

# The Automated Detection and Classification of Variable Stars

Nikola Whallon

*Advisors: Carl Akerlof, Jean Krisch*

## **Abstract**

Variable stars have been an incredibly important discovery since the early days of astronomy. Certain types of variable stars, such as Cepheids, can give not only great insight into the inner dynamics of stars, but also essential parameters of the star, such as luminosity, mass, and the distance from the observer. The number of discovered variable stars has risen dramatically over time, and several catalogues of variable stars (such as The International Variable Star Index, VSX<sup>[1]</sup>) have appeared. These catalogues provide a very useful tool for researchers to gather data on statistically large samples of variable stars, which can lead to more insights, such as the distributions of specific types of variable stars. In fact, stars such as Cepheids, whose distance can be readily inferred from their brightness, are being used to obtain distances to star clusters and galaxies, which helps to map the spatial scale of the universe.

This project aims at writing a series of programs which will automatically detect and classify every variable star visible by the Robotic Optical Transient Search Experiment (ROTSE)<sup>[2]</sup>, and prepare the correct information required to catalogue these stars in VSX.

## **Introduction**

### *ROTSE*

ROTSE is a project run by Professor Carl Akerlof of the University of Michigan. Its goal is to detect transients on the sky, particularly gamma ray bursts and supernovae. However, the techniques used to detect such transients are also applicable to variable stars.

The ROTSE project currently employs four ROTSE-III telescopes. They are situated in Turkey, Namibia, Australia, and Texas. These telescopes are unfiltered optical telescopes with a wide  $1.85^\circ \times 1.85^\circ$  field of view. The telescopes are fully automated, and routinely take pictures of the sky every night. In order to organize the image data taken by the telescopes, the telescopes typically only image pre-selected portions of the sky. These ROTSE-fields are characterized by their celestial coordinates – right ascension (RA) and declination (DEC). Organizing images taken by ROTSE into these fields makes various types of data analysis more convenient. For example, the image subtraction technique, outlined in detail below, requires that two images of the same part of the sky be compared with one another. Thus ROTSE-fields can be directly used and compared for this technique.

Besides taking routine and scheduled pictures of ROTSE-fields in the sky, ROTSE also accepts triggers from the Gamma-ray Coordinates Network (GCN)<sup>[3]</sup>. If any correspondent detects a potential gamma-ray burst, and is connected to the GCN, the GCN will send messages to a list of subscribers. When ROTSE receives one of these triggers, the ROTSE telescopes will autonomously point in the direction of the gamma-ray burst candidate and begin collecting data.

### *Variable Stars*

Variable stars are stars whose apparent brightness changes over time. They have had an

incredibly important role in the developments of astronomy and theoretical astrophysics. Of particular note are Cepheid variable stars. These stars, the first of which was discovered in 1784 by Edward Pigott, were shown in 1908 by Henrietta Swan Leavitt to have a relationship between their period and luminosity. By knowing the luminosity of a star, the distance to the star can be calculated. Thus, Cepheids have been very useful as standard candles – allowing astronomers to calculate the distance to various star clusters and galaxies. This has helped astronomers determine the distance and energy scale of many phenomena, from the distribution of galaxies to the expansion of the universe.

Another noteworthy type of variable star is a star whose brightness changes over time due to an orbiting planet. Carefully analyzing variable stars with signs of orbiting companions has been the major method of detecting exo-solar planets for years, and continues to be an active area of research.

There are two primary categories of variable stars – intrinsic variables and extrinsic variables. Intrinsic variable stars change in luminosity due to internal factors. Extrinsic variable stars change in brightness due to external factors, such as in a binary system when one star eclipses.

There are three main subgroups of intrinsic variable stars: pulsating variables, eruptive variables, and cataclysmic variables. A pulsating variable star changes in luminosity when its radius expands and contracts due to an unstable balance between thermal pressure and gravitational collapse. Eruptive variables change in luminosity due to flares, mass ejections, and other surface eruptions. Cataclysmic variables are stars which undergo drastic changes in luminosity, such as novae and supernovae.

There are two main subgroups of extrinsic variable stars: eclipsing binaries, and rotating variables. Eclipsing binary stars change in apparent brightness when one of the stars in the system partially or completely eclipses the other. Rotating variable stars change in brightness due to their rotation. For example, if a star has a very large sunspot, every time the star rotates the brightness of the star will appear to dip.

There are many types of variable stars within these subgroups, but many of them change in brightness due to similar physical processes. During this study, the vast majority of variable stars found were of the type RR Lyrae (a type of pulsating variable star), and eclipsing binaries. Because of this, a deeper look into the physical processes responsible for these stars' variability is presented here.

### *RR Lyraes*

RR Lyrae variables change in brightness due to a periodic cycle of thermal expansion and gravitational collapse. The core of RR Lyraes is primarily Helium. The rest of the star is primarily Hydrogen. Gravity pulls these outer layers of Hydrogen towards the center of the star, causing the core and surrounding layers to heat up. Once hot enough, a shell of Hydrogen around the Helium core will nuclear-fuse into more Helium. This fusion process releases a huge amount of thermal energy, and cause the star to rapidly expand. By expanding, the core and Hydrogen-fusing shell will cool down once more, and fusion will cease. Once again, the star will gravitationally collapse until fusion starts again. And so the cycle continues. This is the typical process by which pulsating variable stars undergo.

RR Lyrae variable stars have masses of around 0.8 solar masses, and are Population II stars. Population II stars have relatively little metal in them, and are the oldest stars visible today. They are composed almost entirely out of hydrogen and helium, and through their processes of fusion, are believed to have been the source for almost all the other elements in the periodic table as well. The way in which this occurs is in sufficiently massive Population II stars, hydrogen will fuse into helium in the core to the point where helium will then fuse into carbon, which may eventually fuse with more helium to produce oxygen, and so on. The stage of fusion that these stars get to is determined by the mass of the star, with more massive stars capable of later stages of fusion. Each stage of fusion is also a major

stage in a star's lifespan, and results in different types of pulsating variable stars (of which the RR Lyrae type is one).

As Population II stars, RR Lyraes are commonly found in globular clusters, which typically consist of older stars. They are found at all galactic latitudes, although, due to their mass and stage in the stellar life cycle, have been too dim to detect in other galaxies until very recently. However, they are incredibly abundant in this galaxy, and many more RR Lyraes have been discovered than Cepheid variables. In the 1980s, approximately 1900 RR Lyrae variable stars had been discovered in globular clusters. Current estimates predict that there are around 85000 in the Milky Way<sup>[4]</sup>.

While RR Lyrae variable stars are very common, they are also particularly interesting because, similar to Cepheid variables, there is a well known relationship between the period of the star and the luminosity (and thus absolute magnitude) of the star. Knowing the absolute magnitude of a star (and measuring its apparent magnitude) allows the distance to the star to be calculated.

### *Eclipsing Binaries*

The other type of variable star seen in abundance by the ROTSE telescopes is eclipsing binary stars. While the process by which eclipsing binaries appear to change in brightness is quite straightforward, there are a few important notes to make.

A binary star is a system of two stars orbiting each other. When one star partially or fully eclipses the other, the overall brightness of the binary system appears to go down. Thus, as the two stars eclipse each other, two dips in the brightness of the system can be observed per rotational cycle.

For the majority of the stars' orbital cycle, there is no eclipsing, so the dips in brightness due to eclipsing are often short and extreme, giving the lightcurve of binary stars a very distinct shape. However, depending on the angle of the orbital plane relative to our viewpoint on Earth, the eclipses can have widely varying effects on the degree of the brightness dip. Many of the stars discovered in this project were eclipsing binaries, and their lightcurves, showing these brightness dips, are analyzed below.

### *Number of Variable Stars*

Among all types of variable stars, the General Catalogue of Variable Stars (GCVS) as of 2008 contains approximately 46000 confirmed variable stars within the Milky Way, 10000 in other galaxies, and 10000 suspected variable stars<sup>[5]</sup>. When the VSX launched in 2005, it catalogued the entire Combined General Catalog of Variable Stars (GCVS 4.2, 2004), as well as many published variable stars from several projects, including ROTSE<sup>[6]</sup>. Currently, the VSX contains entries on over 200000 variable stars.

Although the total number of variable stars in the galaxy is unknown, one of the more ambitious star-cataloging projects, GAIA, predicts that it will be able to see approximately 18 million variable stars out of the estimated 1 billion stars of less than 20 magnitude it aims to catalogue<sup>[7]</sup>. With these 1 billion stars only representing roughly 1% of the Milky Way, one can imagine the total number of variable stars of all types in the galaxy to be on the order of billions.

With this number so large, individual discoveries of variable stars are somewhat insignificant, and certainly will become more insignificant as projects like GAIA do eventually launch. Because of this, it would be economical to develop methods by which computers can detect and classify large quantities of variable stars without much human intervention. The automated procedure I have developed to do just this can be viewed as a microcosm of the way in which variable stars will be primarily be discovered by projects such as GAIA. I have been able to create a series of computer

programs that can automatically detect and classify 10s of variable stars out of fields containing 1000s of stars, with little to no human intervention, on the order of a day.

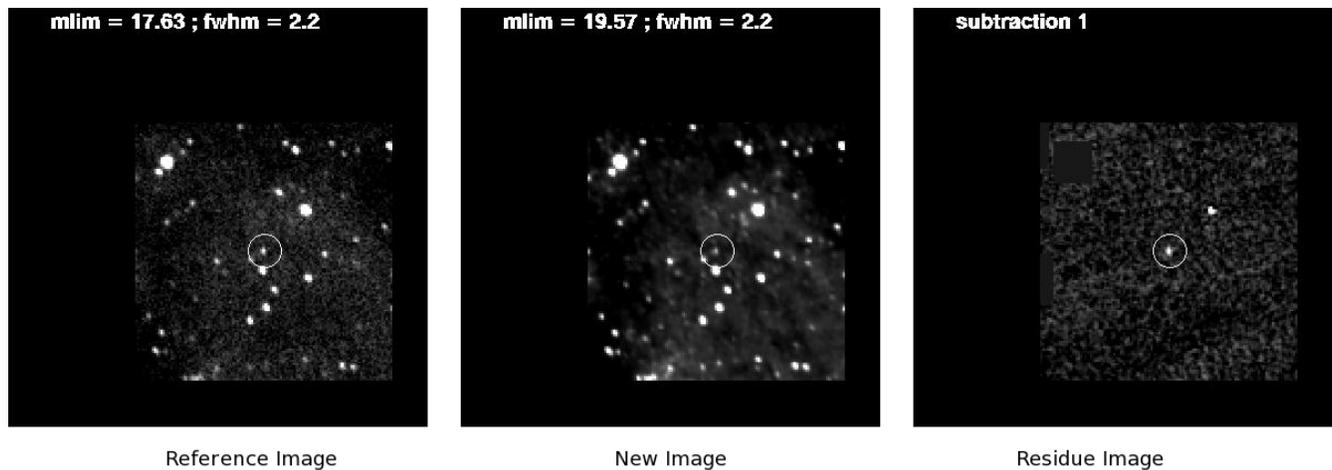
## **Previous Procedure**

### *Image Subtraction*

The general method of detecting transient candidates at ROTSE is an image subtraction technique developed by Akerlof and Yuan. This technique was originally intended for use in discovering supernova, but has also been very effective in detecting variable stars, asteroids, and other transients.

The technique involves subtracting a recent image from a reference image. The algorithm for subtracting images does not simply subtract the pixel values, but takes into account surrounding pixel values as well. The reference image is typically an image taken on a day with relatively little distortion to the image due to weather, telescope defocus, or other external factors, resulting in a particularly clear image. After subtracting the two images, a residue image is left. The majority of the residue image is dark, representing areas where the two images were very close to the same brightness. However, several bright point sources also appear in the residue image. These points then represent areas where the brightness of the two images differed. These point sources in the residue image, then, indicate that there may be a star at that point whose brightness has changed over the time period between when the reference image was taken, and when the new image was taken. This point then becomes a transient candidate.

An illustration of the process is shown in the figure below.



*Figure 1: This is an illustration of the subtraction technique. The Residue Image is created by a method of subtracting the New Image from the Reference Image.*

Computer scripts are run each night to perform this image subtraction technique on all the images taken that night. The resulting transient candidates are then displayed on the website [http://rotse.net/supernova\\_search/](http://rotse.net/supernova_search/). These candidates are then manually checked daily and analyzed to determine whether or not they truly are transients, and if so, what type they are.

## Photometry

The first and most fundamental in the series of ROTSE tools for analyzing image data is the photometry program developed by Quimby. This program measures the magnitude of a point source in an array of FITS files. Since FITS files also provide the time when the image was taken, the program can make a plot and data file of the magnitude measured vs time. This is precisely the lightcurve of the point source, and the shape of the lightcurve can be analyzed to determine whether or not the point source represents a transient such as a variable star. As an example, the lightcurve of VSX J115452.8+411923, a typical periodic variable star discovered by ROTSE, is displayed in the figure below.

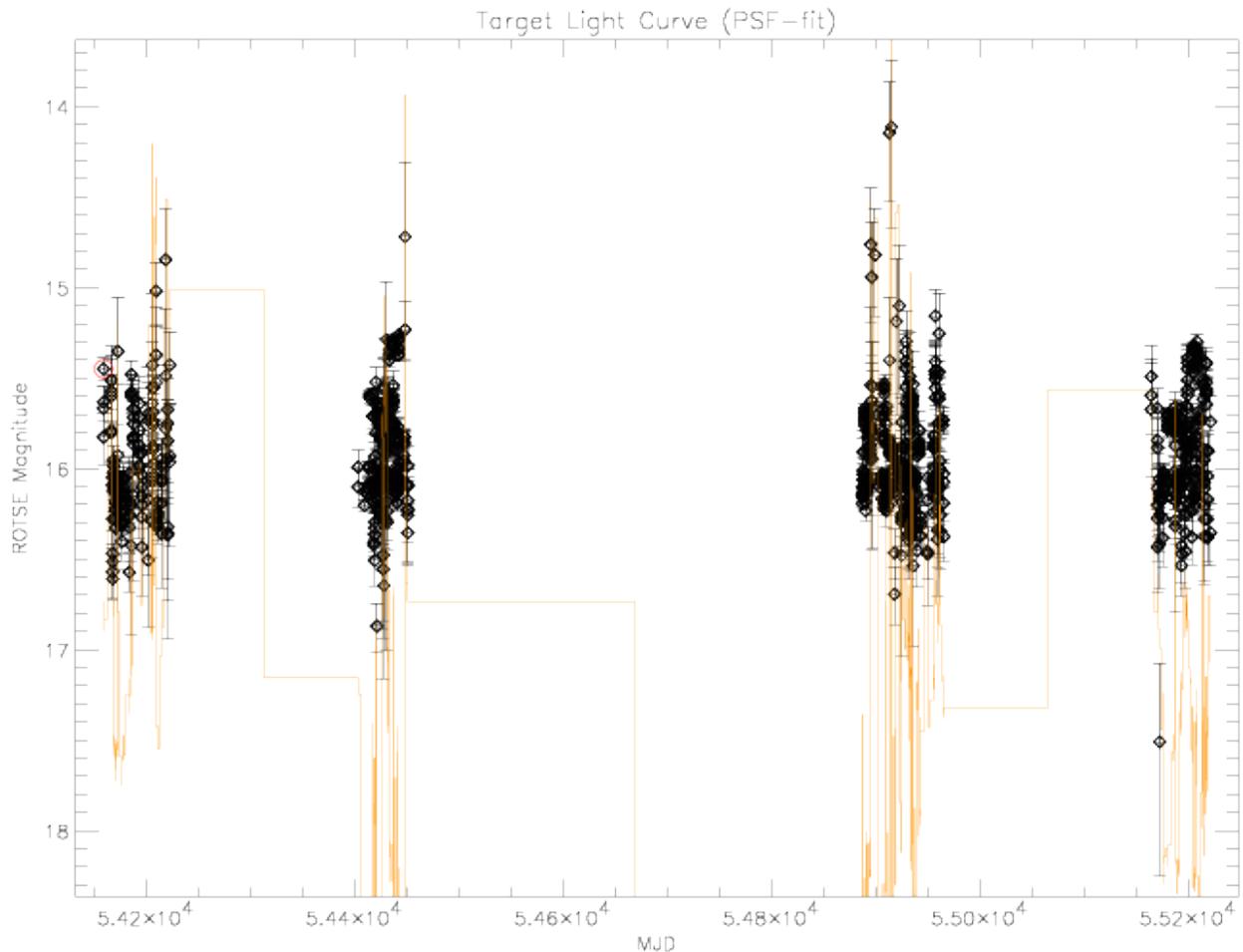


Figure 2: This is the lightcurve produced by the photometry program used at ROTSE of the variable star VSX J115452.8+411923.

While it is apparent from the lightcurve that the magnitude of VSX J115452.8+411923 varies by a large amount ( $\sim 1$  magnitude), it is extremely difficult to analyze the shape of the lightcurve by eye. The reason for this is the irregularity of the sampled data. During some seasons, ROTSE telescopes may not be able to take images of certain parts of the sky, so long periods of time may have no data. By focusing on only sections of the lightcurve which do not have long gaps of data, it begins to be possible to observe a clear shape. However, even after narrowing in on a lightcurve, a clear period

and waveform are still often not readily observed. One of the reasons why the waveform of a star is not always apparent from the lightcurve alone is that for stars with periods of the order of days a gap of data of a single 24 hour day is significant. In order to quantitatively determine the period and waveform of the star, more rigorous analysis is required.

### *Periodicity Detection*

In order to analyze data of a potential periodic variable star which may be sampled too poorly to directly determine a period and waveform, a periodicity detecting program was developed by Akerlof. The way it works is described below.

Assuming the period of the star is  $T_p$ , the program will calculate the magnitude vs phase (from 0 to 1) from a data set of magnitude vs time. The phase of a point in terms of its time,  $T$ , and the period,  $T_p$ , is given by the following formula.

$$\phi = \frac{T \bmod T_p}{T_p}$$

The resulting magnitude vs phase data is then fitted, and  $\chi^2$  calculated, with a cubic spline function with 8 segments. A spline is a smooth, piecewise-defined, polynomial function. (Its origins lie in bending pieces of wood around fixed points.) So in this case, our function is a piecewise function of 8 cubic polynomials evenly spaced over the range 0 to 1, and tied together at each interface such that the function is continuous, its derivative is continuous, and its second derivative is continuous. This results in a piecewise function with 11 degrees of freedom – 3 for the first cubic polynomial, and 1 for each other cubic polynomial. With this many degrees of freedom, it is possible that the function resulting in the best  $\chi^2$  contains several maxima and minima, resulting in many bumps which is not an accurate representation of the waveform of most variable stars. As such, the function resulting in the best  $\chi^2$  is not necessarily the best, and the fitness evaluation of the spline function is calculated using both the  $\chi^2$  and the number of bumps present as parameters.

The program repeats this process systematically over a range of periods, and the resulting data with the best fit is plotted with its period and amplitude. Two full periods are plotted in order to help visualize the periodicity.

As an example, the phase plot produced by this program from the lightcurve data of VSX J115452.8+411923 is shown in the figure below.

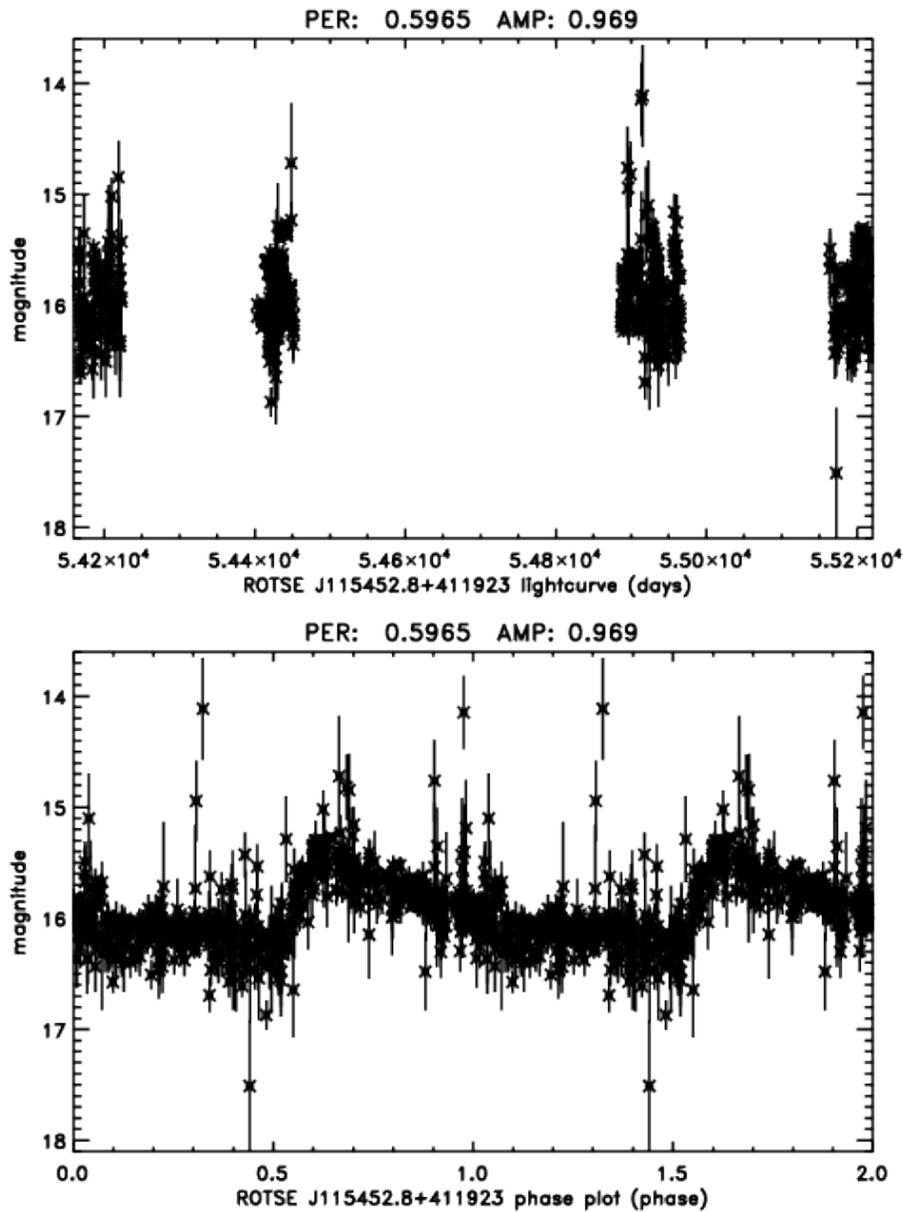


Figure 3: This is the lightcurve (top) and phase plot (bottom) produced by the phase-finding program used at ROTSE of the variable star VSX J115452.8+411923.

From this figure, the period and waveform of VSX J115452.8+411923 is crystal clear. For most stars, making a phase plot like this is the extent of the analysis required to determine periodicity.

However, there are some cases when a clear period and waveform is not found for a star which otherwise shows signs of being a real variable star. For example, a star may vary greatly in brightness, even by a magnitude or two, but fail to produce a clear waveform when analyzed. In most of these cases, the star may be a variable star, but not a periodic one. An example of the phase plot produced for such a star is displayed in the figure below.

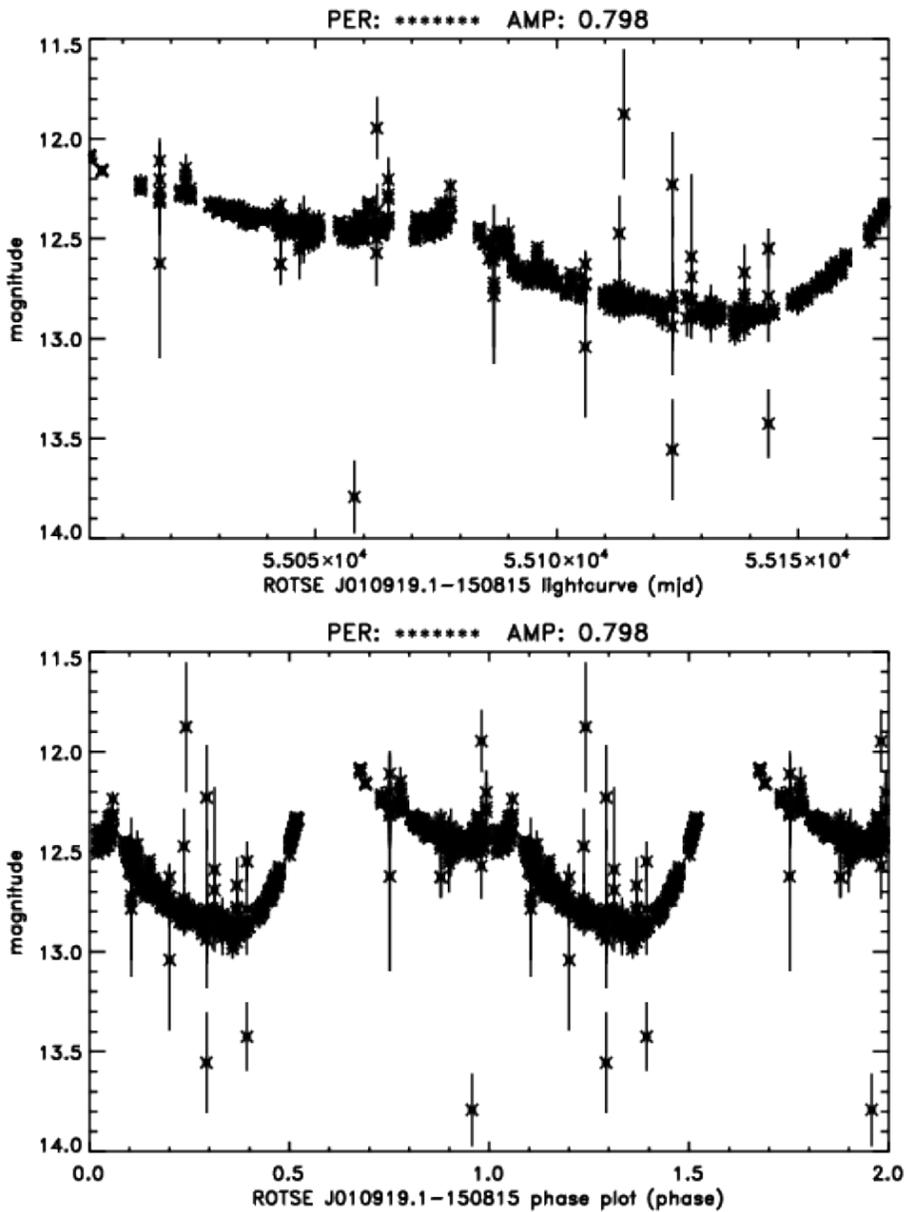


Figure 4: This is the lightcurve (top) and phase plot (bottom) of the variable star VSX J010919.1-150815.

From the lightcurve alone, it is apparent that this star is indeed a variable star, with a greatly varying magnitude and a clear shape. However, because the star is aperiodic, the phase plot was unable to find a real period of the star, and ended up using the whole lightcurve as a cycle. Thus, as expected, a phase plot is not useful at all in analyzing aperiodic variable stars.

### Results

This method of manually sorting through transient candidates in search of variable stars has been used extensively at ROTSE, and many new variable stars have been discovered in this way. When I began discovering many variable stars in this way, I decided to catalogue my discoveries in the VSX,

and a webpage for ROTSE variable star discoveries was also started to display information on these catalogued stars. The webpage can be found at <http://rotse.net/rvvp/>.

Over the year 2011, my co-worker Anastasiya Ramadan and I discovered over 40 variable stars using this technique. Interestingly, almost all of the variable stars were of type RRab, with only a handful of RRC, binary, and irregular variable stars. This fact is important in that the automated procedure to detect variable stars that I have developed was able to find many more stars of different types, and a much larger proportion of binary stars compared to RRab stars. A plot showing the different types of variable stars discovered in 2011 is shown in the figure below.

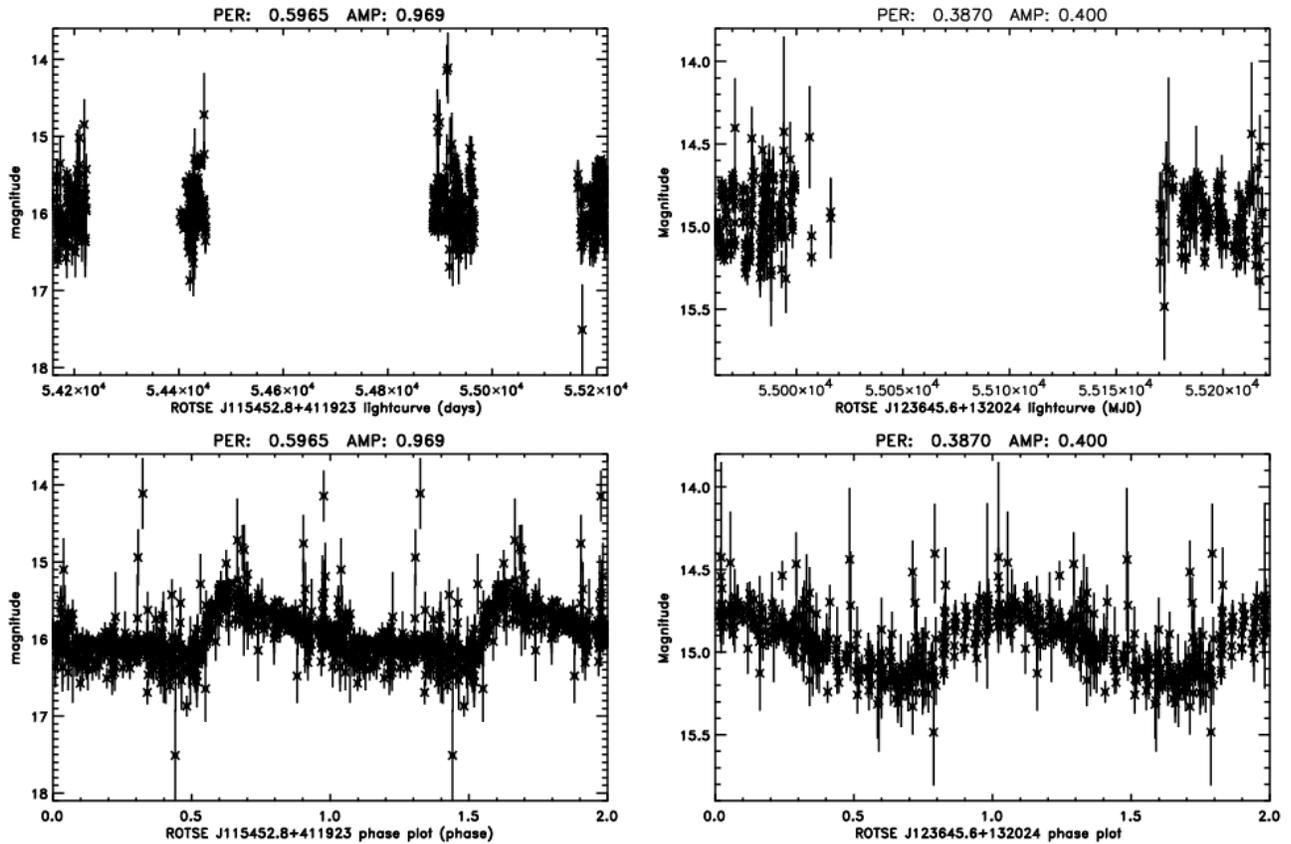


Figure 5: These are the lightcurves (top) and phase plots (bottom) of the RRab VSX J115452.8+411923 and RRc VSX J123645.6+132024. They were used later in the automated procedure as the RRab and RRc reference stars.

These stars are representative of the types of variables discovered using the procedure outlined above. They were in fact used as reference types in the automated procedure outlined in detail below.

### Problems

This process of discovering and analyzing variable stars works well in general. However, after going through this procedure countless times, a few simplifications and enhancements to the process became apparent.

For one, sorting through potential transient candidates for variable stars manually takes a lot of time and probably does not catch each variable star in a ROTSE field either. Next, for each variable star

candidate, the same series of programs are run to analyze the star, requiring more procedural manual input. In order to streamline these processes, a series of scripts were written which would analyze every star in a ROTSE field, providing the user with organized plots and data necessary to confirm a variable star discovery.

Another observation made from doing this procedure many times was the phase plots of successfully identified variable stars showed very distinctive waveforms corresponding to their type. This is definitely to be expected, but to go one step further with this automated system, it occurred to me that by analyzing the waveform of each star quantitatively, a program should be able to classify variable stars as well.

Thus I set out to modify this procedure of detecting variable stars with the goal that a computer automatically detect and classify as many stars in ROTSE fields as possible.

## **Automated Procedure**

### *Waveform Analysis*

Before the whole process of discovering variable stars could be streamlined and automated, one more important tool needed to be made. This was the program for classifying variable stars based on their waveform. Fortunately, the periodicity detection program outputs all the necessary information to define the waveform of a star. Namely, it outputs 256 points, sampled uniformly over one period from the spline fitting function. Using numerical techniques, the waveform data was normalized such that the mean was 0 and such that the integral over one period of the waveform squared equalled 1, as follows.

$$\int_0^1 (\Psi(x))^2 dx = 1$$

Using the normalized data defining the waveform of a star, the waveforms of stars could be directly compared with each other using a cross-correlation function. The cross-correlation of two waveforms is a measure of the similarity of the two waveforms, and is given by the following expression.

$$CC(\phi) = \int_0^1 \Psi_1(x) \Psi_2(x - \phi) dx$$

Since the waveforms  $\Psi_1$  and  $\Psi_2$  are normalized, the maximum value of CC is 1, and the minimum value is 0. If the two waveforms are very similar, CC will tend towards 1, while if they are very different, CC will tend towards 0. Also note that CC is a function of the phase offset,  $\phi$ . This makes sense, as the cross-correlation of two waveforms, even if they are identical, depends on how “in-phase” they are with each other.

Since it was observed that stars of like types had very similar waveforms, it seemed fitting to see if doing a cross-correlation of stars of the same type would yield significantly higher values than doing a cross-correlation of stars of different types. If this was the case, then the idea was to take the cross-correlations of a new star with a set of reference types, and if the cross-correlation value was high enough between the new star and one of the reference types, then the new star would be automatically

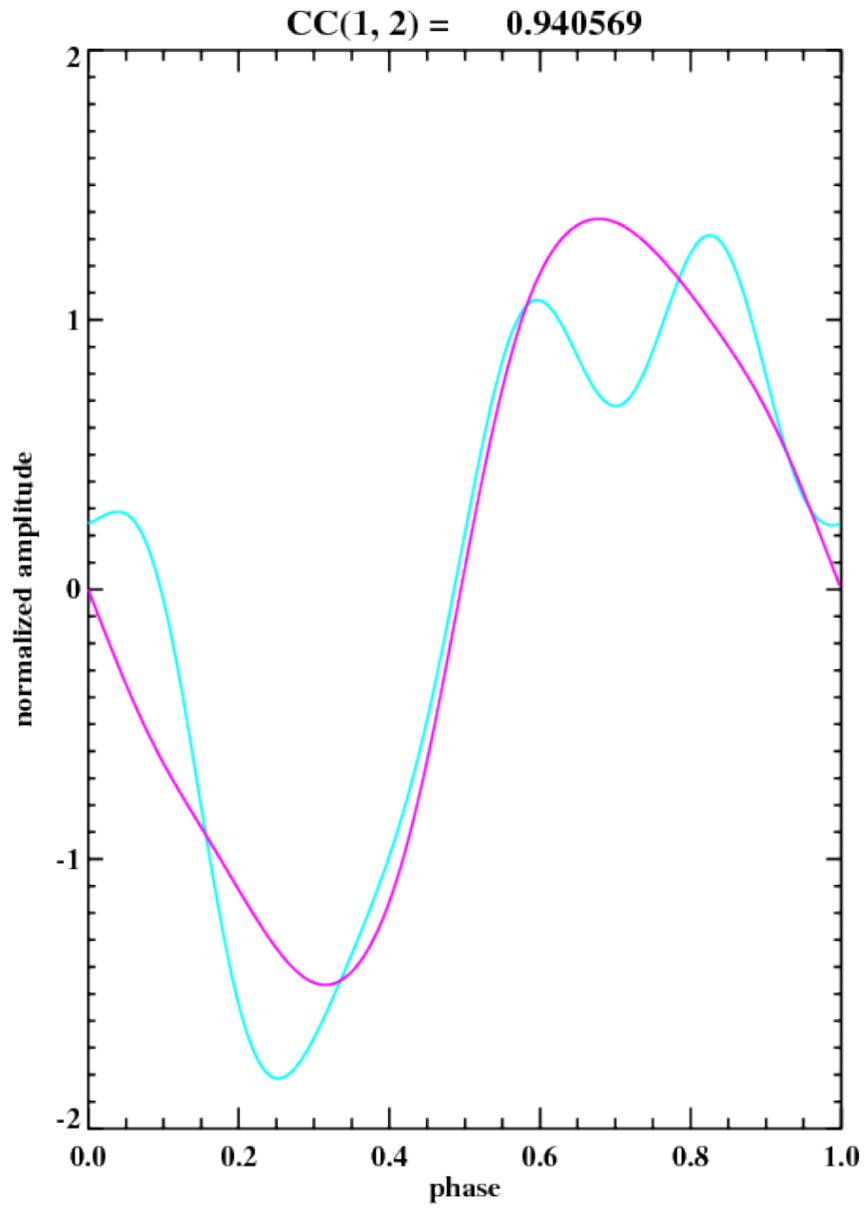
classified under that type.

I wrote a program in IDL, the primary programming language used at ROTSE, to do these integrals and calculate the best cross-correlation of the waveforms of two stars using standard numerical algorithms. The program also made a plot showing the two waveforms at the offset resulting in the best cross-correlation value. Before implementing this program, along with the photometry and phase-finding programs, in the automated system for detecting variable stars, I tested it with all of the variable stars discovered by myself and Anastasiya Ramadan during 2011. This is summarized by a portion of the results displayed in the table below.

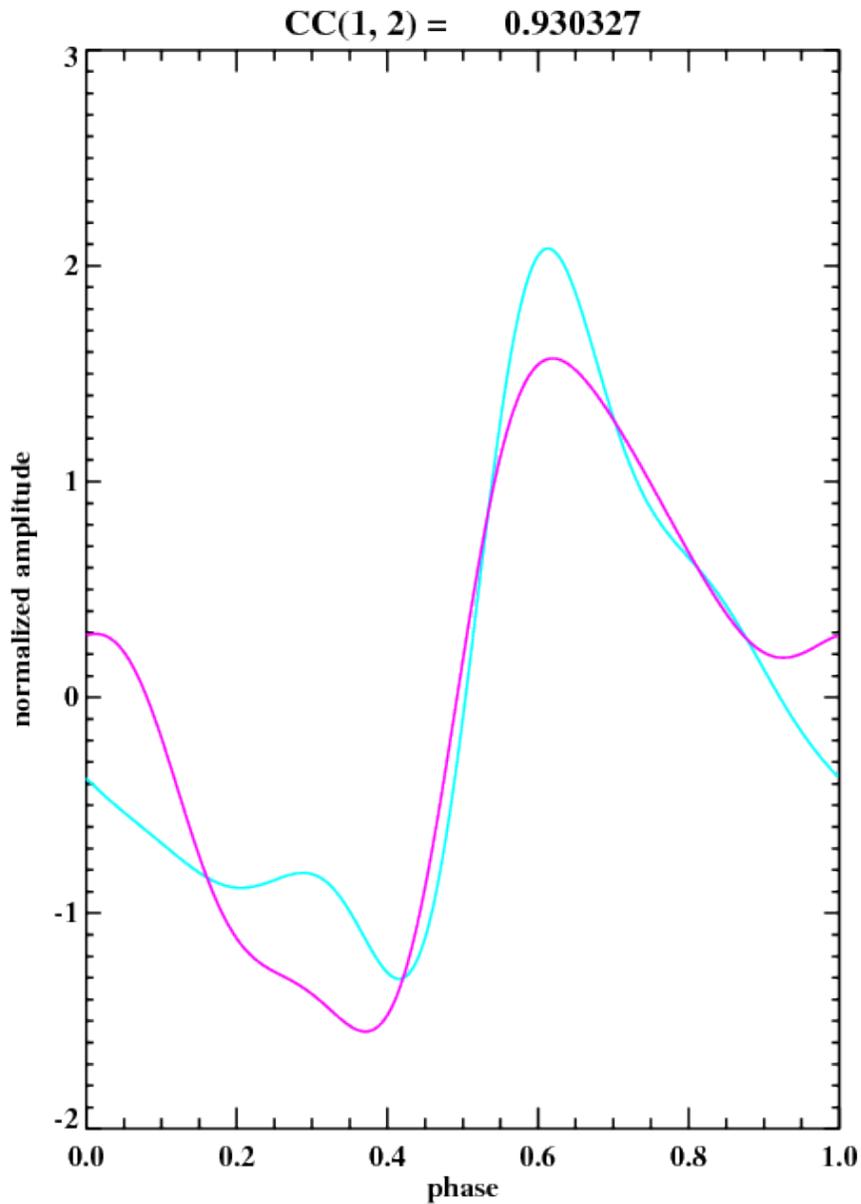
	VSX J203814.6- 242234 (RRab)	VSX J203421.1- 243601 (RRab)	VSX J133221.7- 32518 (RRab)	VSX J125240.5- 132441 (RRab)	VSX J115452.8+4 11923 (RRab)	VSX J123645.5+1 32022 (RRc)	VSX J010919.1- 150815 (L)	Binary 1	Binary 2
VSX J203814.6- 242234 (RRab)	1.00000	0.991445	0.991300	0.983639	0.974875	0.935374	0.928801	0.161642	-0.927332
VSX J203421.1- 243601 (RRab)	0.991445	1.00000	0.977307	0.995701	0.994519	0.910195	0.930327	0.204117	-0.884546
VSX J133221.7- 32518 (RRab)	0.991300	0.977307	1.00000	0.967894	0.957148	0.920923	0.905065	0.0795193	-0.938824
VSX J125240.5- 132441 (RRab)	0.983639	0.995701	0.967894	1.00000	0.995068	0.878754	0.913145	0.245315	-0.877747
VSX J115452.8+4 11923 (RRab)	0.974875	0.994519	0.957148	0.995068	1.00000	0.886386	0.918799	0.256548	-0.850158
VSX J123645.5+1 32022 (RRc)	0.935374	0.910195	0.920923	0.878754	0.886386	1.00000	0.953081	0.159708	0.940569
VSX J010919.1- 150815 (L)	0.928801	0.930327	0.905065	0.913145	0.918799	0.953081	1.00000	0.159708	0.911545
Binary 1	0.161642	0.204117	0.0795193	0.245315	0.256548	0.0543878	0.159708	1.00000	0.205818
Binary 2	-0.927332	-0.884546	-0.938824	-0.877747	-0.850158	0.940569	0.911545	0.205818	1.00000

*Table 1: This table contains the cross-correlation values between 36 pairs of variable stars, a subset of the total preliminary cross-correlation test.*

From the cross-correlation data obtained, it is clear that what constitutes a “large” cross-correlation value must be carefully considered. Many stars of different types, such as the RRc VSX J123645.5+132022 and Binary 2, have cross-correlations that seem to be quite close to 1. To see how this happens, it is useful to look at the cross-correlation plot. The cross-correlation plots for 2 pairs of stars of different types are displayed in the figures below.



*Figure 6: This is the phase plot of the RRc VSX J123645.5+132022 (magenta) vs Binary 2 (cyan) showing their cross-correlation value.*



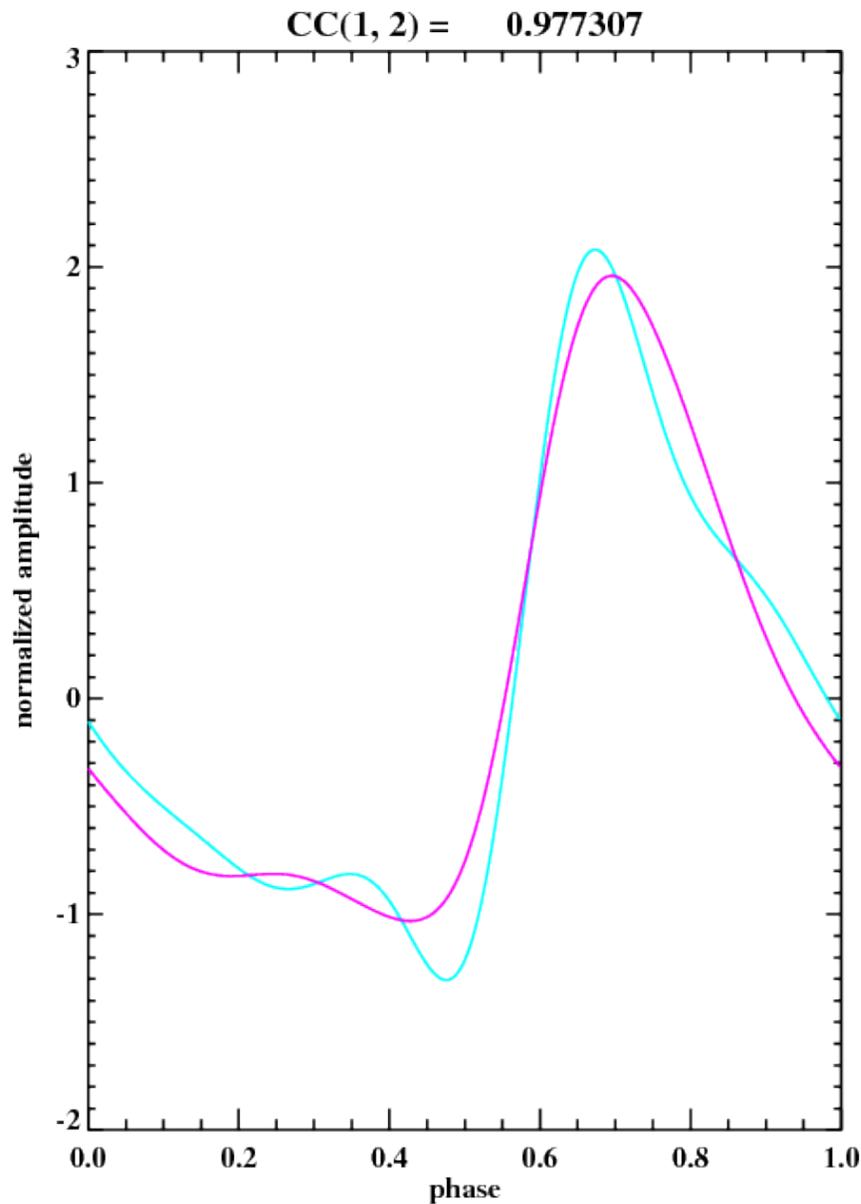
*Figure 7: This is the phase plot of the irregular variable star VSX J010919.1-150815 (magenta) vs the RRab VSX J203421.1-243601 (cyan) showing their cross-correlation value.*

In the case of the RRc VSX J123645.5+132022 vs Binary 2 cross-correlation plot, it is clear that the two stars are of different types by observing the specific shapes of their waveforms, but the general shapes of their waveforms are quite similar, resulting in a “large” cross-correlation of 0.940569.

In the plot shown in Figure 7, the irregular variable star VSX J010919.1-150815 was used. However, because the variable star is irregular, the waveform produced by the phase-finding program is very inaccurate, and it is only because of the similarity of the fitting functions that the phase-finding program uses that results in the similar waveform, and thus “large” cross-correlation.

While these “large” cross-correlation values may seem to present a problem, the cross-correlation values of like-type variable stars are even higher, as can be seen by the multiple

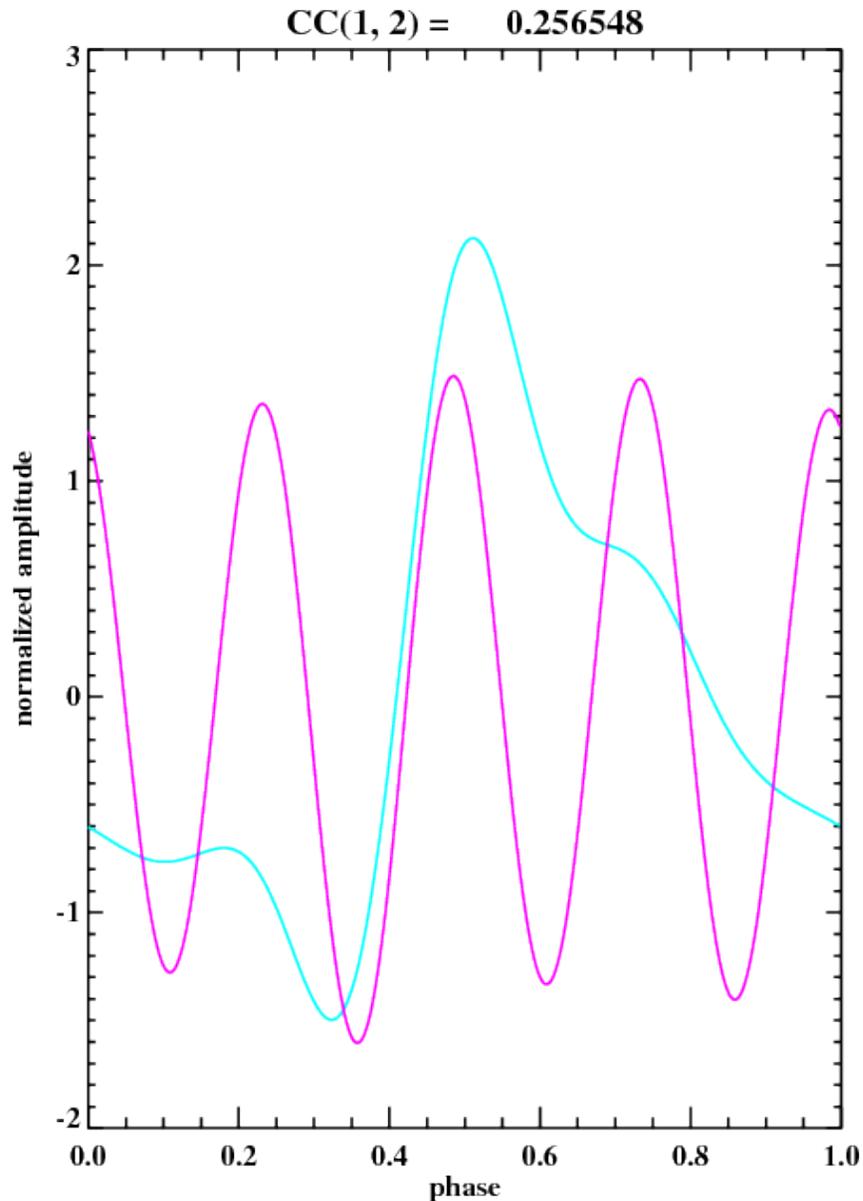
comparisons of RRab stars. In the data presented above, the range of these cross-correlation values is 0.957148-0.995068, and most of the cross-correlation values tend to be 0.99x. The cross-correlation plot of a typical RRab star comparison is shown in the figure below.



*Figure 8: This is the phase plot of the RRab VSX J133221.7-32518 (magenta) vs the RRab VSX J203421.1-243601 (cyan) showing their cross-correlation value.*

Compared to the RRc VSX J123645.5+132022 vs Binary 2 cross-correlation plot from before, here both the general structure and smaller features of the waveform are similar, and it is absolutely clear that the two stars are of the same type. It should be noted that the lowest cross-correlation values between like-types, and the highest cross-correlation values of different types do occasionally overlap at values around ~0.90-0.95, this seldom occurs, and can be quickly sorted through manually.

As a last example, variable stars of different types with completely different waveforms do indeed result in very low cross-correlation values. A plot of such a case is shown in the figure below.



*Figure 9: This is the phase plot of Binary 1 (magenta) vs the RRab VSX J115452.8+411923 (cyan) showing their cross-correlation value.*

All-in-all, the cross-correlation values do reflect the similarity of the waveforms of variable stars in the intended manor. Stars whose waveforms are drastically different result in very low cross-correlation values, and stars of the same type result in cross-correlation values very close to 1. The stars of different type which have somewhat similar waveforms do tend to have large cross-correlation values, but since their values are usually easily distinguishable from values resulting from stars of the same type, they should not present a major problem.

Considering these cross-correlation values of this preliminary test, I decided to modify the

program to compare the waveform of a star with those of reference types, and to automatically classify the star to be the type of a reference star which resulted in the best cross-correlation value. This value was then later used in sorting variable star candidates, with the cross-correlation values higher than 0.90 at the very top of the list. The reference stars chosen were RRab VSX J115452.8+411923, the RRc VSX J123645.6+132024, Binary 1, and Binary 2, representing all the different types of variable stars discovered by Anastasiya Ramadan and myself in the year 2011. After using the automated procedure, however, new types of variable stars, or stars with significantly different and distinct lightcurves, were discovered and additions to the reference star list were made.

Finally, with all the essential programs written, they could be scripted together, using IDL, in order to streamline the process of using them to analyze roughly each star in a ROTSE field.

### *Scripts*

The first step in the process of analyzing every star in a ROTSE field was to obtain a list of the RA and DEC coordinates of all of the stars. This was done by querying the USNO B 1.0 catalogue, and filtering out stars with apparent magnitudes below 17.5. The number of stars per ROTSE field found in this way was usually between 2 and 3 thousand. While the limiting magnitude of the ROTSE telescopes is often dimmer than 17.5, and several variable stars have been detected whose mean magnitudes are around 17.5-18.0, running the photometry program on such stars often ends up in a crash at a particular image. If manually using the photometry program, the bad images can often just be discarded, requiring the photometry program to run again. While this usually presents no problem when analyzing one star, the amount of both human and computer time this would take when analyzing a few thousand stars is entirely impractical. Thus, an apparent magnitude cutoff of 17.5 was implemented.

This list of stars were then fed into the photometry program, phase-finding program, and waveform-analyzing program in a straightforward manor, creating data files containing information about the stars' periods, amplitudes, cross-correlation values, and most likely type, as well as phase plots and cross-correlation plots. This is all the information needed to verify a variable star discovery and classification.

### *Candidate Sorting*

If the end-user were really to look through all of the thousands of plots and data files that the programs and scripts described above create, not much time or effort would be saved in detecting and classifying variable stars. However, all of the information necessary to check if a candidate is real or not, and of what type, is technically given by those programs, so the last step in this automated process was to simply display this information to the end-user in the most quick and convenient way. An html generating script was created in order to make webpages for each ROTSE field, containing the information on each star, listed in an order with the most likely variable star candidates first.

Essentially 3 characteristics of a star were used to determine this likeliness – the period, the amplitude, and the cross-correlation. Early on, it was discovered that the phase-finding program often found a period of  $0.997x$  days (about 1 sidereal day) for many non-variable stars. Though sometimes the waveform of these stars even looked convincing, a period of exactly 1 sidereal day is surely just an artifact of the frequency at which the ROTSE telescopes take images of the sky. The amplitudes found for these stars were small as well, around  $\sim 0.1$  magnitude. After sorting though hundreds of these, it was clear that none of them were real variable stars, so the first thing that the sorting algorithm in the html-generating script did was put all stars with periods of  $0.997x$  days at the end of the list. Also, occasionally the phase-finding program would not find a period, outputting \*\*\*\* instead. These stars,

too, were put at the end of the list.

The next step was to look at the amplitude that the phase-finding program found. No matter what the waveform, or period, of a star, if the amplitude of its variation is significantly large, there is a good chance that it is a variable star. Even stars which appear to have no period, but have a large amplitude of variation are typically variable stars – aperiodic ones. Thus, stars with amplitudes over 0.4 magnitudes were put at the top of the list (which could include those previously sorted to the bottom of the list), and organized in order of their amplitudes.

The last step was to look at the cross-correlation of each star with the reference types. As mentioned earlier on, if a star has a cross-correlation larger than 0.90 with a reference type, the chances are very good that the star belongs to that type. Thus, all stars with cross-correlations of 0.95 with any of the reference types were put at the very top of the list, and sorted by their cross-correlation values with the highest ones first. However, this time stars with periods of 0.997x days, or no period at all, were not included in the sorting.

With the stars sorted, an html file was created which included a list of the amplitudes, periods, cross-correlation and most likely type, and small versions of the phase-plots and cross-correlation plots of each star – all the information required to confirm a discovery.

## **Results**

Running a ROTSE field through these scripts takes approximately 2~3 days, and of the 2000~3000 stars analyzed, only 1~10 variable stars are typically discovered. However, the criteria described above for sorting the stars in order of their likeliness to be variable stars has worked effectively as intended. The 1~10 variable stars found in a particular field usually all appear at the top of the list.

This automated procedure has been used on several ROTSE fields. All of the programs appear to be working exactly as intended, this approach has proven to be very worthwhile

At first using this procedure was somewhat slow as the programs were constantly being tweaked, and solutions to various slowdowns and crashes were being worked out. Once the programs were fleshed out, however, some ~1000 stars could be analyzed a day, and among them usually ~1-2 turned out to be variables. These variable stars were of various types, and not all RRab variable stars, as was the case nearly every time when doing the previous procedure. A table summarizing some of the ROTSE-fields analyzed and the variable stars found within them is displayed below.

ROTSE field	stars analyzed	RRab	RRc	binary	other
sks 0258+0602	2966	0	1	1	1
sks 0326+4045	15561	1	1	31	0

*Table 2: This is a table showing the number of stars analyzed and the number of various types of variable stars detected for 2 ROTSE fields.*

As can be seen from the table, a large proportion of the variable stars discovered were binaries. Also, most of these binaries had lightcurves much different than the handful of binaries discovered using the previous procedure. (The lightcurves were, however, very distinctive of binary stars nonetheless). An example of such a lightcurve is displayed in the figure below.

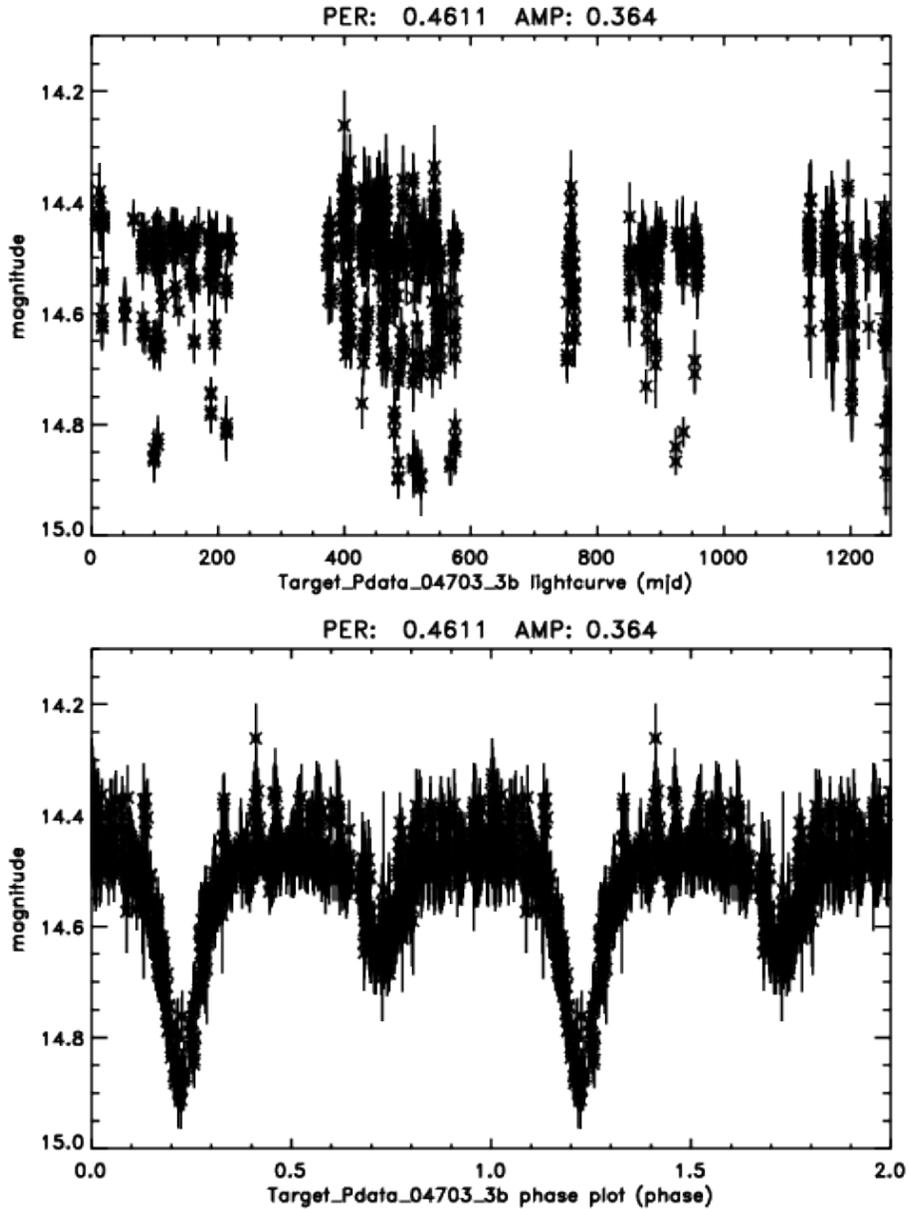


Figure 10: This is the lightcurve (top) and phase plot (bottom) of 04703, an eclipsing binary star.

Since variable stars with these types of lightcurves had not previously been analyzed in our data, they did not have high cross-correlation values with any of the reference types. Also, many of their amplitudes were smaller than the cut-off 0.4 magnitude used for sorting. Thus, several of these variable stars fell through the cracks in the initial sorting process. However, because they tended to have high  $\chi^2$  values calculated by the phase-finding program, some of them did appear near the top of the list of viable variable stars. Still others were manually discovered when taking the time to sort through all of the analyzed stars in the ROTSE field.

At first, the great variation in the shape of lightcurves for binary stars appeared to present a problem in that they could not all be identified by doing a cross-correlation with the lightcurve of a single variable star. In fact, the automatic classification of such binary stars based on a somewhat

different and specific approach has been studied by Jonathan Devor in the DEBiL project. However, for this project, after finding a variable star of a new type, or one with a different and distinctive lightcurve like the binaries shown above, it was simply added to the list of reference stars in order to more easily and automatically detect stars of the same type in the future. After several stars with new distinctive lightcurves were added to the reference list, a study of the effectiveness of the waveform-analyzing program similar to that done on the previously discovered variable stars was carried out. Specifically, 10 variable stars ranging all of the new types, and some old types, from ROTSE field sks0326+4045 were cross-correlated with each other. The phase plots for these variable stars is displayed in the figure below.

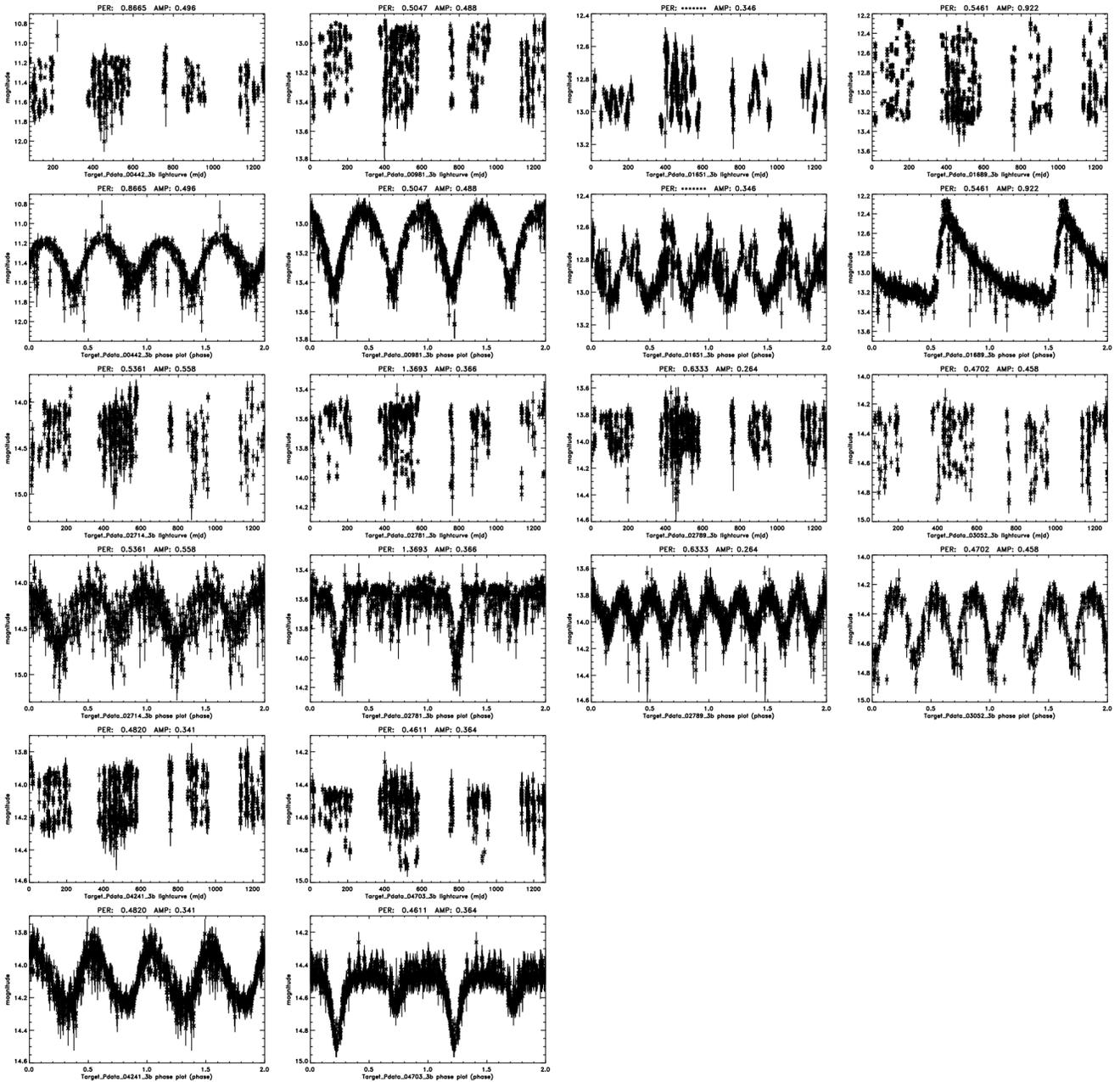


Figure 11: These are the lightcurves and phase plots for the 10 variable stars from ROTSE field sks0326+4045 used in a new cross-correlation study. From left to right, top to bottom, they are sks0326+4045\_00442, sks0326+4045\_00981, sks0326+4045\_01651, sks0326+4045\_01689, sks0326+4045\_02714, sks0326+4045\_02781, sks0326+4045\_02789, sks0326+4045\_03052, sks0326+4045\_04241, and sks0326+4045\_04703.

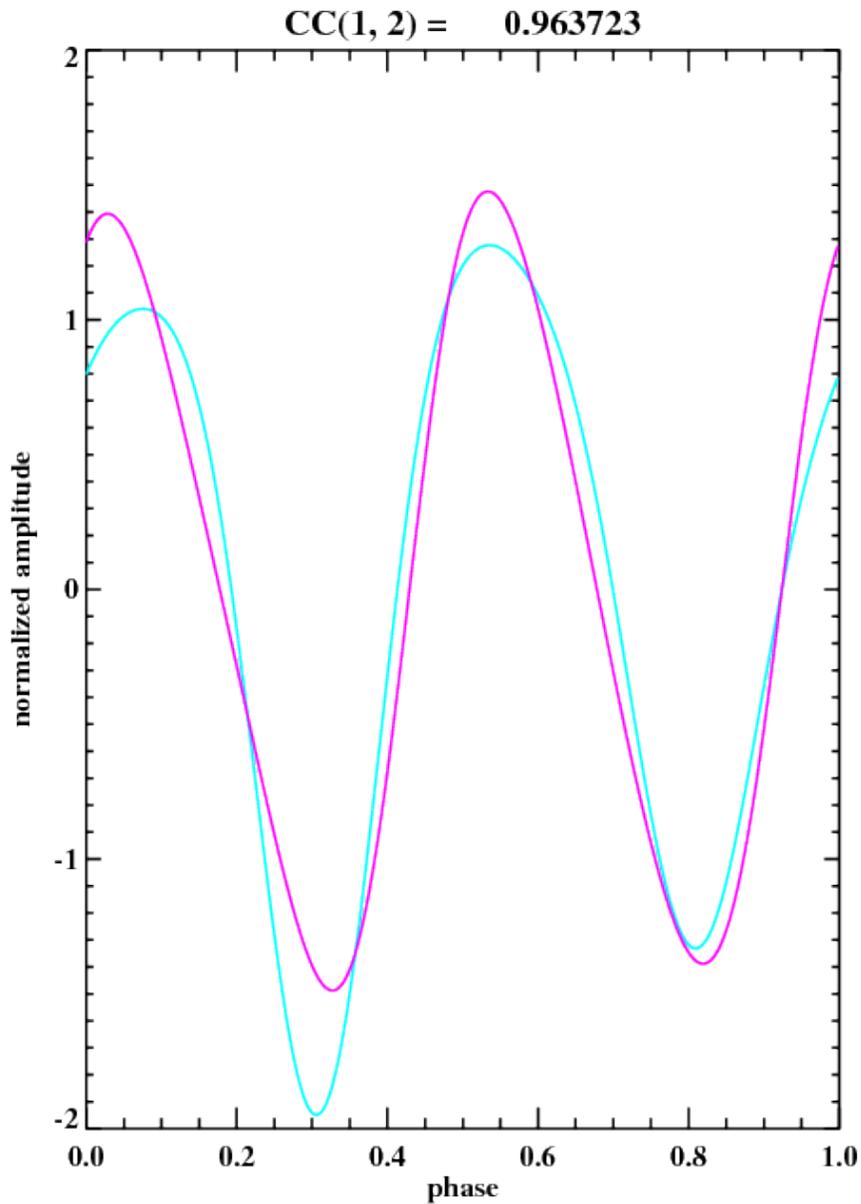
While most of the variable stars above are clearly binary stars, several of them have much different, yet still distinctive, waveforms than the binaries found before. Based on both simply the visible waveforms of the stars above, and more specifically on the cross-correlation values between them displayed below, I have labeled the type of binaries with waveforms similar to sks0326+4045\_00442 as Binary 3, those with waveforms similar to sks0326+4045\_04703 as Binary 4,

those with waveforms similar to sks0326+4045\_03052 as Binary 5, and those with waveforms similar to sks0326+4045\_02789 as Binary 6. The cross-correlation results of this study are summarized in the table below.

	sks0326+4045_00442 (Binary 3)	sks0326+4045_00981 (Binary 3)	sks0326+4045_01651 (Binary 5)	sks0326+4045_01689 (RRab)	sks0326+4045_02714 (Binary 3)	sks0326+4045_02781 (Binary 4)	sks0326+4045_02789 (Binary 6)	sks0326+4045_03052 (Binary 5)	sks0326+4045_04241 (Binary 3)	sks0326+4045_04703 (Binary 4)
sks0326+4045_00442 (Binary 3)	1	0.987859	0.283446	0.494334	0.954262	0.793834	-0.233401	-0.226197	0.963723	0.900516
sks0326+4045_00981 (Binary 3)	0.987859	1	0.201412	-0.440943	0.947087	0.762356	-0.301400	0.163264	-0.960775	0.896420
sks0326+4045_01651 (Binary 5)	0.283446	0.201412	1	0.429734	0.193999	-0.535744	0.0910613	-0.974881	-0.184927	-0.391232
sks0326+4045_01689 (RRab)	0.494334	-0.440943	0.429734	1	-0.642220	-0.798715	0.210164	-0.395981	0.432851	-0.698927
sks0326+4045_02714 (Binary 3)	0.954262	0.947087	0.193999	-0.642220	1	0.868831	-0.278073	-0.172997	0.913643	0.940789
sks0326+4045_02781 (Binary 4)	0.793834	0.762356	-0.535744	-0.798715	0.868831	1	-0.418042	0.490624	0.678662	0.957909
sks0326+4045_02789 (Binary 6)	-0.233401	-0.301400	0.0910613	0.210164	-0.278073	-0.418042	1	0.0519799	0.196560	-0.487584
sks0326+4045_03052 (Binary 5)	-0.226197	0.163264	-0.974881	-0.395981	-0.172997	0.490624	0.0519799	1	-0.157587	0.352044
sks0326+4045_04241 (Binary 3)	0.963723	-0.960775	-0.184927	0.432851	0.913643	0.678662	0.196560	-0.157587	1	-0.808595
sks0326+4045_04703 (Binary 4)	0.900516	0.896420	-0.391232	-0.698927	0.940789	0.957909	-0.487584	0.352044	-0.808595	1

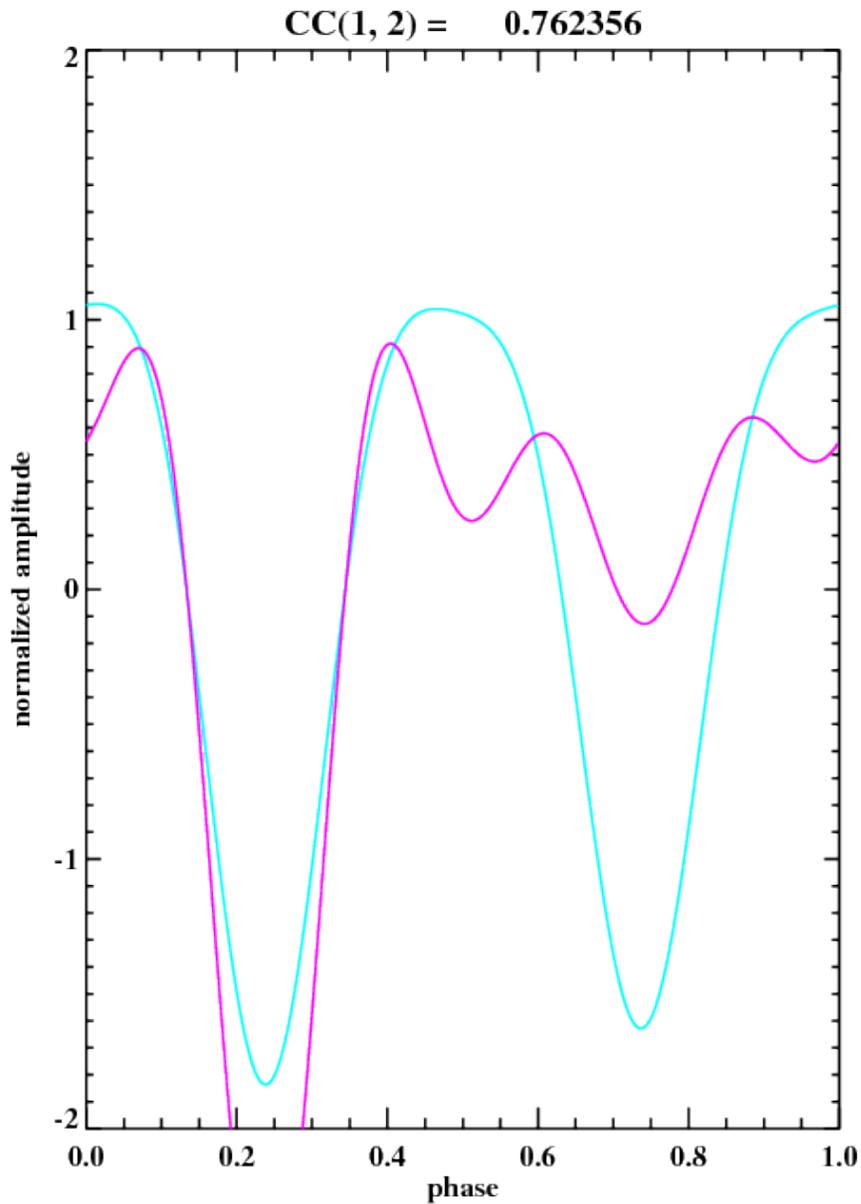
*Table 3: This table contains the cross-correlation values of 45 pairs of stars used in analyzing the new types of variable stars in ROTSE field sks0326+4045.*

As can be seen by the cross-correlations of these types of stars, variable stars of like and different types show the same trends seen in the first cross-correlation study. Particular analyses on a few of these cross-correlations is given below.



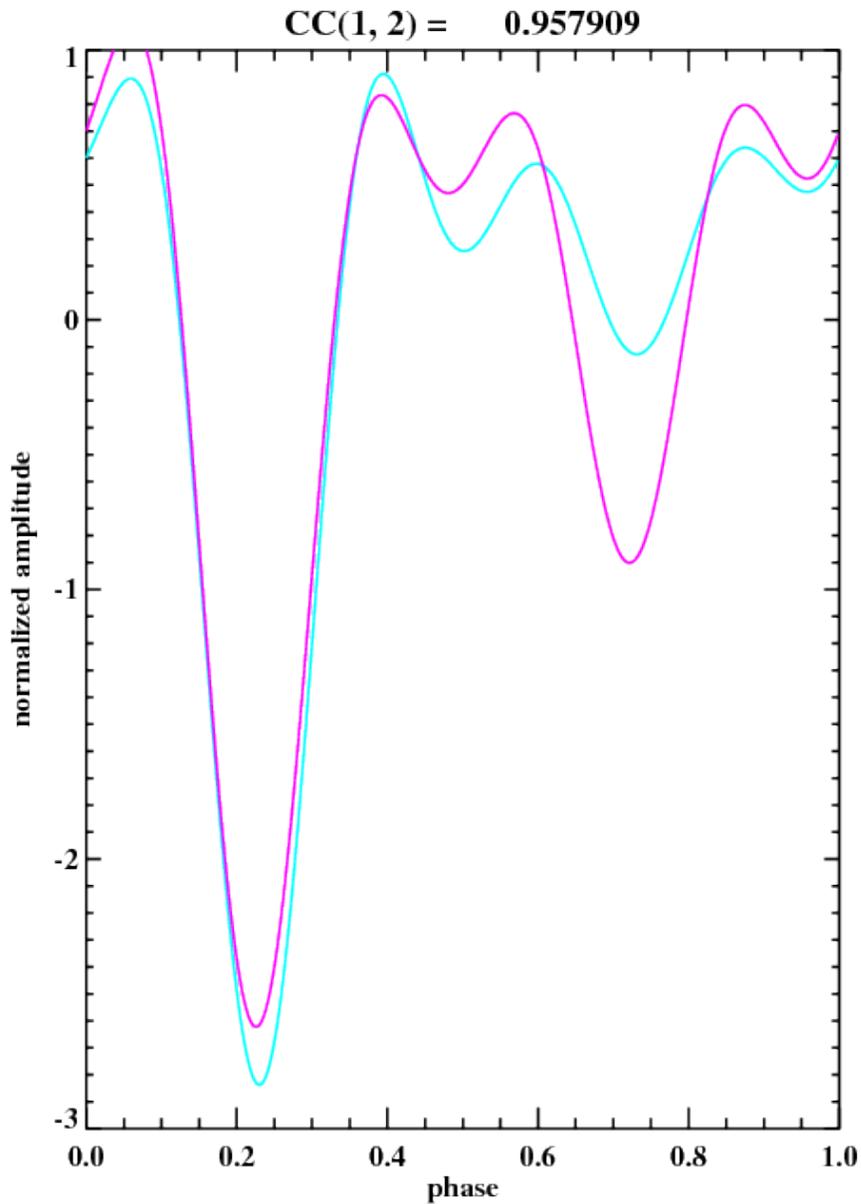
*Figure 12: This is the phase plot of the binary stars *sks0326+4045\_0042* (magenta) vs *sks0326+4045\_04241* (cyan) showing their cross-correlation value.*

Here we see the modified phase plot of the two binary stars *sks0326+4045\_00442* and *sks0326+4045\_04241*, and their cross-correlation value of 0.963723. From the original phase plots, it appeared visually that the stars were of the same type, and this very high cross-correlation value confirms it. If either of these stars is picked as a reference star (as *sks0326+4045\_00442* was), similar stars should be easily automatically classified.



*Figure 13: This is the phase plot of binary stars *sks0326+4045\_00981* (cyan) vs *sks0326+4045\_02781* (magenta) showing their cross-correlation value.*

Here we see the modified phase plots of *sks0326+4045\_00981* and *sks0326+4045\_02781*, including their cross-correlation values of 0.762356. This value is not terribly low, and these two stars have been confirmed to both be types of binary stars, but their waveforms differ significantly, both visually, and in terms of the cross-correlation value. Since the cutoff for automatic classification was a cross-correlation value of 0.90, they would not be automatically identified as the same type of star.



*Figure 14: This is the phase plot of the eclipsing binaries `sks0326+4045_02781` (cyan) vs `sks0326+4045_04703` (magenta) showing their cross-correlation value.*

Here is the modified phase plots of `sks0326+4045_02781` and `sks0326+4045_04703`, showing their cross-correlation value of 0.957909. This cross-correlation value suggests that the stars are similar enough to be of the same type, and a quick look at the original phase plots confirms this. These two stars are eclipsing binaries where the dips in magnitude occur when one star in the binary system eclipses the other. `sks0326+4045_04703` was chosen to be the reference for detecting these types of stars.

Another major benefit of this automated procedure is its use in finding stars of irregular type. Even though the phase-finding program and waveform-analyzing program require that a star be periodic in order to effectively analyze the star, stars which were not periodic but still appeared to have

a large amplitude of variation (arbitrarily cut-off at 0.4 magnitudes) were sorted near the top of the list. Thus, stars whose brightness clearly changed significantly, but did not share the characteristics of periodic variable stars were still easily detected.

An example of this type of variable star found is shown in the figure below.

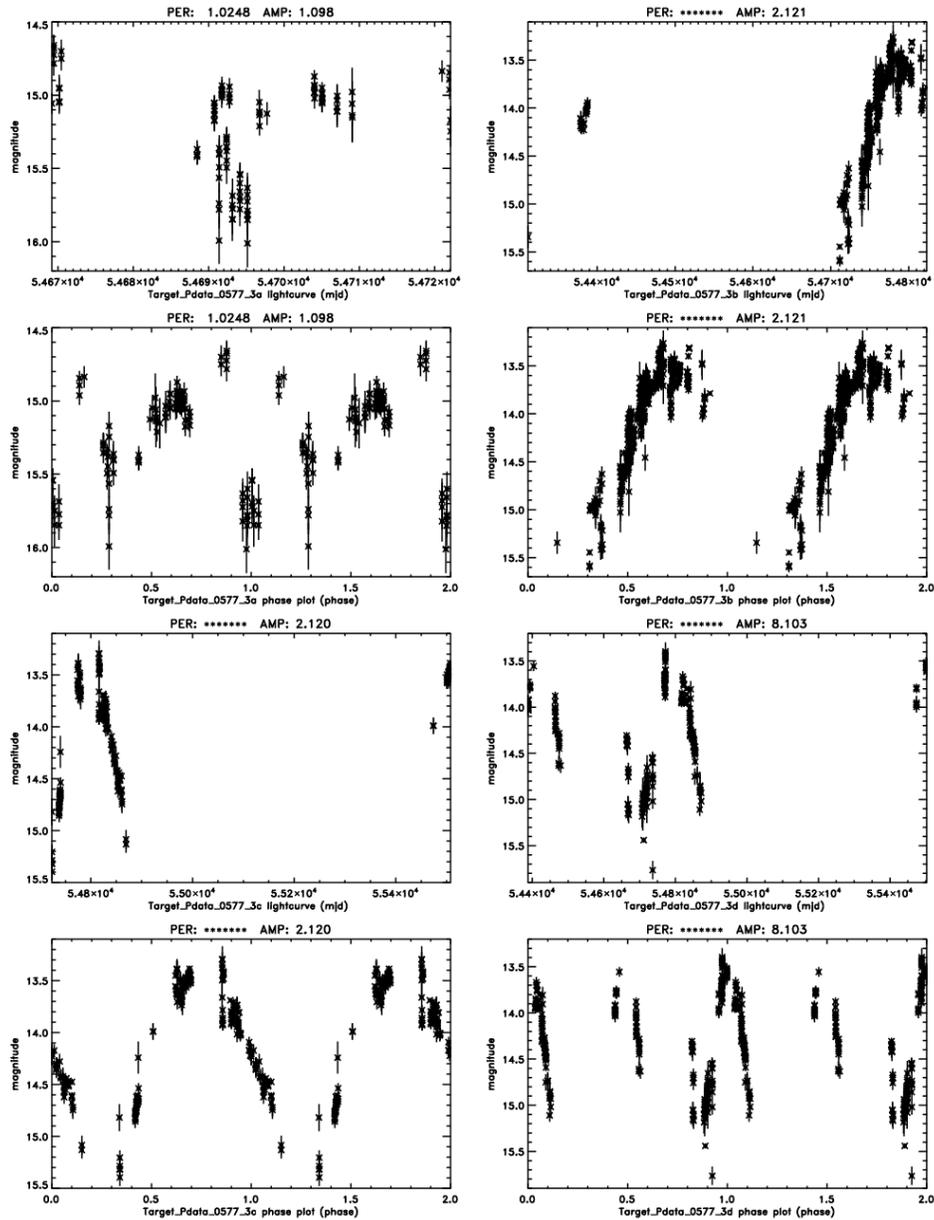


Figure 15: These are the lightcurves and phase plots for the irregular variable star *sks0258+0602\_0577*. The 4 sets of lightcurves and phase plots are of data from the 4 ROTSE telescopes, ROTSE a, b, c, and d.

Like the irregular variable, VSX J010919.1-150815, from Figure 4 above, the magnitude of *sks0258+0602\_0577* varies greatly. Thus, even though no period was successfully determined for this star, it was classified as a variable star successfully based on its amplitude alone.

## **Conclusion**

The type of detailed analysis of potential new types of variable stars to add to the system is important in making sure variable stars of those types are effectively detected and classified automatically in the future. The automated system has worked very well, and has greatly increased the output of variable stars discovered, with much less human interference. While the number of variable stars discovered per unit time using this new procedure is an order of magnitude greater than could be found manually, it is still many orders of magnitudes away from the estimated total number of variable stars in the galaxy. However, this type of procedure would be ideal for tackling the issue of analyzing large amounts of photometric data of stars in search of variables.

Overall, this automated procedure for discovering and classifying variable stars worked very well. While there certainly were some problems along the way, and remaining grey areas of stars which may be variable but were not effectively analyzed during this process, the merits of this procedure make it a quite worthwhile approach.

## **References**

[1]<http://www.aavso.org/vsx/>

[2]<http://rotse.net/>

[3]gcn<http://gcn.gsfc.nasa.gov/>

[4]Smith, Horace A. *RR Lyrae Stars*, Cambridge Astrophysics Series 27, 1995

[5]<http://www.sai.msu.su/gcvs/gcvs/index.htm>

[6]<http://www.aavso.org/vsx/index.php?view=about.top>

[7]Eyer, Laurent and Jan Cuypers. *Predictions on the number of variable stars for the GAIA space mission and for surveys as the ground-based International Liquid Mirror Telescope*, 2000, <http://arxiv.org/pdf/astro-ph/0002383.pdf>