longer concerned with which cycle, so every day the phase should be same. But as we can see $\phi = 1.5$ is not same to the previous day i.e., noon when the phase was $\phi = 0.5$.

But if we look carefully, we will realize that it is in fact the same phase. $\phi = 1.5$ can be called "one-and-a-half cycles," or we can say it is "halfway through the next cycle." And as already mentioned, we are not concerned with which cycle, we can simply omit the "next cycle", and say "halfway". And hence we can see that the phase is in fact 0.5.

This concept can be extended further and therefore each time one has to compute phase, it can be converted into "standard phase" by omitting the "which cycle" part. For example, if we have $\phi = 7.28$, then we can see that we are 28% into the 7th cycle. And so, we can omit the "seven cycles" part, and say the phase is 0.28. All standard phases are between 0 and 1.

The concept explained above may sound strange as it is stated that 0.28,1.28,3.28 etc. are all essentially the same phase. However, it is mathematically correct. We are basically taking all numbers modulo 1. This is known as modular arithmetic, where we can still do arithmetic and compute numbers, but always have to omit the integer part. Therefore, we can write,

$\phi =$ decimal part of $\left[\frac{t-t_0}{P}\right]$

Another thing to keep in mind are the negative phases. For example, if we want to know 66 hours before epoch, then in this case, time is t = -66, therefore phase is:

$\phi = \frac{-66-0}{24} =$ -2.75

Now again, we simply omit the "2" and the standard phase becomes 0.75. This is because as mentioned, we are doing arithmetic modulo 1, which gives us permission to add or subtract any integer without changing the result. Therefore, we can add three to the given phase: $\phi = -2.75 + 3.00 = 0.25$. Now, 0.25 lies between range 0 to 1, and hence is the standard phase.

Now that we are clear about phases, periods and epoch, we can move on to epoch folding and the folded light curves.

Folded Light curves

A folded light curve is basically a light curve whose magnitude instead of being plot as a function of time, is plotted as a function of phase. [10] That is, we can perform the similar experiment with a variable star as we did for a clock. Again, we will find the period for its variation, choose an epoch and calculate the phase for all its data points.

So mathematically, what is happening,is the cycles are getting superimposed on each other. That is, instead of plotting each data with time, we plot its phase and that way all cycles get "folded" on top of each other, and with enough data we can get an accurate picture of what the cycle looks like. If, however we have picked the wrong period, the folded light curve will look flat and boring as all the maxima and minima in each "phase bin" would get cancelled out with each other. Each light curve in the SuperWASP archive is processed using a computer program to search for possible periodic variations in brightness which can also be at wrong period and hence a manual analysis is required. Now once we have obtained a light curve with sharp maxima and minima, an astronomer must assess how significant the resulting light curve is. This is generally done by looking at the spread of values, or errors, in the typical bin, and comparing how much higher the high bins are than the standard error.

However, in astronomy a couple of things to keep in mind is that every time a folded light curve is plotted, rather than plotting it from 0 to 1, we plot it from -1 to 1 or 0 to 2, so we can see a complete cycle of variation containing the complete picture of both the maxima and minima.

To show this via an illustration, below is the process of epoch folding for an X-ray binary star system.

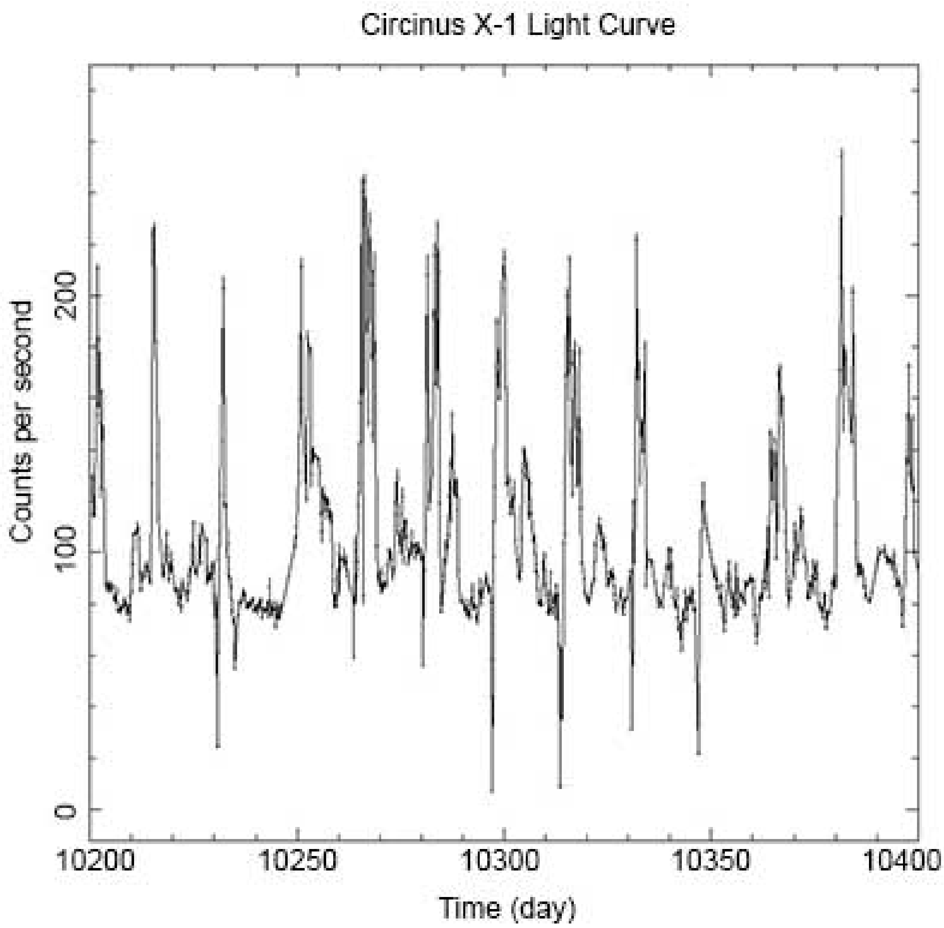


Figure 1.4: Light Curve of X-ray binary system called Circinus X-1 using data from the All-Sky Monitor aboard the RXTE satellite. [11]

First, we have the original light curve. We can clearly see the maximum intensity repeated at regular intervals.

Looking closer at the data, we can visually identify the period as 16.6 days. So to construct its folded light curve we plot these point with respect to their phase having period 16.6. Pictorially, it can be seen as cutting the light curve at every 16.6 gap as shown in fig 1.5.

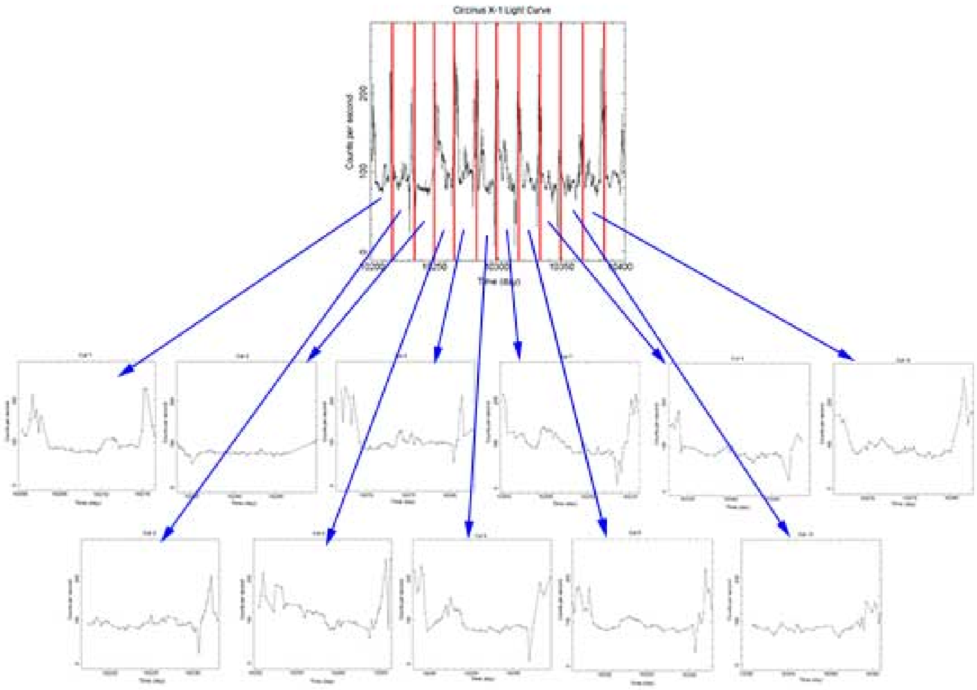


Figure 1.5: Light curve of X-ray binary system Circinus X-1 cut into 16.6 day segments. [11]

Now, in order to perform epoch folding, we "fold" or "add" all the given light curves and stitch it together to obtain a solo light curve having period of 16.6 days. Where all the intensities of the light curved gets added to give the final intensity in each time bin.

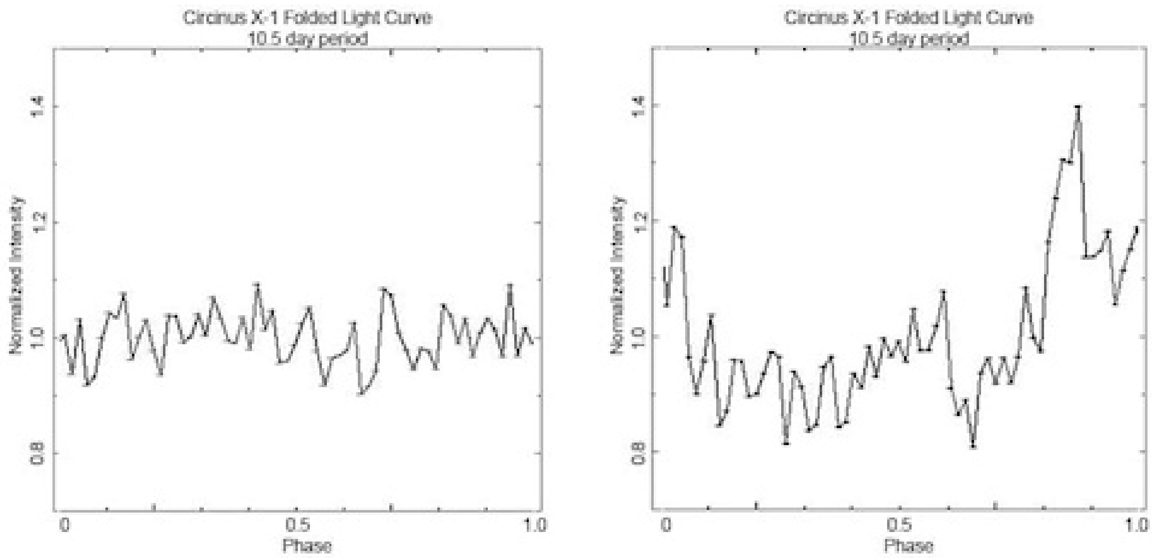


Figure 1.6: Two folded light curves for Circinus X-1. On the right, the folding period used was 16.6 days, which is close to the real period of the system. On the left, the folding period used was 10.5 days, quite different from the real period of the system. [11]

We can see from the above figures that light curve having 10.5 day period is pretty noisy and does not indicate strong features of minima and maxima. Hence we can conclude that the 10.5 day period is not the real period of the system.

Each light curve contains a red line over-plotted on each image. This is actually the average of the folded light curve. In each of 100 small ranges of the cycle (phase bins), all data points are averaged to produce a single value. This average folded light curve therefore indicates the average variation of brightness of the star throughout the many years of observations. [12]

1.3.2 Fourier Analysis and the Fourier transform

Fourier analysis is a mathematical technique which uses an infinite number of sine and cosine functions with different periods, amplitudes and phases to represent a given set of numerical data or analytic function. This helps astronomers to estimate the period(s) of variability by determining which of the sine-cosine functions are statistically significant. This also means that fourier analysis works best when the signal under study is quite identical to a sine or cosine wave. A Fourier transform is a mathematical operation that is used to change the data from time domain to frequency domain. The aforementioned amplitudes and phase are determined using this Fourier transform itself. If we have a give set of time-varying data x(t), the its fourier transform is given by the integral:

$$F(\nu) = \int_{-\infty}^{+\infty} x(t) exp(-i2\pi\nu t) dt \tag{1.1}$$

Where v is frequency, defined as 1/P and sine and cosine functions are defined by the Euler's formula,

$$exp(-i2\pi\nu t) = cos(-2\pi\nu t) + isin(-2\pi\nu t) \tag{1.2}$$

As shown in the figure 1.7, if the data set contains several signals with different frequencies, then its fourier transform will have local maxima at each frequency, with a global maximum at the frequency having largest amplitude.

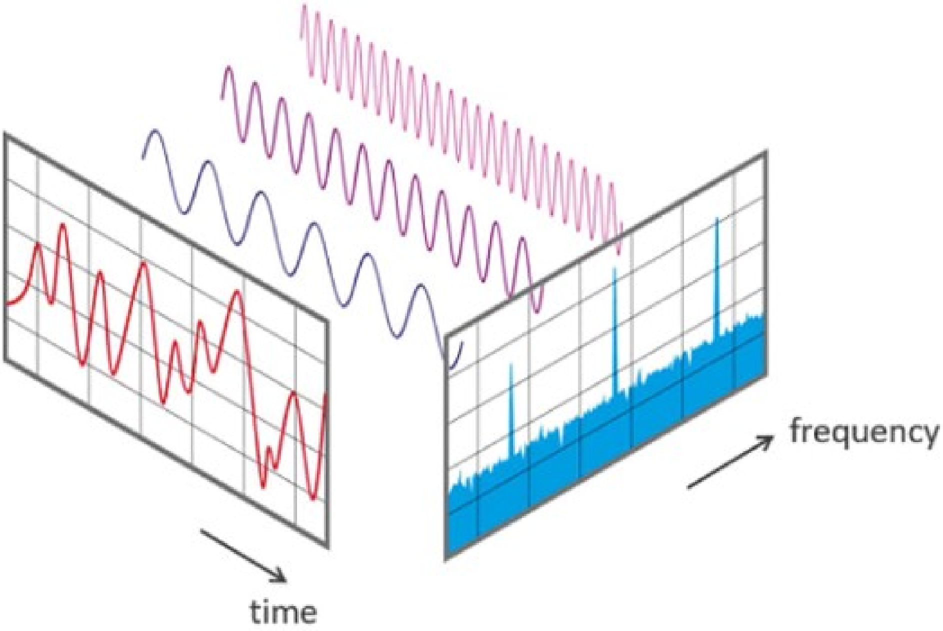


Figure 1.7: Pictorial depiction of Fourier Transform: The red wave represents the incoming signal;the purple waves show how the Fourier transform performs; the blue wave represents the transformed data output. [13]

Despite of being such a powerful tool, Fourier transform also has its limitations [14] which depend on the data set available.

The amount and quality of data places several limits on the usefulness of the transform, including maximum and minimum period testable, the accuracy of the period determination, and the minimum statistically significant amplitude that can be found etc.

This can be explained using an example,

Let us say we have,

Span of data= 5000 days

Sampling rate= 10 data points per day

Therefore, maximum period that can be detected is 5000 days having one cycle, however this is highly unreliable as we cannot be sure if the detected variation is actually periodic or simply a short-lived instability. Therefore, a more trustworthy period could be of 2500 days or 1250 days having 2 and 4 cycles respectively. Therefore, the span of the data determines the accuracy of period detection.

Another important factor responsible for the effectiveness of a Fourier transform, is its resolution. The resolution of a Fourier transform is defined as the precision up to which the frequency (or period) of the given data set can be determined. This resolution is defined by

$$d\nu = \frac{1}{N\Delta} \qquad or \qquad dP = \frac{P^2}{N\Delta} \tag{1.3}$$

Where N is the total number of samples i.e., 50,000 and Δ is the space between the samples i.e., 0.1 day. As it can be clearly seen in the figure 1.8, a longer span clearly provides a much more precise determination of the period.

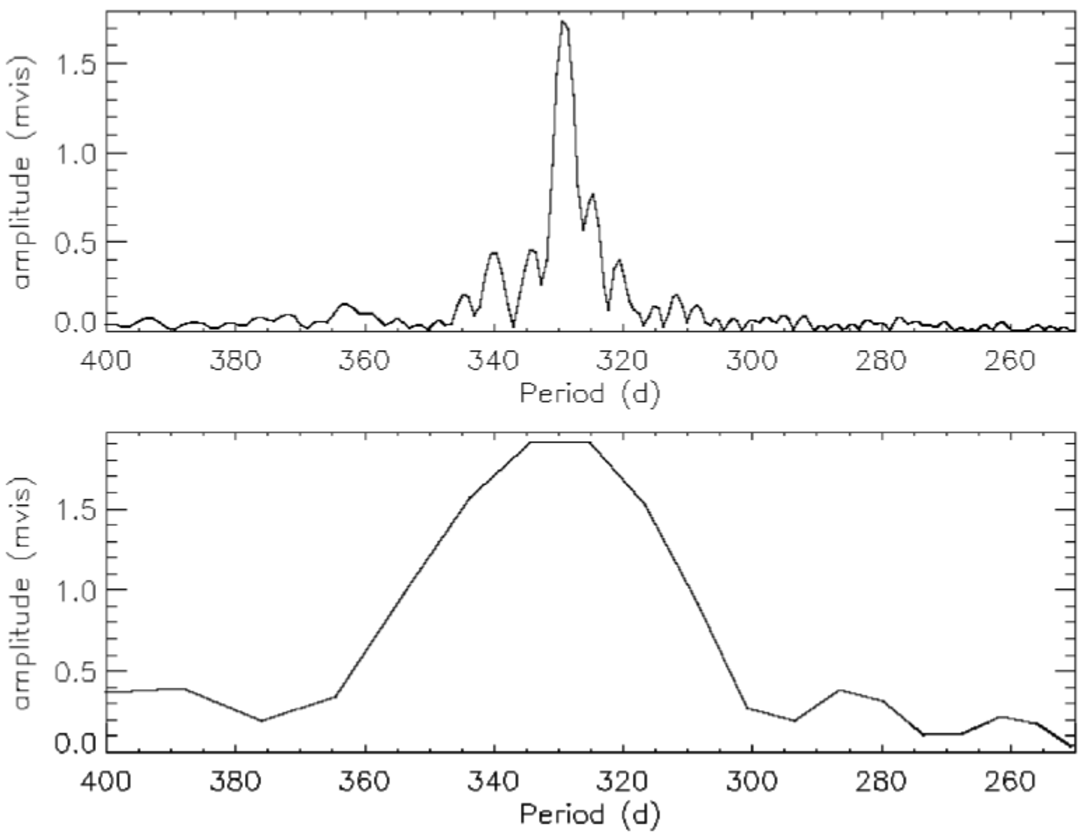


Figure 1.8: Fourier Transform of two data sets of R CVn star. We can see that both data sets show the spectral peak at nearly same period, however thw pulsation period is much better resolved in above picture (span of 32674 days) as compared to case below (span of 2997 days). [14]

Now, minimum period that can be detected would be 0.2 day, corresponding to max frequency or Nyquist frequency of 5 cycles per day. Sometimes however, it is possible to detect frequencies higher than the sampling rate itself. When this happens, the transform undergoes aliasing, in which several peaks appear in the transform along with the real one. These alias peaks are separated from the true frequency by the integral multiples of the sampling frequency, such that the transform looks like a “picket fence”.

In case of even sampling, the alias peaks will have equal statistical significance to the real peak and therefore it is impossible to tell which peak in the spectrum is the correct one.

In case of uneven sampling, the alias peaks have different strengths and generally would be lower than the correct one. As one can see from figure 1.9 of δ Scuti star that due to its strictly period variation, uneven sampling and complete phase coverage, it was easy to determine its period. However, in general it is not possible to uniquely determine periods shorter than the sampling rate.

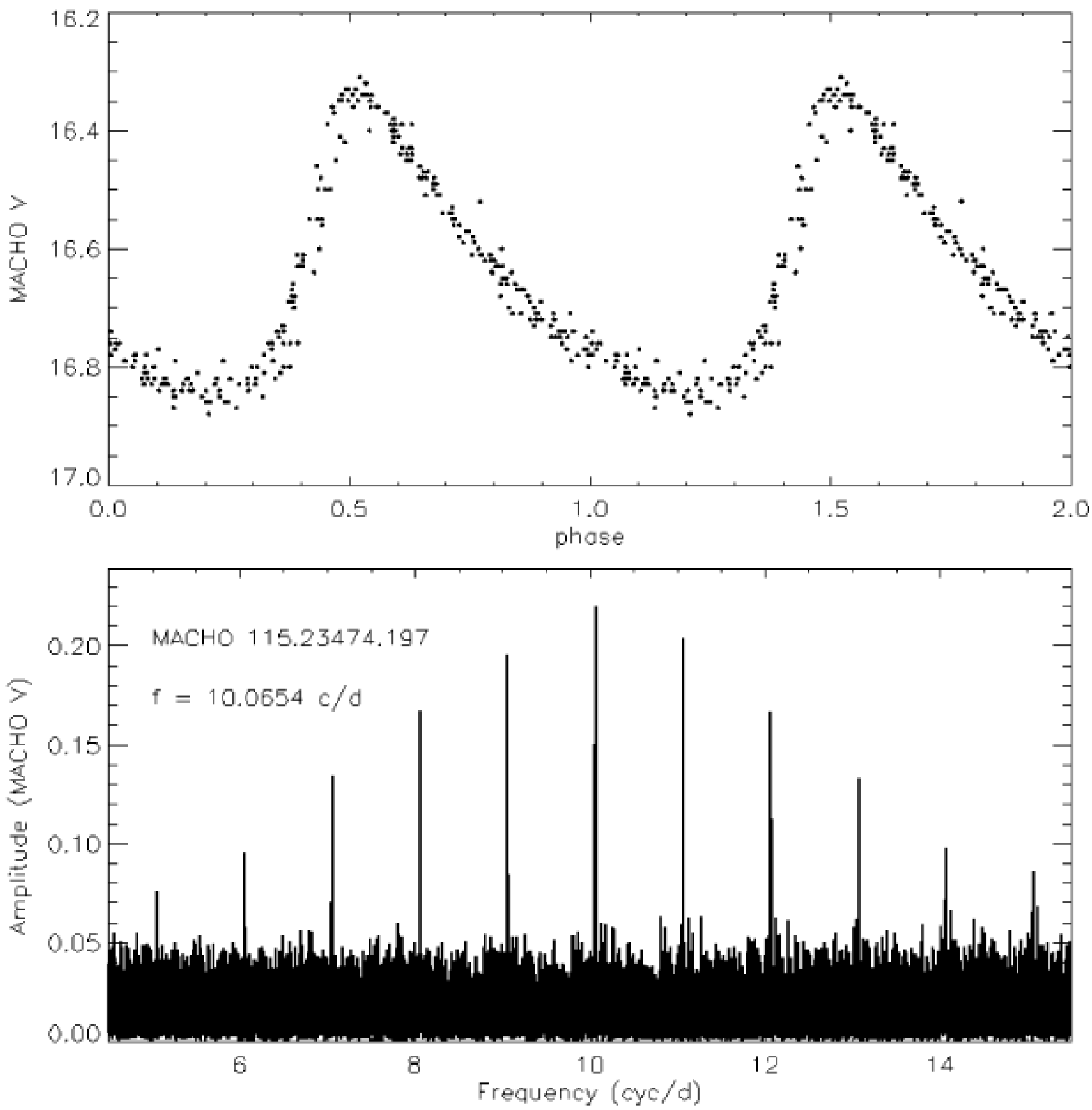


Figure 1.9: Folded Light Curve and Fourier Transform of high amplitude δ Scuti Star having pulsation frequency of 10.06 days with sampling rate of 1 day. As the sampling is so low, the transform suffers from aliasing shown by the secondary peaks offset from the main peak by integer multiples of 1 cycle per day. [14]

Another major consideration in Fourier analysis is noise. This noise can be of two types-Intrinsic noise of photometric observations caused by sky background and instrumental effects and measurement error of the data itself. Measurement of noise is extremely important in Fourier Analysis as, Fourier analysis for a given data set assumes that the whole data set is a signal. And thus noise will appear in the Fourier spectrum. And therefore once we find about about the noise level, it allows one to regulate the reliability and accuracy of the results.

For example, figure 1.9 contains light curve from the photometric data provided by MACHO, which typically has errors between 0.05 magnitude and 0.2 magnitude. It depends upon the field crowding and brightness levels of the stars. Now as we can see from figure 1.9, the errors get manifested as a dense forest of peaks within 0.05 limiting magnitude. "The mean value of these peaks provides some indication of the detection limit of any periodic signal. If the signal strength is large relative to the scatter in observations (for example , Mira Variables), then the noise level of the Fourier Transform will be relatively low. If signal strength is small relative scatter observation (for example, Cepheids or δ Scutis), then the noise level will be higher. In case of data obtained from MACHO, any signal amplitude near or less than 0.05 magnitude would be indistinguishable and undetectable." [14]

Thus, it is important to keep in mind the noise level and errors in the photometric data when performing Fourier analysis.

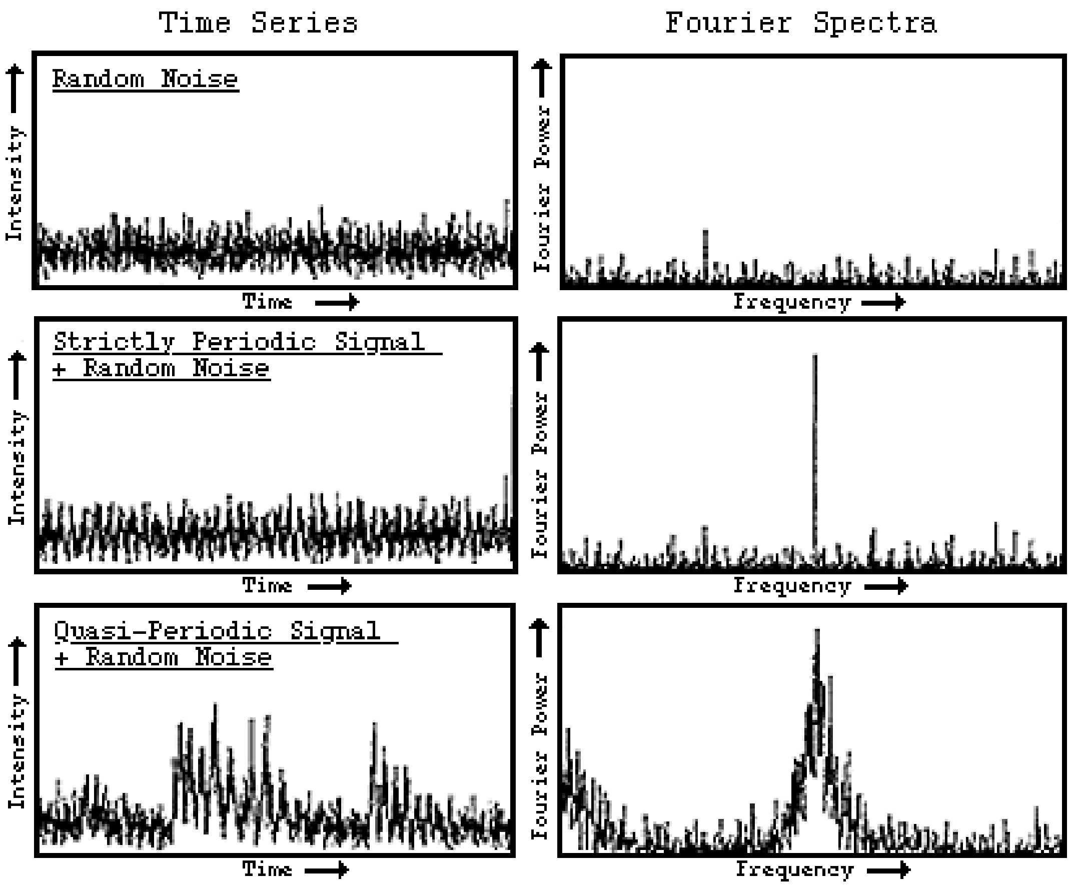


Figure 1.10: Time Series Signals with their corresponding Fourier Spectra. [14]

1.3.3 Wavelet Analysis

While Fourier transform and epoch folding are great at identifying systems like pulsars having consistent periodicity, but they tell us only about the frequency present in the whole data set and nothing about the time dependency of these frequencies. But as we know that that there are many sources of interest having time varying periodic signals like, long-period Mira stars exhibit slightly varying periods from cycle to cycle, while R Aquilae stars are known to exhibit strong varying periods which indicate the evolutionary changes of the star. Other stars, for example Semi-Regular variables and RV Tauri stars, do not show a constant period but instead vary with one or more characteristic periods which become incoherent when viewed over the full light curve. [14] Some stars also exhibit temporary periods or quasi periods. Therefore, due to the need to analyze transient and/or non-sinusoidal signals having multiple frequencies, wavelet analysis and wavelet transform was developed.

The wavelet transforms of a set of time-series data, x(t), is given by, [15]

$$W(\omega, \pi; x(t)) = \omega^{\frac{1}{2}} \int x(t) f * (\omega, (t - \pi)) dt \tag{1.4}$$

where ω is a test frequency, π is "lag" time or a position within the light curve, and the function, f, is called the "mother wavelet"-a function which determines how the signal should vary with time, frequency and position within the light curve. The wavelet transform is a highly flexible tool as the mother wavelet can be of the form of any mathematical function (it can be specific waveform like sine wave or a time-varying weighting function like sliding window). And through this wavelet analysis one can find the given frequencies in the signal as well as the evolution of their spectrum with respect to time. It basically decomposes a signal into time and frequency space simultaneously.

However, although wavelet analysis is a powerful tool, it also has limitations and astronomers must use them with caution. This is because the extra information that can be gained comes at the cost of increased difficulty in evaluating whether the result is significant or not. For example, similar to Fourier analysis, the data set for wavelet analysis should also be long and well-sampled enough such that one can adequately study the periods of interest. That is, if the data spans 2000 days and we have the period of interest as 500 days, then the wavelet window covers only five cycles and the wavelet analysis will not give meaningful information about the time evolution of the signal as nearly all of the data will lie within the window for any chosen value of π.

1.4 Variable Stars [6, 16–19]

A star is basically a massive ball of gas. It exists by creating a delicate balance between nuclear force at its core, trying to blow it apart and the force of gravity trying to compress all its mass into the center.

Throughout the stellar evolution, these forces advance and recede, swelling and shrinking the star. And finally, when a star dies, it is because one of these forces overpowers the other and the star either collapses into a black hole, or explodes supernova.

Every star that we see in sky either has been or will be variable in its light output at one time or another. It is inevitable. If one could live long enough, we could see that every star is a variable star.

And knowing about the variability of a star, one can know how far away, how old, how massive, or small the star is. This is the reason why variable stars are so important in astronomy.

Now what is variability? Simple stated, a variable star is star who apparent magnitude, that is, its brightness fluctuates or variates as seen from Earth. Now this variability can be of two types depending on the cause of its existence:

Extrinsic variables:

These are stars whose variability is caused by external factors like rotation or eclipses, i.e., the amount of light that reaches Earth. This means, that the total energy output of the star or its luminosity is not varying, but the amount of light we see from our vantage point on Earth varies.

The primary types of extrinsic variables are:
(1) Eclipsing variables
(2) Rotating variables
(3) Micro-lensing variables

Intrinsic variables:

These are stars whose variability is caused by changes in the physical properties of the stars themselves.

The primary types of intrinsic variability are:
(1) Pulsating variables
(2) Eruptive variables
(3) Cataclysmic variables
(4) X-Ray variables

With a total of 211 variable star types and sub-types listed in the International Variable Star Index5 [20], it is difficult to categorize them solely on the basis of their light curves, specially with the noise level of the SuperWASP light curves.

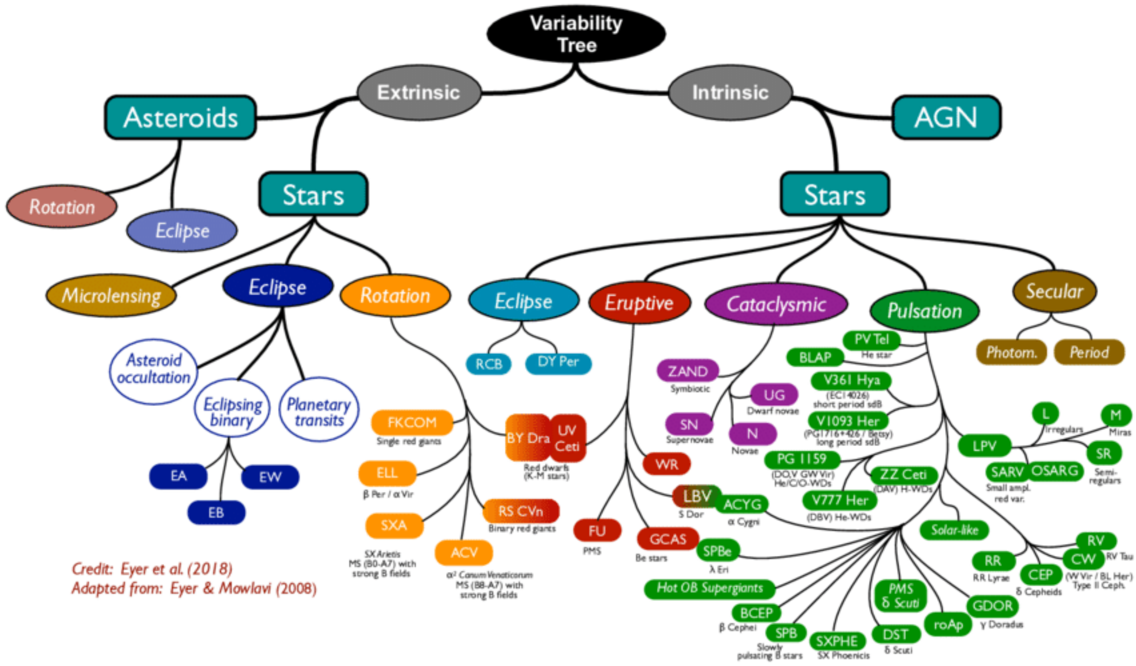


Figure 1.11: Variability Tree [21]

Therefore, for the purposes of this project, the light curves are categorized into the generic and overachieving variable types: [18]

(1) Pulsators
(2) EA and Eb
(3) EW
(4) Rotators
(5) Unknown
(6) Junk

1.4.1 Pulsating stars

Pulsating stars are those type of variable stars whose brightness variations are caused by changes in the area and temperature of the star's surface layers. These pulsations can be of two types. First is known as the radial pulsation. In such pulsations the whole volume of the star changes periodically. This happens due to the spherical and symmetrical expansion and contraction of the outer layers of a star around a hydro-static equilibrium state, that is, the gravitational pull on the mass elements of normal stars is balanced by gas pressure. There are four proposed mechanisms [22] for pulsations-

- **The ϵ mechanism [23]**

  According to this mechanism if a nuclear burning region is compressed, its pressure increases and as a result temperature also increases. Increased temperature increases the rate of nuclear reaction producing more energy and causing expansion. But once expansion happens, the pressure drops, the temperature drops and as a result the rate of nuclear reactions and energy produced drops causing compression and this cycle keeps repeating causing steady pulsations.

- **The κ - γ mechanism [24]**

  These are non-adiabatic pulsations which occurs in stars having stellar interiors consisting of partial ionization of Hydrogen and Helium and sometimes other elements as well. When there is an increase in the compression of the atmosphere of the star, it causes an increase in temperature and density. Now in layers where this increase in temperature causes the radiative opacity κ to increase and/or decrease of the third adiabatic exponent Γ_3, the energy flux coming from the inner layers of the star gets stored temporarily. Now this stored heat energy causes an extra buildup of pressure which pushes back the layers which are trying to reach their equilibrium state by expanding. This causes the star to expand beyond its equilibrium radius. When the material recedes, energy is again stored in the stellar interior, and the whole cycle repeats as the layers repeatedly move inward and are pushed back outward. This mechanism is also called the Eddington Valve. The δ Cepheid, RR Lyrae and the δ Scuti stars draw their pulsation power from the Heii ionization zone. roAp stars are believed to be excited in the Hi and Hei ionization zones Mira variables are excited in the Hi ionization zone. β Cephei and SPB stars are triggered in the ionization zone of the iron-group elements.

- **Convective Blocking/ Convective Driving [25]**

  In this scheme, similar to κ Mechanism, the base of convection zone blocks the energy flux from the interior temporarily, releasing the energy stored during compression in the subsequent expansion phase. The pulsations of white dwarf stars (spectral types-DA, DB) and γ Doradus stars are thought to be excited (at least partly) via this mechanism.

- Finally, for stars that are intrinsically stable (not self-excited) like the sun pulsate due to disturbances in their surface convective zones. The vigorous convective motion in the outer surface layers generates acoustic noise in broad frequency range, which excites solar-like oscillation modes.

Second type of pulsations are non-radial in nature. These non-radial pulsations produce small amplitude variations. These are such pulsations in which the star changes its shape, but not its volume, that is, some parts of the stellar surface protrude outwards while some shrink inwards simultaneously. Depending up on their type and excitation mechanism these pulsations traverse various internal portions of the star which cause these periodic displacements in a characteristic pattern of brightness variations which can be theoretically described with a spherical harmonic function. The areas which pulsate in the same phase and other which pulsate in opposite phase are separated by horizontal and vertical node lines. Node lines are lines along which there is no pulsation. These pulsations can be described with three quantum numbers (l, m, and n) where l is the degree of the mode that is equivalent to the number of node lines on the stellar surface, and the azimuthal number m which tells the number of node lines that lie in meridional direction (there are 2l +1 possible m-values for one l-value) and radial number "n" which tells about the number of nodes in the radial component of displacement from the center to the surface of a star. For n=0, the star oscillates in the fundamental mode, n=1 is the first overtone and n=2 the second overtone. Study of the pulsations and their hidden frequencies is currently the only means to probe and study about the internal structure of stars and obtain information about their local physical conditions like pressure density and temperature. The study of all this is known as asteroseismology. There are some other stars as well, like, β Cepheid and δ Scuti stars which pulsate in both radial and non-radial modes.

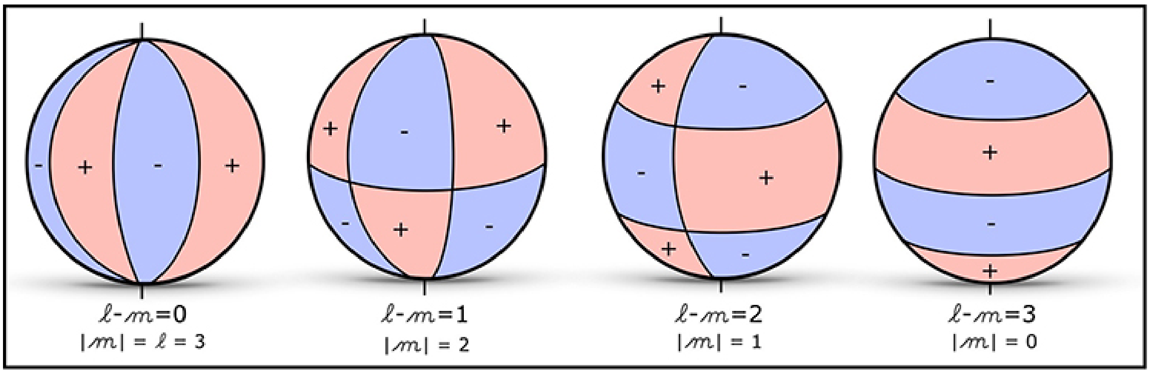


Figure 1.12: Non-Radial Modes of Pulsation. [26]

Table 1.2: Pulsating variable stars arranged in alphabetical order by designation [17]

Variable		Designation (and subclasses)	
53 Persei	*	**53 Per**	O9-85 non-radial pulsating stars
α Cygni		**ACYG**	Be-Ae (emission) pulsating supergiants
β Cep		**BCEP**	classical β Cephei stars
		BCEPS	short-period β Cephei stars
BL Boo	*	**BLBOO**	anomalous Cepheids
Cepheids		**CEP**	radially pulsating F lb-II stars
		CEP(B)	double mode pulsators
W Virgins		**CW**	
			CWA population II, period $>8^d$
			CWB population I, period $<8^d$
δ Cepheids		**DCEP**	classical Cepheids, population I
		DCEP(S)	classical Cepheids, with overtone
δ Scuti		**DSCT**	A0-F5III/V pulsating stars
		DSCTC	low-amplitude δ Scuti stars
γ Doradus		**GDOR**	early-type F dwarfs showing multiple periods
Slow irregular variables		**L**	
			LB late-type giants
			LC late-type supergiants
λ Bootis	*	**LBOO**	"p", non-magnetic, A-F, population I dwarfs
Mira		**M**	long-period late-type giants
Maia			stars predicted to exist but none have been found
mid B variables	*	**Mid B**	B3-B6 stars; periods = 1-3 days; amplitudes of up to 0.12
PV Telescopii		**PVTEL**	helium supergiant Bp stars
RPHS		**RPHS**	Very rapidly pulsating hot subdwarf B stars (EC 14026 stars)
RR Lyrae		**RR**	
			RR(B) double-mode RR Lyrae stars
			RRAB RR Lyrae stars with asymmetric light curves
			RRC RR Lyrae stars with symmetric light curves
RV Tauri		**RV**	
			RVa radially pulsating supergiants with constant mean magnitude
			RVb radially pulsating supergiants with variable mean magnitude
Semiregular variables		**SR**	
			SRA M, C, S, Me, Ce, Se giants/small amplitudes
			SRB M, C, S, Me, Ce, Se giants/poorly defined periods
			SRC M, C, S or Me, Ce, Se supergiants
			SRD F, G, and K giants/supergiants
			SRS semiregular pulsating red giants with short periods.
SX Phoenicis		**SXPHE**	population II pulsating subdwarfs
UU Herculis	*	**UUHer**	high-latitude F supergiants
ZZ Ceti stars		**ZZ**	
			ZZA hydrogen pulsating white dwarfs
			ZZB helium pulsating white dwarfs
			ZZO showing He II and C IV absorption lines.

* Indicates a designation found within literature but absent in GCVS

CEPHEID VARIABLES:

These are one of the most important and famous pulsating variables. They pulsate radially, varying with temperature as well as diameter and resulting in brightness changes. They have strictly regular periods and amplitude. Because of their strong direct period-luminosity relation, they have been used as the standard candles upon which the knowledge of galactic and extra-galactic distances is built on.

The Cepheid variables are mainly of two types:

1. **Classical Cepheids/Population 1 Cepheids/type 1 Cepheids/δ Cepheid Variables:** These are bright yellow, highly luminous (100,000x of sun), giants and super-giant variables. They have a direct relation with their period and spectral type, i.e., the longer the period, the later the spectral type. These are used to determine galactic distances that lie within the Local Group and have helped identify a lot of characteristic of our Milky Way Galaxy, for example, Sun's height above the galactic plane and the Galaxy's local spiral structure. These Cepheids can also oscillate in fundamental, 1^{st} or 2^{nd} overtones. *They usually have asymmetric light curves that have steep rise in maxima and a slow decline. They also show a bump in their light curves which progresses backwards in phase with increasing period. The longest period classical Cepheids have almost sinusoidal light curve.*

2. **Type 2 Cepheids/Population 2 Cepheids:** These are older stars which are metal-poor and have low masses (approx. 0.5 to 0.6 solar mass). These stars tend to reside away from the disk of the galaxy and are used to establish distances up to the galactic center, globular clusters and galaxies. Accurately measuring their period-luminosity relation is extremely important as once in 1950s this error was discovered and re-calibrated then the distance scale of the universe doubled! These are sub-divided into three types depending on their age.

   (a) **BL Herculis Stars:** These stars have shortest periods. *Their light curves also show Hertzsprung progression like classical Cepheids where the secondary bumps progress backward in phase, starting from the descending branch of shortest period light curves and ascending branch for longest period light curves.*

   (b) **W Virginis:** Stars having period up to 5-20 days belong to the W Virginis subclass. They show irregular periods and amplitudes from cycle to cycle. *The light curves of CW stars sometimes exhibit humps on the descending branch of the light curve, or sometimes a broad flat maximum. Their light curves increase in their symmetricity with their period. That is, the shortest period stars show a steep rise and slow fall which becomes almost symmetric for stars with 8-day periods.*

   (c) **RV Tauri Stars:** These stars are yellow super giants which are older and have low mass. Their defining character is their unusual light curves for being a radial pulsator. These are further divided into RVa and RVb. RVA do not vary in mean magnitude and RVB vary in their mean magnitude by up to 2 magnitudes in V, with periods of 600- 1500 days. *These stars show alternating shallow and deep minima and therefore can be confused with eclipsing binaries.*

3. **ANOMALOUS CEPHEIDS:** These type of cepheid variables have relatively more mass (1-2 solar mass) and are metal-poor. They lie in between the classical cepheids and type 2 cepheid in the period-luminosity diagram. Two theories that suggest the origin of these starts are that either they are intermediate age stars that have exceptionally low metallicity or they are coalesced old binary stars. *Their light curves are asymmetric with steep rise and slow decline. The fundamental-mode pulsators have a small bump just before the rise. Their light curves are very similar to that of classical Cepheids (periods >1 day) and RR Lyrae stars (<1 day), and differentiating between them is quite difficult without more information.*

A small proportion of Cepheid variable also fall under the category of **Double mode Cepheids** which pulsate in two or modes at the same time.

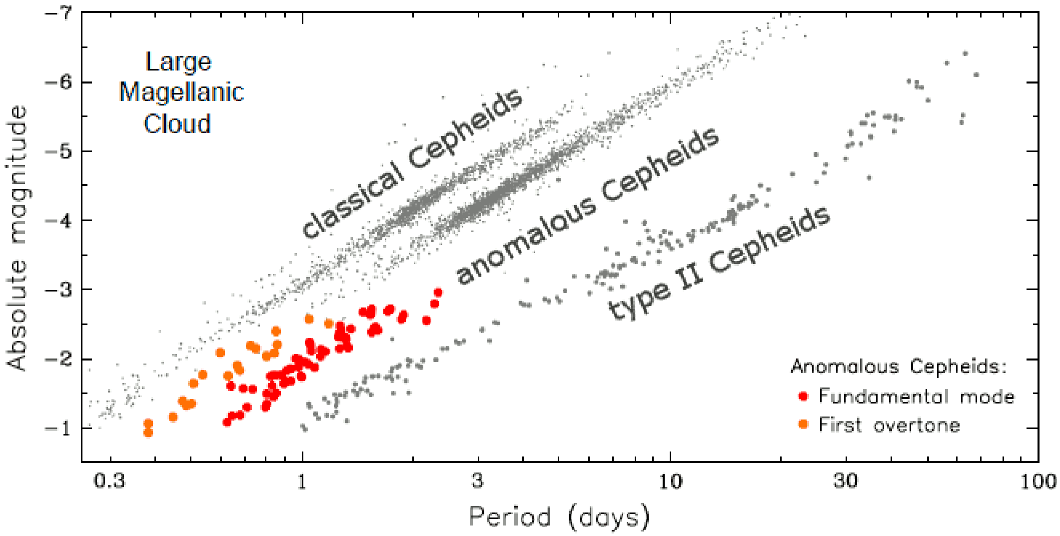


Figure 1.13: Classes of Cepheid Variables [19]

δ SCUTI VARIABLES:

δ Scuti stars have the same period-luminosity relation as classical Cepheids and are quite similar to them in a lot of other aspects. They are faint stars with short periods and small masses and pulsate in radial as well as non-radial modes.

δ stars having magnitude >0.3 in V are called **High-Amplitude δ Scuti stars or HADS**. These are fundamental radial pulsators having asymmetric, sawtooth light curves which also resemble that of classical cepheids.

Multi-mode pulsations are however very common in δ Scuti stars with FG Virginis having over 79 different modes. *Their light curves are however much more symmetric and hence difficult to differentiate with other types of variable stars.*

RR LYRAE:

These are low-mass, radially pulsating stars. These are stars old that have exhausted almost all the hydrogen in their cores and are now burning on helium and hence are excellent tracers of dynamical and chemical properties of oldest stars and helping us get insight into early history of stars. They are a numerous class of variable stars in some globular clusters and were also known as "cluster variable" earlier. They have a relatively small range of mean values which makes them great sources to calibrate the distances in both our galaxy and nearby ones. On the basis of their pulsation modes, they are subdivided into 3 types:

1. **RRab or RR0 (fundamental-mode):** They have periods from 0.3 to 1.2 days and amplitudes ranging from 0.5 to 2 magnitudes in V. *They have asymmetric light curves that has a sharp rise and slow decline. It also commonly contains bump, but its detailed morphology depends on star's chemical composition.*
2. **RRc or RR1 (first-overtone):** They have periods from 0.2 to 0.5 days having low amplitude variation.*They have almost symmetric light curves and sometimes can be nearly sinusoidal, which makes their identification difficult. They also sometimes show small secondary bumps on their ascending branch.*
3. **RRd or RR01 (double mode):** These stars simultaneously pulsate in both fundamental and first overtone. They have a period ratio of 0.74 days and fundamental periods of around 0.5 days. *After separation of their light curves into modes, the first overtone light curve has similar amplitude and shape as RRc stars while the fundamental overtone light curve has small amplitude and is nearly sinusoidal- different from RRab light curves.*

   RR lyrae stars also show a phenomenon known as Blazhko effect which means the stars show long-term modulation of the amplitudes and the phases of their light curves.

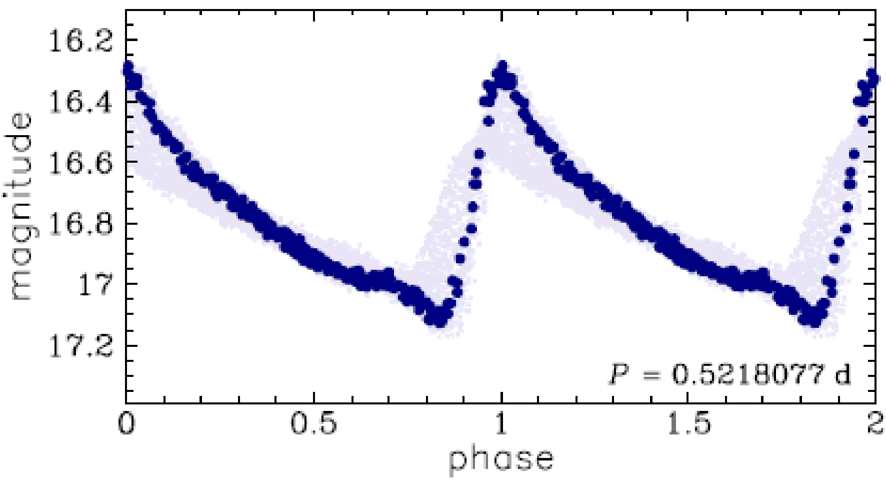


Figure 1.14: Illustration of one Blazhko cycle (155 days in this case). The spread in the curve can be see due to change in the phase and amplitude of the light curve [19]

1.4.2 Long Period Variables

1. **Mira Type Variables:** These are the coldest and largest red giant stars with high luminosity and low temperatures. They are in a much later stage in their evolution and masses approx. 0.6x solar mass. They also have huge radii around several hundred times the solar radii. Many of them are slowly ejecting a steady stream of matter into the surrounding space; this mass loss can have very dramatic consequences for their future evolution.

2. **Semi-regular stars (SR):** These stars are giants or super-giants of that show noticeable pulsation periodicity, sometimes accompanied or interrupted by irregularities. Hence, their light curves are different and variable.

    (a) **SRA:** These are semi-regular late-type giants with persistent periodicity and usually small < 2.5 magnitude in V with periods ranging from 35-1200 days. These stars differ from Mira Variables only due to their small light amplitudes.

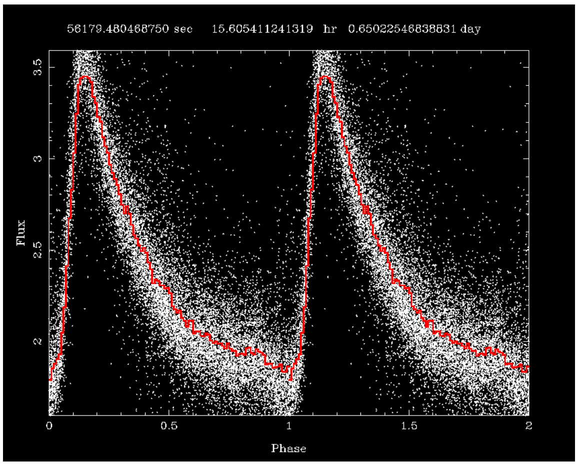
    (b) **SRB:** These are also semi-regular late-type giants but have poorly defined periodicity with mean cycles in the range of 20 to 2300 days or with alternating intervals of periodic and slow irregular changes. Every star of this type is usually assigned a mean period cycle.

    (c) **SRC:** These are semi-regular late-type super-giants having amplitudes of approx. 1 magnitude in V and periodicity ranging from 30 days to several thousand days.

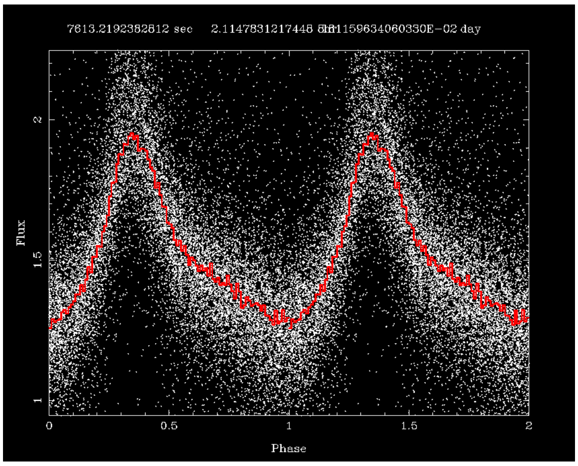
    (d) **SRD:** These are semi-regular variable giants and super-giants of F, G, or K spectral types. They have 0.1 to 4 magnitude in V and period ranging from 30 to 1100 days.

However, it is difficult to categorise in between different types of pulsators just by their light curves hence under this project we categorise them only on the basis on pulsators as a whole and the general trend of their light curves will be noted as, [12]

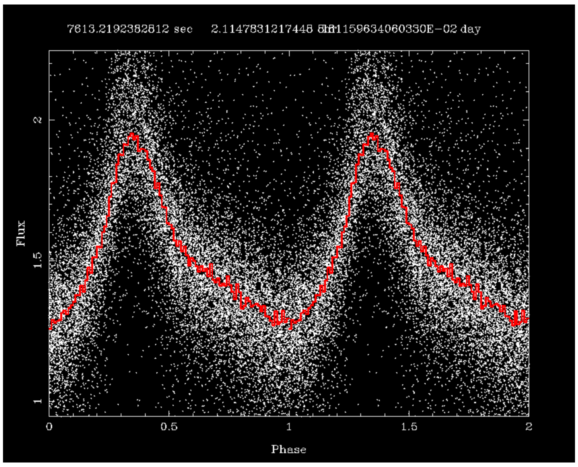
- The folded light curves of pulsating stars have a single maximum per cycle.
- Their asymmetric pulse profiles usually have a steeper rise in brightness and a shallower decay.
- Their pulsation periods can range from a few hours to hundreds of days, for different types of pulsator
- Sometimes the computer software identifies an incorrect period for the star. Their light curves then show some regular structure, but often have overlapping pulses.



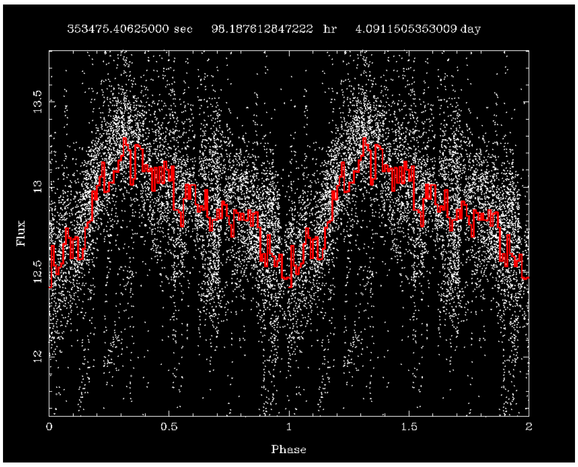
(a) High Amplitude Pulsator



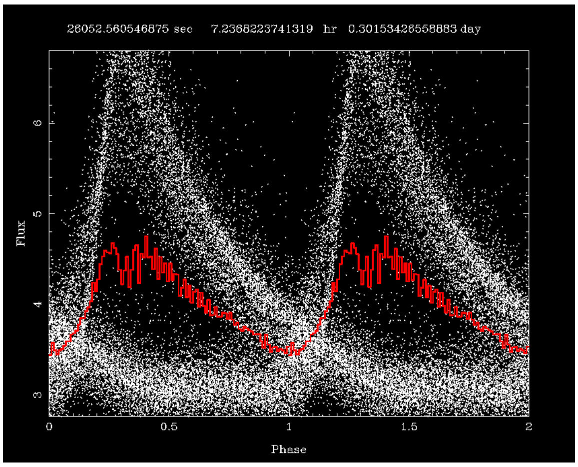
(b) High Amplitude Pulsator



(a) Low Amplitude Pulsator



(b) Long Period Pulsator



(a) Pulsator Folded at wrong period

Figure 1.17: Light Curves of various Pulsating Variable Stars [12]

1.4.3 Eclipsing Binary Variables

Eclipsing variables stars are close binary or multiple star systems whose orbital plane, unlike in rotators, lines up with our line of sight. As one-star orbits in front of the other star, it obstructs the light from the other causing an eclipse and a dip in the combined brightness of the pair. The maximum light is lost when the fainter of the two-star eclipses in front of the brighter. These are known as primary eclipses. In many cases the brighter star can also pass in front of the fainter, resulting in a secondary eclipse. The time lapse between primary eclipses is equal to the orbital period of the system.

The light curves of eclipsing binaries give important clues in determining their physical properties, such as the size, mass, luminosity and temperature of the stellar components. These light curves depend on the sizes and brightness's of the stars, their separation from each other, and the geometry of our view from Earth.

1. **EA-Algol variables:** These are those eclipsing systems that contain spherical or slightly ellipsoidal components and may have missing secondary eclipses. Their orbital periods range from 0.2 to greater than 10,000 days, and their amplitudes of variation can be up to several magnitudes. The beginning and ending of their eclipses are precisely defined in their light curves. The light curve is essentially flat between eclipses, or may vary slightly due to ellipsoidal or physical variability of the components.

2. **EB-β Lyrae type variables:** These are those eclipsing variables that contain ellipsoidal components but the precise times of the beginning and ending of their eclipses are impossible to determine. This happens due to continuous varying of their brightness. They have orbital periods longer than 1 day with amplitudes up to 2 magnitude in V. There is always a secondary minimum present in their light curves. The depth of their secondary eclipses is usually considerably shallower than the primary eclipse.

3. **EW-W Ursae Majoris type variables:** These are those eclipsing variables that contain ellipsoidal components which are almost in contact with each other and orbit around a common center of mass. In these systems as well, the exact times of beginning and ending of the of eclipse are impossible to determine due to the continuously varying brightness of the system. They have orbital periods usually less than 1 day and have amplitude variation typically less than 0.8 magnitude in V. The depths of the primary and secondary eclipses in their light curves are almost equal.

4. **EP- Exoplanet transiting variables:** These are those eclipsing variables whose infinitesimal light variation is caused by the eclipse of one or more of its planets transiting in front of the star as seen from our line of sight.

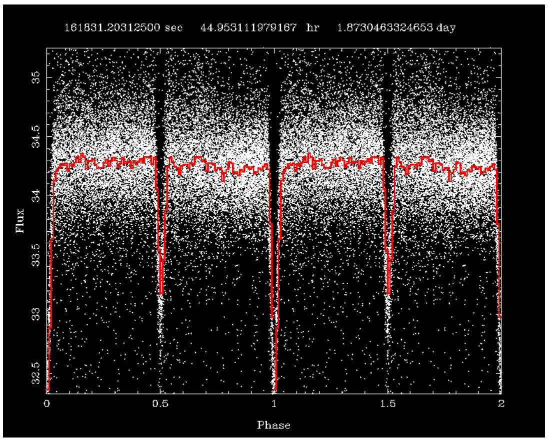
Table 1.3: Eclipsing variable stars arranged in alphabetical order by designation [17]

Variable Type	Designation (and subclasses)	
Algol type	**EA**	Algol-type eclipsing system
β Lyr type	**EB**	β Lyrae-type eclipsing sytem
Planetary eclipsing type	**EP**	Stars showing eclipses of their planets
W UMa type	**EW**	W Ursae Majoris-type eclipsing variables
RS Canum Venaticorum	**RS**	RS Canum Venaticorum-type systems
- additional classification according to the component's physical characteristics		
	GS	one or two giant components
	PN	one components is the nucleus of a planetary nebula
	RS	RS CVn system
	WD	systems with a white dwarf component
	WR	systems with a Wolf-Rayet component
- additional classification based on the degree of filling of filling of filling of inner Roche lobe		
	AR	AR Lac-type detached binary
	D	detached systems with components not filling their inner Roche lobes
	DM	detached main sequence system
	DS	detached system with a subgiant
	DW	detached system like W UMa systems
	K	contact system with both components filling the inner critical surfaces
	KE	contact system/early spectral type
	KW	contact system of late spectral type
	SD	semidetached system in which the surface of the less massive component is close to its inner Roche Lobe.

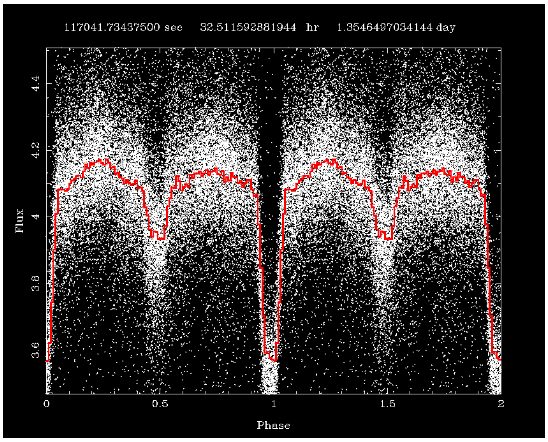
Note: The combination of above 3 classification systems results in the assignment of multiple classifications. These are separated by ["/"] in the data field. Examples are: E/DM, EA/DS/RS etc.

Common characteristics [12] shown by Ea and Eb stars are: -

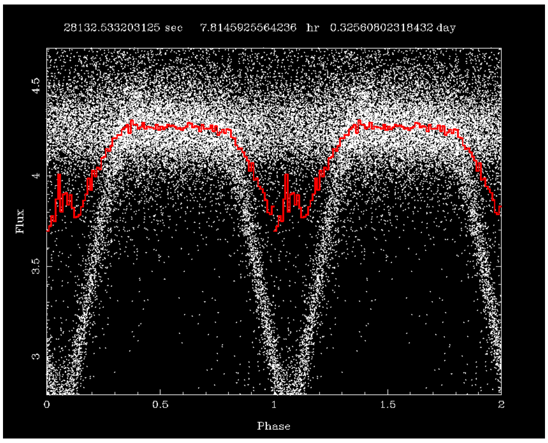
- EA and EB type eclipsing binaries both have two V-shaped minima per cycle, usually of different depths.
- The regions between the eclipses may be relatively flat (EA) or show some variation of brightness (EB).
- The beginning and end of eclipses are clearly defined.
- The minima are narrower than the maxima.
- The orbital periods can range from hours to days.
- Sometimes a light curve is folded at half the true period, so two minima overlay one another and only one minimum is visible. These are still classified as EA or EB.



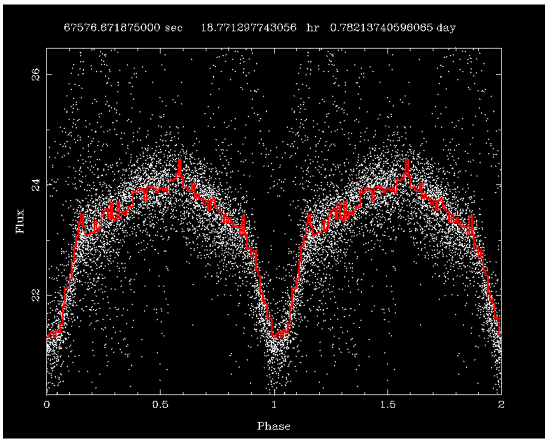
(a) An EA type star folded at correct period



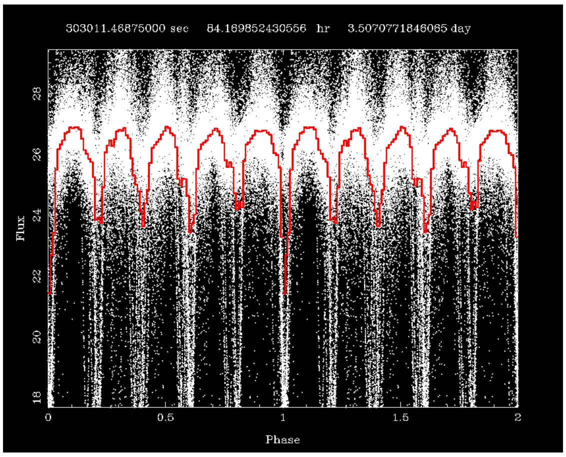
(b) An EB type star folded at correct period



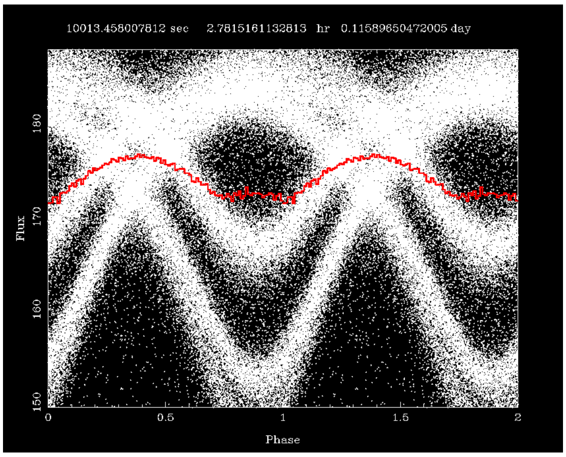
(a) An EA type star folded at half period



(b) An EB type star folded at half period

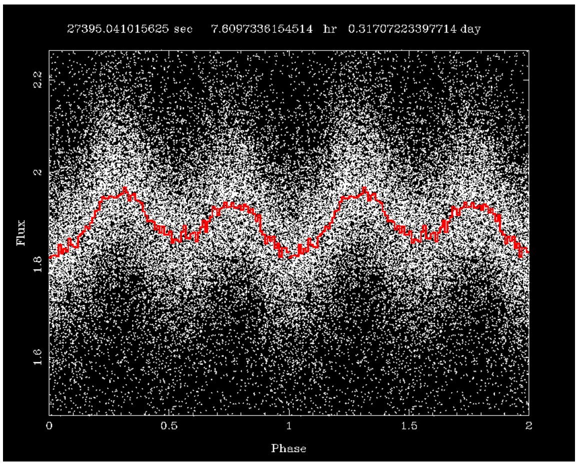


(a) An EA/EB type star folded at the wrong period

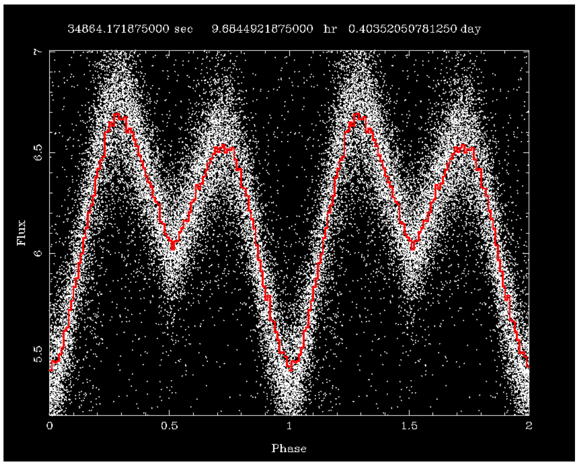


(b) An EA/EB type star folded at the wrong period

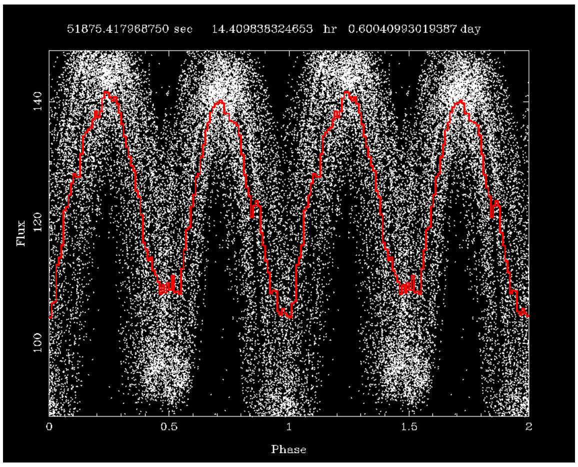
Figure 1.20: Light Curves of various Eclipsing Binary Stars(EA/EB) [12]



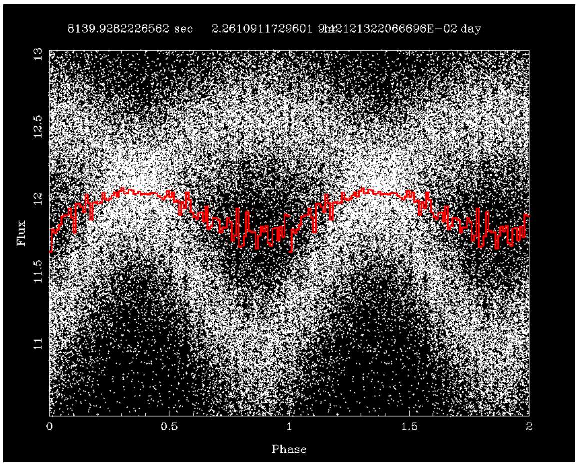
(a) An EW type star folded at correct period



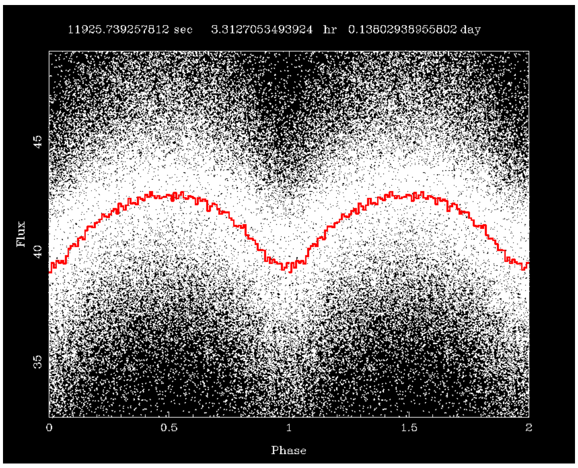
(b) An EW type star folded at correct period



(a) An EW type star folded at wrong period



(b) An EW type star folded at wrong period



(a) An EW type star folded at half correct period

Figure 1.23: Light Curves of various Eclipsing Binary Stars(EW) [12]

Some characteristics [12] shown by EW stars are: -

- EW type eclipsing binaries have two minima per cycle, often of similar depths.
- They have continuously varying brightness throughout the cycle.
- The minima are usually slightly narrower than the maxima.
- The minima may be flat-bottomed.
- The maxima may be different heights.
- The orbital periods are less than 1 day.
- Often, these stars will be identified at half their true period and so only one maximum and minimum per cycle will be seen when the two minima overlay one another. These should still be classified as EW.

Sometimes the automated software identifies an incorrect period for an EA, EB or EW eclipsing binary star which is genuinely periodic at some other period. These light curves show some regular structure, but often have overlapping eclipses or multiple structures at all phases. Such objects are classified as EA, EB or EW eclipsing binary stars folded at the wrong period.

1.4.4 Rotating Variables

Rotating Variable Stars are variable due to irregular or non-uniform brightness in their stellar surface and/or because they have been distorted into ellipsoidal shapes and rotate axially with respect to an observer. The irregular brightness on the stellar surface could be due to presence of spots caused by the star's magnetic field which in turn could cause some thermal or chemical in-homogeneity. And the ellipsoidal shape is due to the strong gravitational tug by a neighboring star. So, we are able to see the variation in brightness as these non-uniformities and deformities rotate in and out of view. These stars show very small amplitude variation and hence are difficult to observe. Up till now there are over 900 stars classified as rotating variable stars in the General Catalog of Variable Stars. [27]

1. **BY Dra stars:** These are red dwarf stars with non-uniform stellar surface and have quasi-periodic variability. They have period typically ranging from hours to 120 days. Their amplitude varies from 0.001 to 0.5 magnitude in V. The large spots present on the surface of BY Dra stars are believed responsible for the UV Ceti type flares, making some of them both rotational and eruptive variables simultaneously.

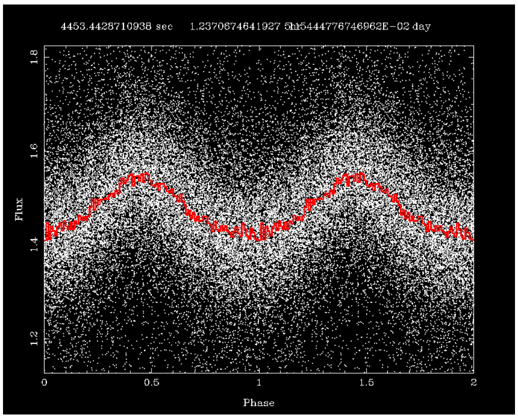
2. **Rotating ellipsoidal variables (ELL):** These stars are close binary systems that vary with periods equal to their orbital motion. These stars are distorted in shape due to tremendous gravitational stress put on them by their ultra-close companion star. Although, mathematically, the probability of the orbital plane of these systems being coincident with our line of sight, causing one star to eclipse the other is greater than zero, but by definition, an ellipsoidal variable cannot show eclipsing behavior.

These stars are all grouped together under the "Rotators" category under the SuperWASP project and their light curves show the following specific characteristics [12]:

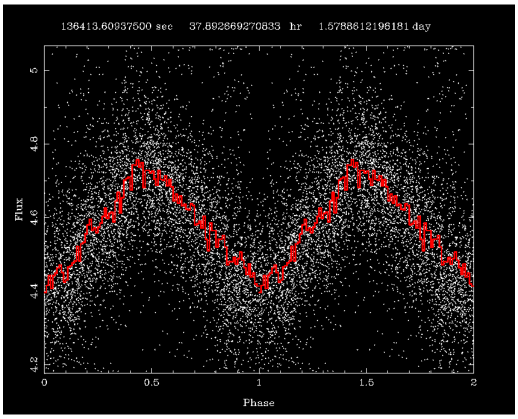
- Their light curves are usually symmetric and sinusoidal light curves that vary continuously.
- They may look somewhat triangular in shape with symmetrical rise and fall.
- They have one maximum and one minimum per cycle, of equal width.
- For short period rotators (periods less than a day or so), because the modulation is essentially sinusoidal, the folded lightcurve will usually look broadly the same shape if it is turned upside down. This is a useful check to distinguish such a lightcurve from that of an EW type eclipsing binary that is folded at half the true period.
- For long period rotators (periods more than a few days), it can be difficult to distinguish between rotational modulation and pulsation. If the profile is symmetrical, it is likely a rotator; if the profile is asymmetrical, it is likely a pulsator.

Table 1.4: Rotating variables stars arranged in alphabetical order by designation [17]

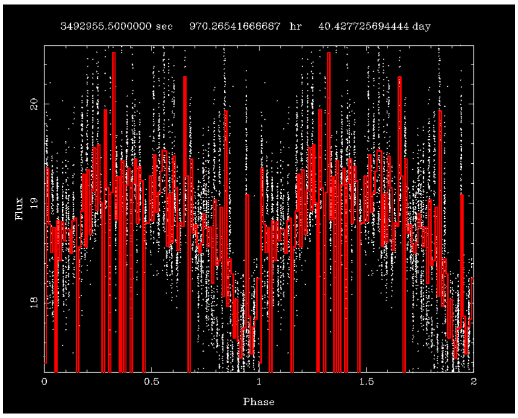
Variable Type	Designation	
α^2 Canum venticulum	ACV	B8p-A7p stars showing small variations due to large "star spots" generated by intense magnetic fields.
	ACVO	Rapidly oscillating α^2 CVn variables.
BY Draconis	BYDRA	dKe-dMe stars showing quasi-periodic light changes.
Ellipsoidal Variables	ELL	rotating ellipsoidal variable stars. May be any spectral class.
FK Cornae	FKCOM	G-K III stars that are rapidly rotating, spectral, and in some cases they are binary systems.
Pulsars	PSR	optically variable pulsars
SX Arietis	SXARI	B0p-B9p stars with intense He I and Si III lines



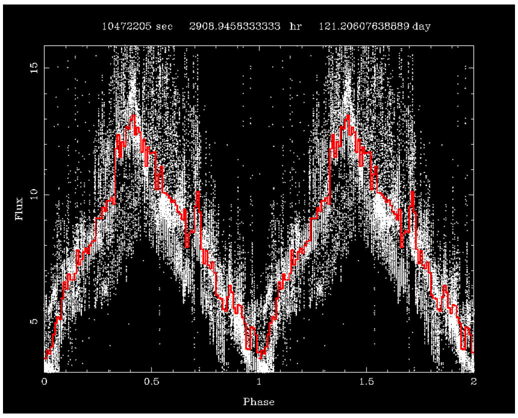
(a) Short Period rotator with sinusoidal profile



(b) Short Period rotator with triangular profile



(a) Long Period rotator with triangular profile



(b) Long Period rotator with triangular profile

Figure 1.25: Light Curves of various Rotating Variable Stars [12]

1.4.5 Unknown [12]

The unknown category simply means that the folded light curves indicates some form of periodicity however, it doesn't fall into any of the above mentioned categories. These can include Semi Regular Variables, Irregular Variables or long period variables.

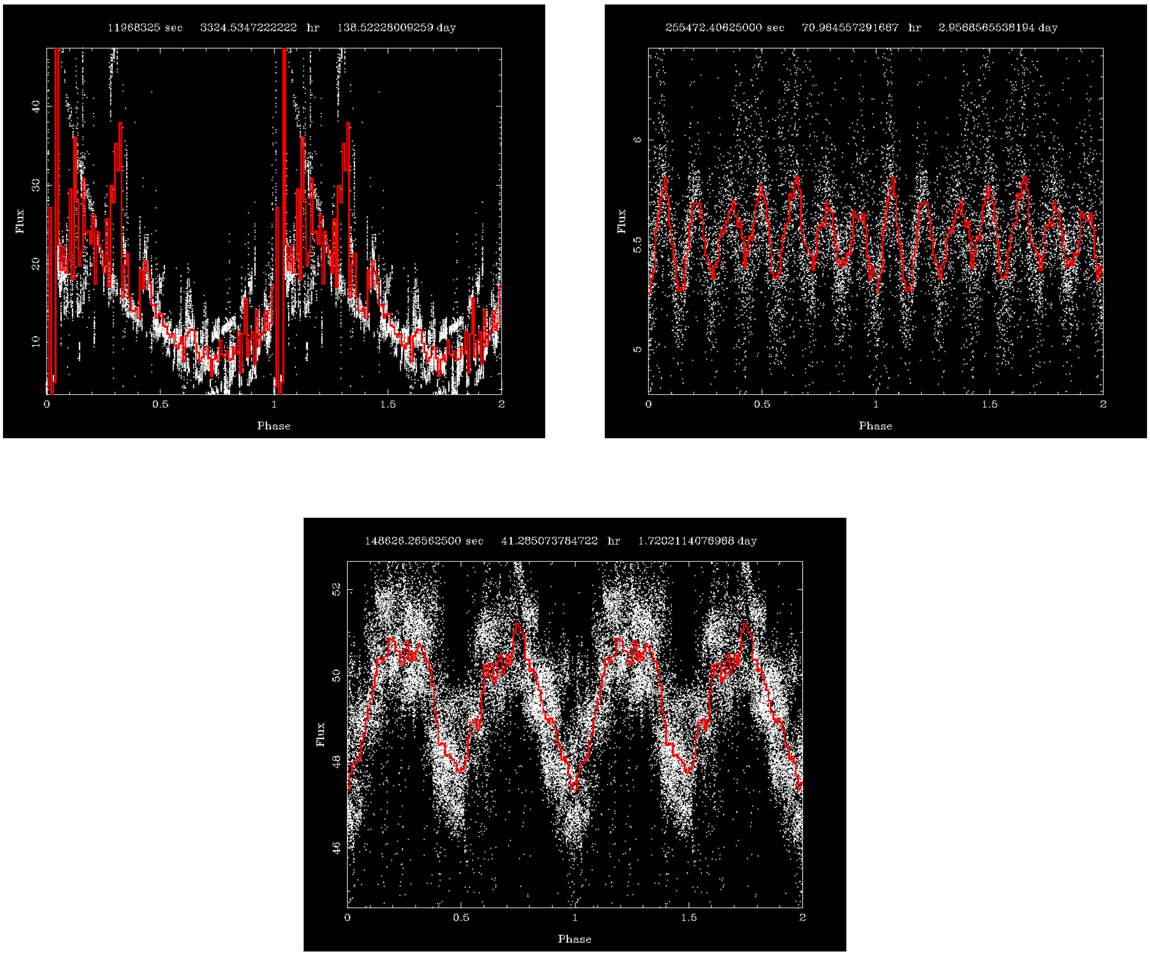


Figure 1.27: Light Curves of Unknown Variable Stars which show a periodic behaviour but don't fall into the above mentioned categories [12]

1.4.6 Junk [12]

Light curves which show no genuine or apparent periodicity are categorised as junk. These are mostly not real celestial objects and only get identified due to data dropouts, systematic artefacts, instrumental errors etc. A junk light curve showing data dropouts can be confused for eclipsing binary variables.This is because,although the average light curve (in red) seems to have a V-shaped dip, but if we carefully look at the individual data points, they do not follow a smoothly varying eclipse.

Some light curves might show some periodicity but as a significant number of period detections result from the systematic effects in the SuperWASP photometric data, such light curves having periods close to integer fractions or multiples of a sidereal day or lunar month (i.e. 1, 1/2, 1/4 d, etc.) are also classified as Junk.

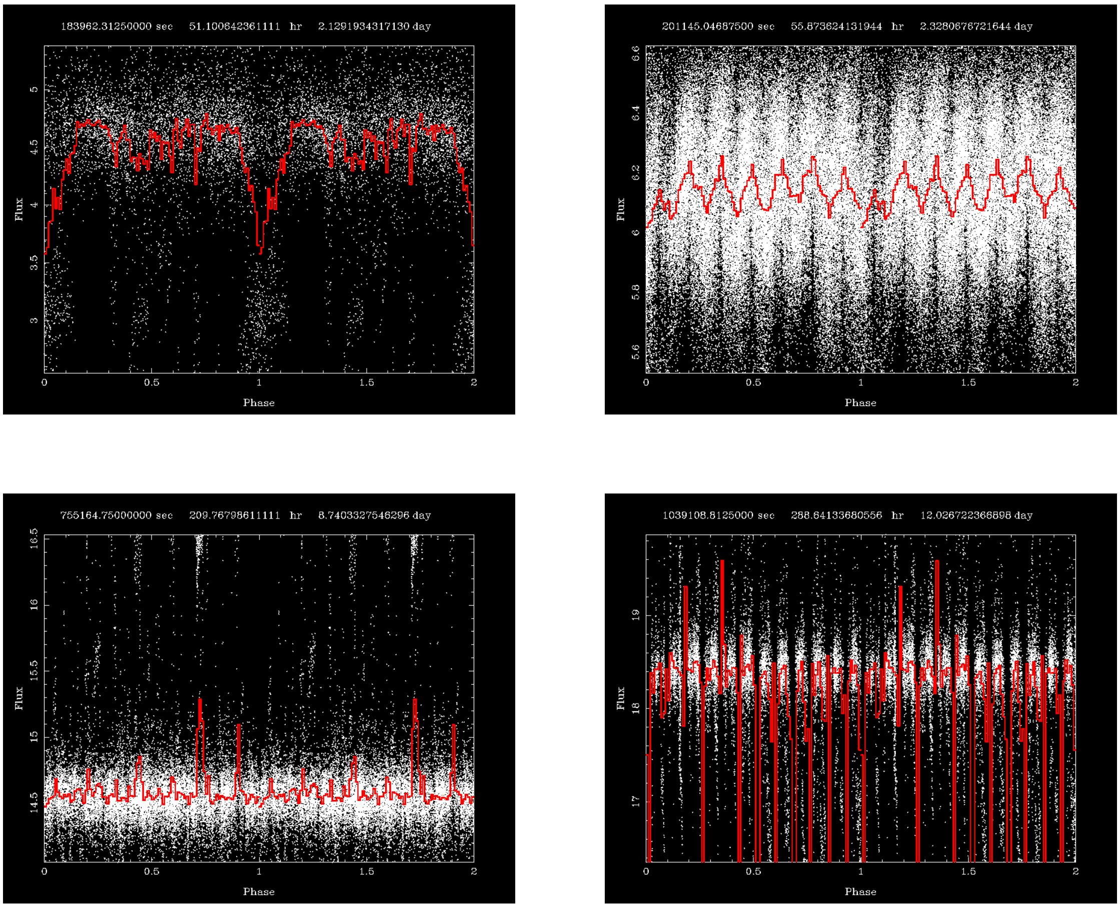


Figure 1.29: Some examples of Junk Light Curves [12]

Chapter 2

Working Methodology

- 500 Light-curves were obtained and classified using the Zooniverse Citizen Science platform and cataloged with their SuperWASP IDS.

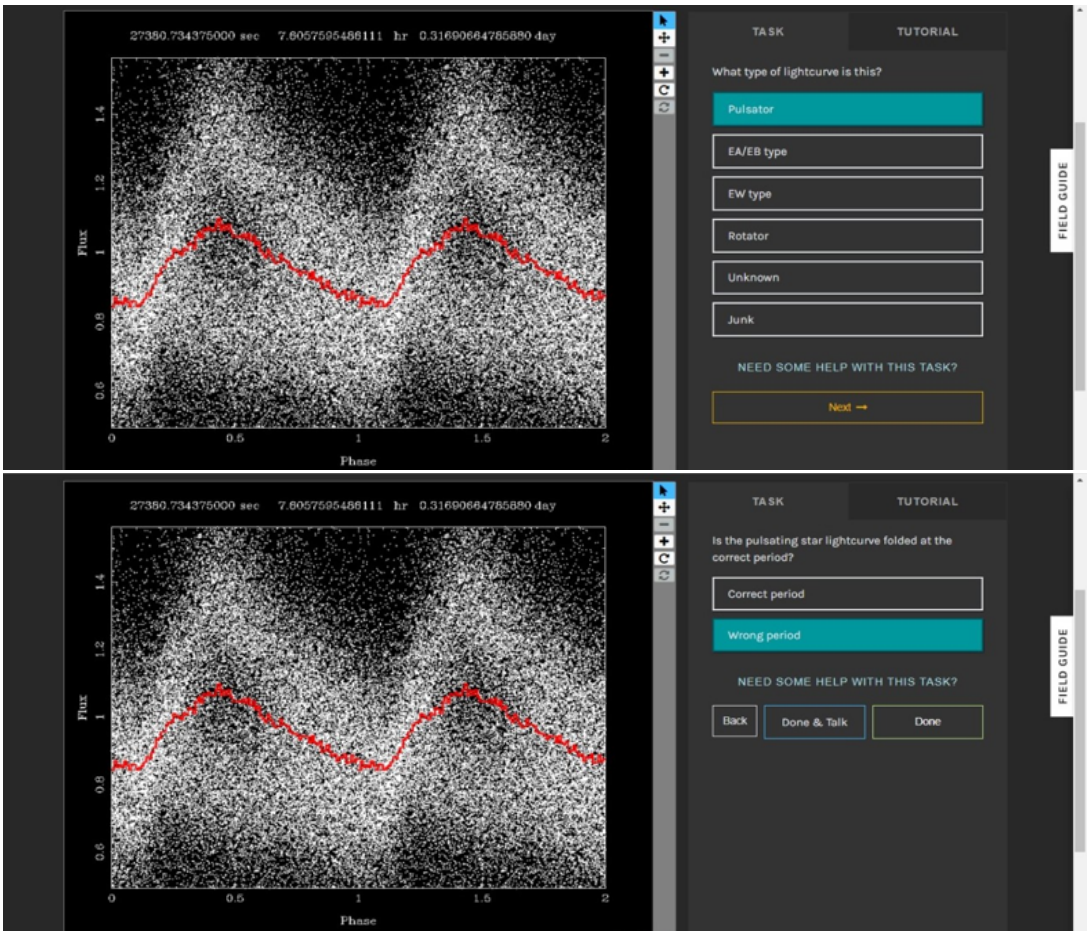


Figure 2.1: Analysing Light Curve on Zooniverse Platform

- The SuperWASP IDs for each light curve was used to obtain its corresponding RA and DEC values from the cerit survey. [28]

- Once all the RA and DEC equatorial coordinates were obtained, they were entered into the Vizier Catalog Service [29]. In the Catalog service, the coordinate is compared with around 2177 catalogs and displayed for use.
- To search for same objects through various catalogs, a criteria to look within circle of 1 arc min radius (assuming the given RA and DEC values to be the centre) was applied.

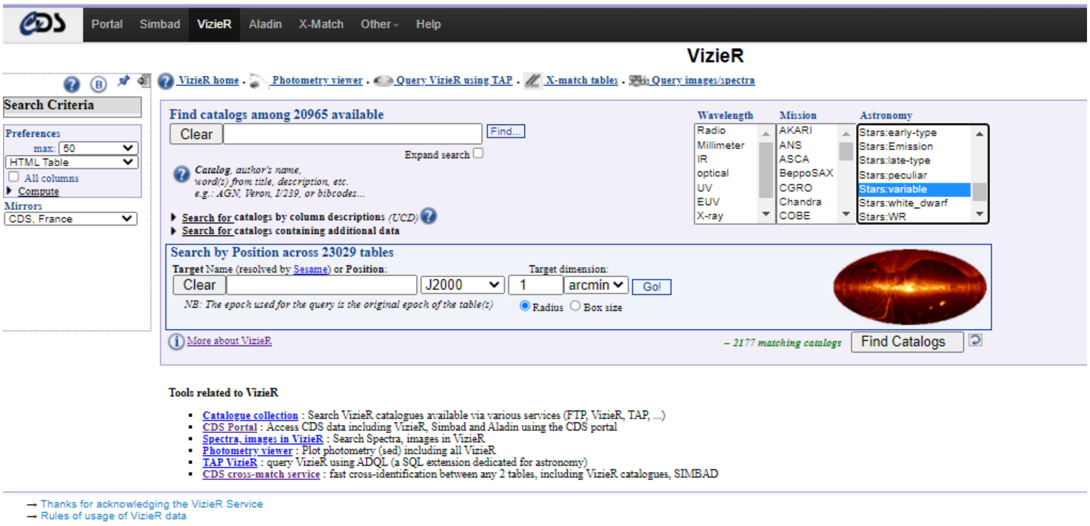


Figure 2.2: Vizier Catalogue Service Platform [29]

- According to the data found through catalogs,
  - Majority of classification of light curves was confirmed.
  - New Data such as magnitudes, corrected period, variability sub-types were obtained and entered into the catalog.
- On completing the catalogue, different characteristic features of the data, like periods and magnitude for each variable star, RA and Dec spread for all stars was plotted to be analyzed.

Chapter 3

Results And Discussion

A total of 500 light curves have been analyzed and cataloged as a part of this project. As mentioned, all the light curves were obtained from www.zooniverse.org

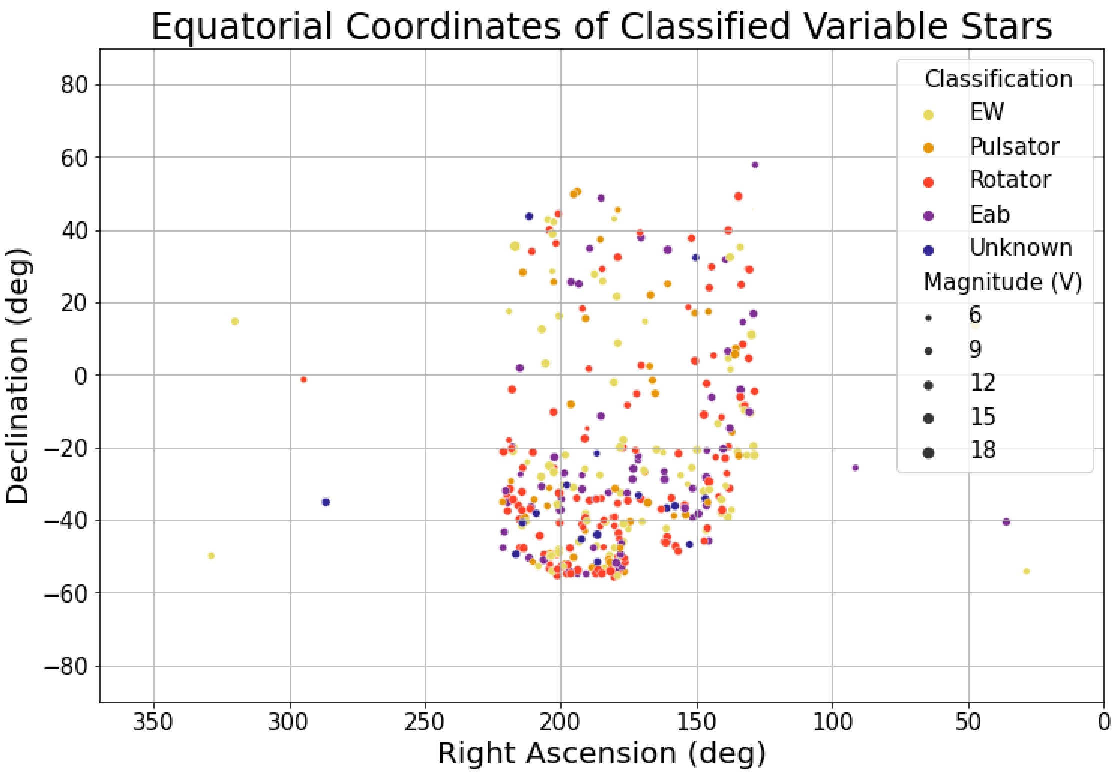


Figure 3.1: Map of all the identified Variable stars

Figure 3.1 shows the equatorial coordinates and positions of all the identified stars. The stars seem to cluster in between RA: 230°to 130°and Dec: -60°to +60°. The reason for the uneven distribution of these classifications is because via the Zooniverse platform,only a few degrees of sky data is available to the people at any particular time.

We can also see a pretty defined curve below -40°declination and this is because, this is the location of the galactic plane. The SuperWASP cameras cannot resolve the dense regions of the Galactic plane and hence we don't obtain any data from that region.

3.1 Preliminary Identification

The preliminary identification was as follows:

Ea/EB	EW	Pulsators	Rotators	Unknown	Junk
67 (13.4%)	88 (17.6%)	44 (8.8%)	123 (24.6%)	18 (3.6%)	160 (32%)

Table 3.1: Distribution of classifications done using Zooniverse SuperWASP Light Curves

According to the preliminary identification, we can see that the majority of analyzed light curves are junk. This can easily be explained due to the presence of systematic defects of the SuperWASP photometric data.

Next most populated stars are the eclipsing binaries with 31% of total. This again is expected as over 80% of the stellar structures in our universe are a part of multiple star system.[30]

About one-fourth stars were identified as rotators and about 9% as pulsators.

Once preliminary identification of all the 500 light curves were done, they were cross-referenced from around 2177 variable star catalogs (eg. Simbad [31], ASAS-SN [32], VSX [20] etc) present in the Vizier catalogue service.

Doing this step had three advantages; first, it became possible to extract out variable stars which weren't classified amongst these catalogues, second, a confusion matrix could be obtained to identify the % of match and shortcomings, if any, present in the identification of folded light curves, third, more information like magnitude, period, photometric data etc about these stars could be obtained from these catalogs.

For cross referencing through Vizier catalogue service, the RA and Dec values for all the light curves was obtained via cerit catalogue. Then these values were input into the vizier platform and all objects within the radius of 1 arc min were identified.

A total of 45 object ids did not have any corresponding variable type classification in the catalogues present on the Vizier Catalogue Service. However, further investigation is still required. Table 1.5 shows the list of all these stars.

Sch.No.	Personal ID	SuperWASP ID	Tentative Classification
1	34_Unknown	J103447.91-473632.0	Unknown
2	51_Rotator	J094154.88-292936.4	Rotator
3	56_Rotator	J122616.62-530113.9	Rotator*
4	62_Eab_HP	J094408.68-421638.5	EA/EB
5	80_Eab_CP	J085207.89+082359.0	EA/EB
6	81_Rotator	J091729.01+314730.4	Rotator*
7	82_Pulsator_CP	J125511.93-535226.2	Pulsator*
8	91_Rotator	J115626.20-545037.0	Rotator*
9	122_Rotator	J101036.84-473857.1	Rotator*
10	133_Unknown	J083913.95-110421.5	Unknown
11	154_Rotator	J083348.25+574833.7	Rotator
12	155_Rotator	J091252.00-194945.6	Rotator
13	181_Eab_HP	J094520.78-022457.9	EA\EB
14	183_Eab_HP	J085427.16+245211.5	EA\EB
15	191_Eab_CP	J125555.07+502500.0	EA\EB
16	193_Eab_HP	J083618.94+165129.5	EA\EB
17	199_Rotator	J141951.36+015125.9	Rotator
18	209_Rotator	J113422.41-284911.7	Rotator
19	210_Eab_HP	J142700.73-352934.2	EA\EB
20	244_Unknown	J141414.04-490748.5	Unknown
21	249_Rotator	J114612.72-503422.6	Rotator*
22	267_EW_CP	J102517.62-214412.9	EW
23	281_Rotator	J094121.73+234510.5	Rotator
24	288_Pulsator_CP	J122849.97-545138.6	Pulsator
25	297_Unknown	J130115.18+494324.6	Unknown
26	304_EW_CP	J122817.41-552101.8	EW
27	307_Rotator	J122458.64-533942.6	Rotator*
28	320_Rotator	J133150.36-540803.1	Rotator*
29	321_Rotator	J124410.30-430408.9	Rotator
30	332_Unknown	J115747.54+213842.7	Unknown
31	340_Rotator	J121936.08-553300.6	Rotator
32	343_Rotator	J122709.64-214300.1	Rotator
33	357_Rotator	J130508.95+253800.0	Rotator
34	381_Rotator	J090839.58-371549.4	Rotator
35	399_Eab_HP	J094322.15-351151.0	Eab
36	401_EW_WP	J124135.05-401835.5	EW*
37	404_EW_CP	J124728.52-420526.2	EW
38	408_Unknown	J211936.55+144446.8	Unknown
39	417_Unknown	J022334.57-403521.2	Unknown
40	443_EW_HP	J094921.54-382212.9	EW
41	445_Rotator	J094400.69-020324.8	Rotator
42	447_Unknown	J122543.87+112034.5	Unknown
43	449_Rotator	J092654.45-361238.1	Rotator*
44	458_Eab_HP	J092631.91-150054.6	Eab

*These are identified Large Amplitude variable stars in Gaia DR2 [33, 34], however no variability type is mentioned in any catalogue

3.2 Confusion Matrix

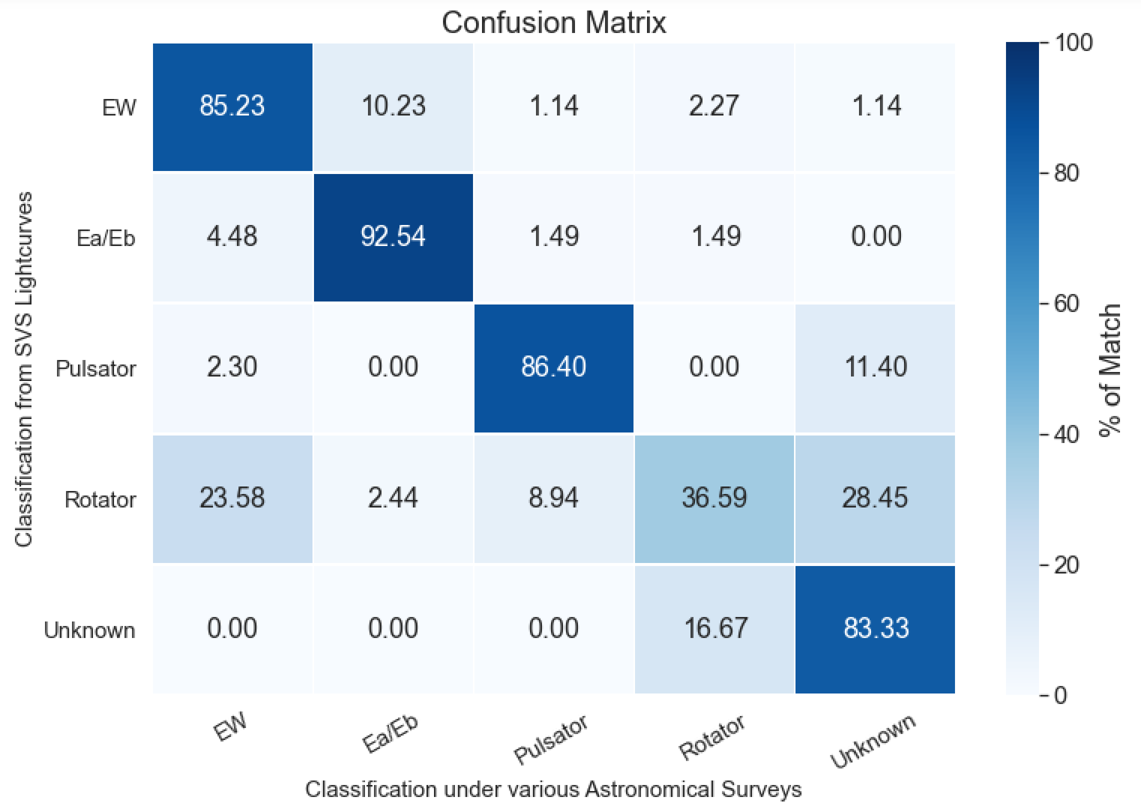


Figure 3.2: Confusion Matrix between Identified Variable Stars from Zooniverse SuperWASP and identified variable stars from other Surveys from Vizier Catalogue Service

We can see from the confusion matrix that except rotators, all other types of identified light curves have an average accuracy of 86.875%.

- The least accurately identified variable stars are rotators. Maximum confusion occurred between rotators and unknown type stars (Irregular and Semi Regular Variables in several catalogs). The reason for this is mainly because many of these light curves were folded at incorrect periods(usually less than 100 days for rotators) whereas, the given Irregular and Semi Regular Variables in various surveys mostly had periods greater than 100 days.
- A high confusion level is also seen between EW type eclipsing binaries and rotators. Again this can be due to the presence of EW type eclipsing binaries folded at half correct period. This is because half folded EW loose a lot of features characteristic to them, for example presence of a V-shaped minima can become broader. This can lead to the confusion between a typical short period sinusoidal rotator and half folded EW star.
- Another similar case is the confusion between where several EB type eclipsing binary stars got confused with EW due to the loss of clarity by folding the light curve at half period.

- Confusion amongst unknown variable types, pulsators and rotators is also noticeable. This is probably because the SuperWASP program used to identify the periods of the light curve identified the initially visible shorter periods (say 50d) and missed the irregularity that occurred at the longer time span(say 500 day). This is because most of the SRs and Irregular stars found in the catalogs have pretty long periods (usually >100 days).

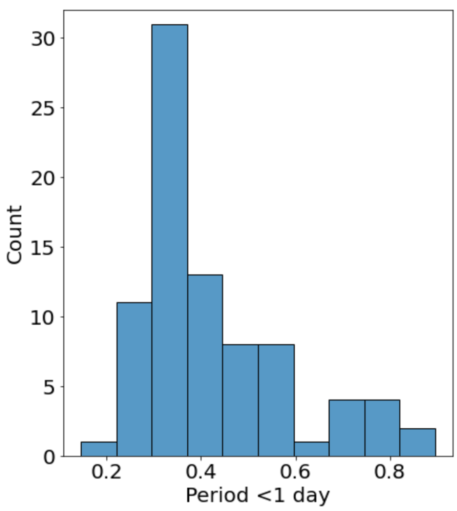
3.3 Variability Periods

An important factor in cataloging variable stars is being able to identify and extend the extremes of each type i.e. variable stars with extremely short or long periods, or extremely high or low amplitudes. This is important because, studying the extreme variable types of each category in more detail can reveal new insights about the stellar structure, stellar evolution etc.

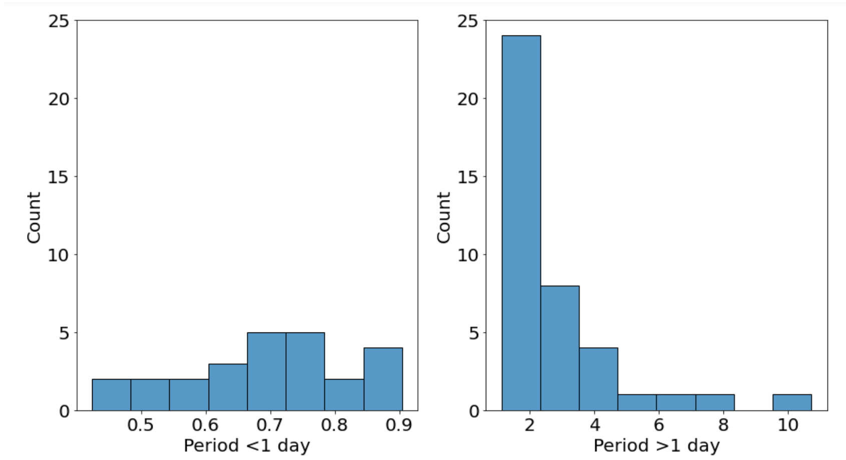
Now to achieve this we need a criteria as the standard definitions of periods,therefore the criteria [18] been followed here is :

1. **EA/EB**: 0.3 d <P <10 d
2. **EW**: 0.22 d <P <1 d
3. **Pulsator**: δ Scuti (<0.3d), RR Lyrae (0.44 - 0.88d), W Virginis (0.8<P <35d), Cepheids (weeks to months),Mira (>100 d).
4. **Rotator**: P >0.5 d (periods range from hours to months

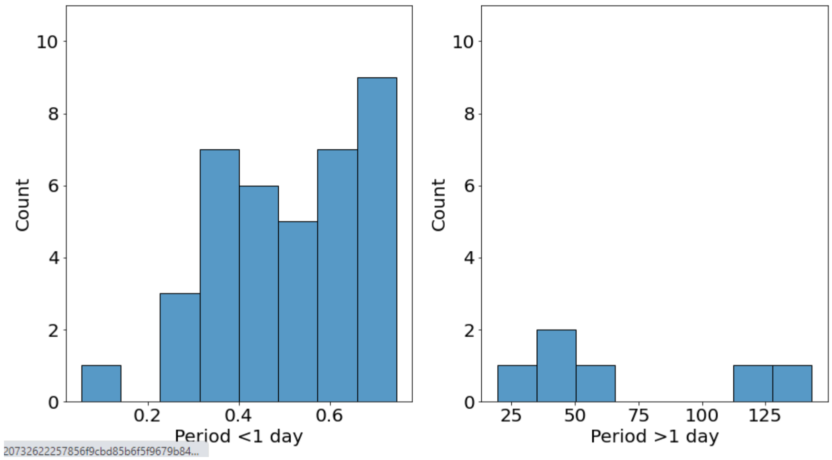
From figure 3.3 we can see the spread of the periods for each variability types. All the periods plotted are correct periods given in Zooniverse SuperWASP light curves. However, for the stars having light curves folded at half or wrong period were obtained form various other catalogs.



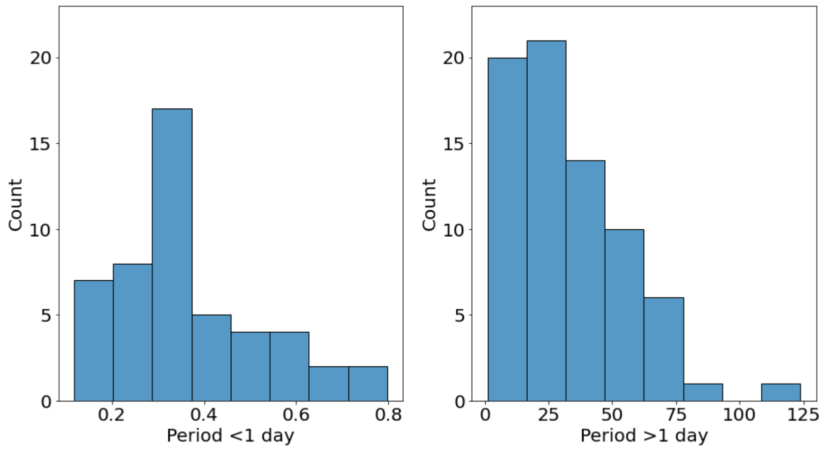
(a) EW type Eclipsing Binary Stars



(b) EA and EB type Eclipsing Binary Stars



(c) Pulsating Variable Stars



(d) Rotating Variable stars

Figure 3.3: Histogram of identified Periods for all four variable star classifications

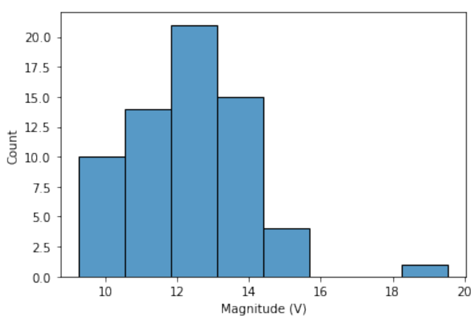
Studying the graphs we can see some gaps in the histograms for example, there is 1 eclipsing binary star falling outside its range. Similarly, the results from all histograms are summarized below.

Variability Type	Sch. No.	Personal ID	SuperWASP ID	Period
EA\EB	1	434_Eab_HP	J143549.75-180224.1	10.72
EW	1	135_EW_CP	J090220.67+071309.9	1.818
Pulsator	1	120_Pulsator	J142027.40-395253.4	0.298
Rotator	1	330_Rotator	J114119.13-344444.0	0.118
	2	321_Rotator	J124410.30-430408.9	0.122
	3	449_Rotator	J092654.45-361238.1	0.135
	4	387_Rotator	J085746.88-221913.8	0.18
	5	356_Rotator	J135541.69-381454.2	0.183
	6	342_Rotator	J122721.71-341913.7	0.274

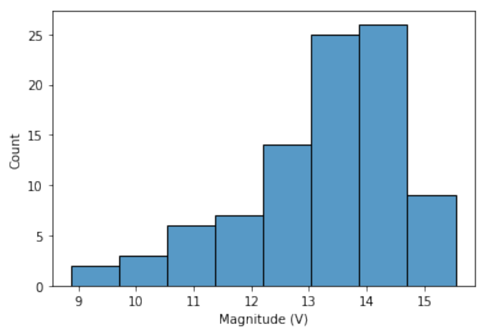
Table 3.2: Variable stars having extreme periods

3.4 Magnitude

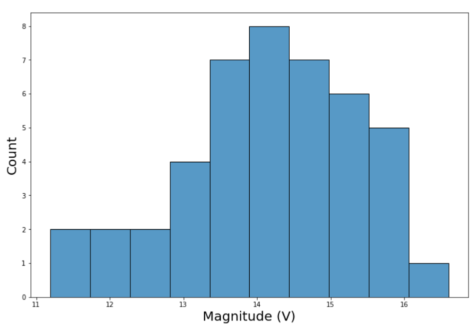
The distribution of NOMAD V magnitude of SVS stars with a variable-type classification and correct period classification are given in fig 1.34.



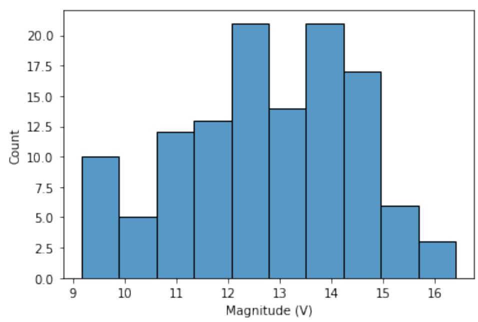
(a) EA and EB type Eclipsing Binary Stars



(b) EW type Eclipsing Binary Stars



(a) Pulsating Variable Stars



(b) Rotating Variable stars

Figure 3.5: Histograms of identified Magnitudes for all four variable star classifications

The Magnitude Range for

- **Pulsators:** 11.19 V <Mag <16.6 V
- **EA/EB:** 9.24 V <Mag <19.54 V
- **EW:** 8.88 V<Mag <15.54 V
- **Rotator:** 9.16 V<Mag <16.4 V

Therefore for all the light curves analyzed we can say that their NOMAD V Magnitude lies between 8.88 <Mag <19.54

3.5 Junk Light Curves

The results of cross-referencing the Junk light curves with various other catalogs are shown in table 1.7.

Star	EA/EB	Pulsator	Rotator	Irr/SR	Galaxy	Quasar	SN
26	2	2	3	20	8	6	1

Table 3.3: Distribution of Junk Light Curves

We can see that around 53 out of 160 identified junk light curves are actually stars. The reason these stars could not be identified using the folded light curves is probably because they have quite low amplitude variation which gets lost amongst noise leading to a light curves showing no genuine periodicity. Another reason could be the folding of light curves at wrong periods which leads to loss of variability information. We can also see from the magnitude distribution of Junk Light Curves in fig 1.35 that the magnitude peaks around 30 days (period of lunar month) and 2 days which show data variation in multiples of sidereal day i.e. 1,1/2,1/3 etc.

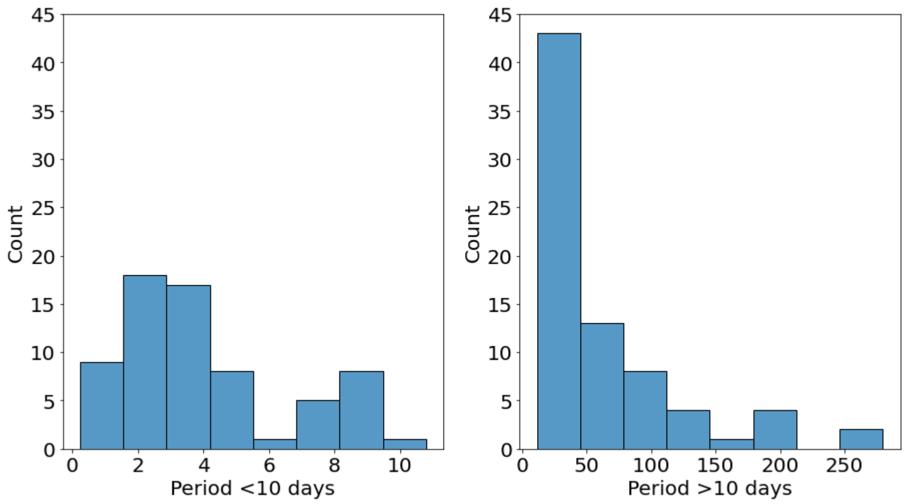


Figure 3.6: Distribution of Detected Period for Jink Light Curves

Chapter 4

Future Prospects

In the present work, we performed the preliminary study for the Light Curves analysis. We identified their variability types, found out the range period and magnitudes for the identified variable stars, ending up with some extreme stars that fall outside their respective variability type ranges. We also discussed certain-short comings that come while analysing these folded light curves.

Building on this work, we can now work on each of the 44 identified Variable Stars. As they were not available in the Vizier Catalogue Service we can attempt to work on them further and discover at least some previously unidentified and unclassified variable star. The 8 stars having extreme periods and falling out of the standard criteria of their respective period ranges should be now studied in depth using additional information like its spectroscopy data, temperature and colour data etc. This would help us gain more insight as to whether the extreme period behaviour is due to some underlying physical phenomenon or not. Obtaining such answers can also help one gain more insight in the stellar evolution and stellar structure of such stars.

The catalogue can even be appended with even more columns of information, for example, adding photometric data of optical bands. Obtaining these magnitude values, we can calculate the colour of these stars and plotting color-color diagrams to identify their spectral type as well as the temperature. Similarly we can even segregate the catalogue for each variability type and study their trends separately.

Under the Zooniverse Citizen Science Project of SuperWASP variable stars, even more light curve classifications can be done which would lead to the completion of the SuperWASP catalogue containing over 2 million light curves. This will also aid in their aim to develop a web portal in order to give researchers and public access to the output of this project. Moreover, this data will be used to train a convolution neural network (CNN) which will then be used to categorize even more light curves for different stars.

Bibliography

[1] P. et al., **118**, 1407 (2006).

[2] S. R. Kane, A. Collier Cameron, K. Horne, D. James, T. A. Lister, D. L. Pollacco, R. A. Street, and Y. Tsapras, Monthly Notices of the Royal Astronomical Society **353**, 689 (2004).

[3] *Dealing with ccd data, lecture 12*, `http://slittlefair.staff.shef.ac.uk/teaching/phy217/lectures/instruments/L12/index.html`.

[4] E. Høg, C. Fabricius, V. V. Makarov, S. Urban, T. Corbin, G. Wycoff, U. Bastian, P. Schwekendiek, and A. Wicenec, **355**, L27 (2000).

[5] M. et al, **125**, 984 (2003).

[6] *Variable star astronomy: variable stars, light curves, and variability, chapter 11*, `https://www.aavso.org/education/vsa`.

[7] M.-H. Grondin, R. W. Romani, M. Lemoine-Goumard, L. Guillemot, A. K. Harding, and T. Reposeur, The Astrophysical Journal **774**, 110 (2013).

[8] *Superwasp variable stars from zooniverse citizen science platform*, `https://www.zooniverse.org/projects/ajnorton/superwasp-variable-stars/classify`.

[9] *Nasa: imagine the universe, light curves and what they can tell us*, `https://imagine.gsfc.nasa.gov/science/toolbox/timing1.html`.

[10] *Variable star astronomy: variable stars and phase diagrams, chapter 12*, `https://www.aavso.org/education/vsa`.

[11] *Nasa: imagine the universe, timing analysis*, `https://imagine.gsfc.nasa.gov/science/toolbox/timing2.html`.

[12] *Superwasp variable stars project*, `https://www.zooniverse.org/projects/ajnorton/superwasp-variable-stars/`.

[13] *Fourier transform*, `https://commons.wikimedia.org/wiki/File:FFT-Time-Frequency-View.png#filelinks`.

[14] M. Templeton, **32**, 41 (2004).

[15] G. Foster, **112**, 1709 (1996).

[16] *Variable star classification and light curves,an aavso course*, https://bit.ly/3gF9rU2.

[17] G. Good, Observation of Variable Stars, 10.1007/978-1-4471-0055-3 (2003).

[18] H. B. Thiemann, A. J. Norton, H. J. Dickinson, A. McMaster, and U. C. Kolb, Monthly Notices of the Royal Astronomical Society **502**, 1299 (2021).

[19] I. Soszyński, *Ogle atlas of variable star light curves*.

[20] C. Watson, A. Henden, and A. Price, Journal of the American Association of Variable Star Observers (JAAVSO) **35**, 414 (2007).

[21] L. Eyer, L. Rimoldini, M. Audard, R. Anderson, K. Nienartowicz, F. Glass, O. Marchal, M. Grenon, N. Mowlavi, B. Holl, G. Clementini, C. Aerts, T. Mazeh, D. Evans, L. Szabados, A. Brown, A. Vallenari, T. Prusti, and J. de Bruijne, (2019).

[22] G. Handler, Planets, Stars and Stellar Systems, 207 (2013).

[23] S. Rosseland and G. Randers, Astrophysica Norvegica **3**, 71 (1938).

[24] N. Baker and R. Kippenhahn, Z. Astrophys. **54**, 114 (1962).

[25] A. Brickhill, Monthly Notices of the Royal Astronomical Society **251**, 673 (1991).

[26] H. J. Lamers and E. M. Levesque, in *Understanding stellar evolution*, 2514-3433 (IOP Publishing, 2017), 21-1 to 21–14.

[27] N. N. Samus', E. V. Kazarovets, O. V. Durlevich, N. N. Kireeva, and E. N. Pastukhova, Astronomy Reports **61**, 80 (2017).

[28] O. W. Butters, R. G. West, D. R. Anderson, A. Collier Cameron, W. I. Clarkson, B. Enoch, C. A. Haswell, C. Hellier, K. Horne, Y. Joshi, S. R. Kane, T. A. Lister, P. F. L. Maxted, N. Parley, D. Pollacco, B. Smalley, R. A. Street, I. Todd, P. J. Wheatley, and D. M. Wilson, **520**, L10, L10 (2010).

[29] O. F. et. al, *The VizieR database of astronomical catalogues*, http://vizier.unistra.fr/viz-bin/cat/J/A+A/610/A20).

[30] *Most stars are binary stars*, https://www.atnf.csiro.au/outreach/education/senior/astrophysics/binary_intro.html.

[31] M. Wenger, F. Ochsenbein, D. Egret, P. Dubois, F. Bonnarel, S. Borde, F. Genova, G. Jasniewicz, S. Laloë, S. Lesteven, and R. Monier, **143**, 9 (2000).

[32] T. Jayasinghe, C. S. Kochanek, K. Z. Stanek, B. J. Shappee, T. W. .-S. Holoien, T. A. Thompson, J. L. Prieto, S. Dong, M. Pawlak, O. Pejcha, G. Pojmanski, S. Otero, N. Hurst, and D. Will, **503**, 200 (2021).

[33] G. Collaboration, A&A **616**, A1 (2018).

[34] G. Collaboration, A&A **595**, A1 (2016).