Chasing Down Variables from a Decade-Long Dataset

Nikki G. Noughani[1] and Ralf Kotulla[2]

[1]Student, [2]Researcher (Department of Astronomy, University of Wisconsin-Madison) Madison, WI, USA

Abstract

Finding and characterizing variable stars is at the heart of a relatively modern branch of astronomy named time-domain astrophysics. Variable stars are ordinary stars that change in brightness on widely varying degrees. Characterizing the details of how this brightness fluctuates and on what timescales, as well as the star's average properties (mean brightness, temperature, etc.), is greatly important to understanding the process of stellar evolution. For this project, we use data over ten years from the 0.5 meter Sloan Digital Sky Survey telescope. We will develop tools that can identify several variable stars, estimate their variability periods and timescales, and determine the shape of their brightness variations over this time.

1. Introduction

For as long as we've looked up at the stars, we've known our night sky is constantly changing. However, it is only in our recent history we've had the means to truly measure the changes we've watched in a meaningful way. The first large classification of stars was done by Annie Jump Cannon, who worked under Edward C. Pickering and who created the Harvard Classification Scheme. This was the first serious attempt to organize and classify stars based on their temperatures and spectral types.

This data was what then allowed Hertzsprung [1] and Russell [2] to create the Hertzsprung-Russell diagram, which shows the relationship between stars' absolute magnitudes (or their brightness) against their effective temperatures. By plotting this relationship, the two found a clear division of stellar classes (see Fig. 1). This is one of the most important diagrams to understanding stellar evolution to date, as patterns can be traced using the clear classifications the graph shows. The diagram allows observers to trace the evolutionary paths of stars and it effectively proved the classification of a star is indicative of the star's surface temperature.

All stars vary as they evolve, but what distinguishes the normal variations of a star from those we identify as specifically as "variable stars" is the timescale. Normal stellar evolution is very long, varying over many decades or hundreds of years, whereas variable stars change on comparably short timescales, typically ranging from several hours to several months.

Variable stars are categorized into two groups: intrinsically variable and extrinsically variable. An intrinsically variable star varies due to properties of the star itself, for example if the envelope of the star swells and shrinks over time. An extrinsically variable star varies because of an external

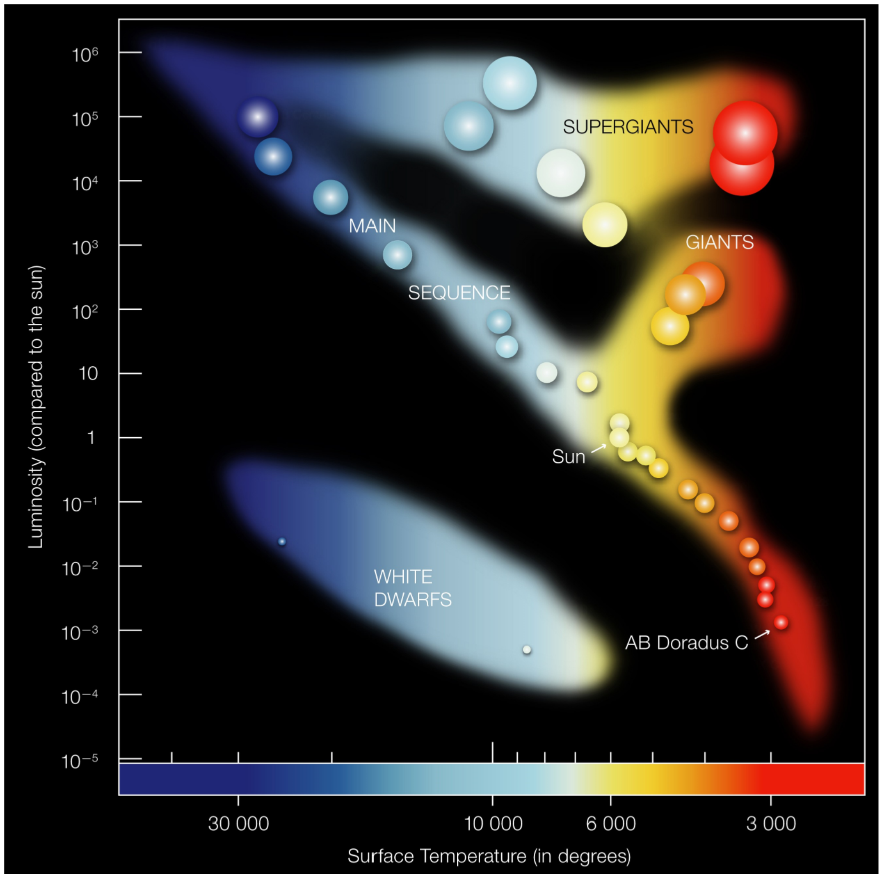

Figure 1: Hertzsprung-Russell Diagram, showing the main groupings of stars in the luminosity vs surface temperature plane. Credit: ESO press release eso0728c (19 June 2007), released under CC by 4.0.

obstruction to the amount of light that can reach our detection, for example a star that has a planet whose orbit periodically obstructs the star from our view.

Variable stars are useful for many reasons. In the case of intrinsically variable stars, there is a group of stars within this classification called Cepheid variables, who vary because of the swelling and shrinking of their envelopes as a result of the star's mass resonance. In the early 20th century, the astronomer Henrietta Leavitt discovered the relationship between the luminosity and the period of Cepheid variables, whose regular variations allowed her to calculate the distance of these stars as a result of this relationship. Cepheids are still used as a "standard candle" in astronomical observation to measure distance. As a recent example, they were used to demonstrate the shape of our galaxy is similar to that of a pringle chip by measuring Cepheids from all directions around our Milky Way [3]. Extrinsic variables are equally important. For example, binary star systems are a group of stars that fall under the categorization of extrinsically varying sources, and they are the most common stellar systems in our universe. By observing their variable patterns, we can collect more data and better understand the physics behind them.

2. SDSS Photometric Telescope

In the year 2022, the Large Synoptic Survey Telescope (LSST) will begin operation. This will be the world's largest digital camera taking the most thorough survey ever of the Southern sky. It will scan the sky several times a night each night for a decade, with the data collection being fully completed in the year 2032. This will be an incredibly important data set for those who study variable stars, however the 2032 end date is quite a way off, and this is only an estimate. Construction has already pushed back the operation date by several years, so there is no guarantee the data will indeed have been collected by 2032. And this is where our project gains importance.

The Sloan Digital Sky Survey (SDSS) photometric telescope in New Mexico is 0.5 meters in diameter and collects calibration data for the larger 2.5 meter telescope. This means for nearly a decade (2001 - 2009) the 0.5 meter telescope observed five standard star fields (SA fields) night after night, and each several times a night. This is exactly the data set necessary for collecting variable star information, and so the basis of our project was formed. We are analyzing the SDSS data in order to identify variable sources, some previously known and some previously unknown, so that we may create a program which can use machine learning to identify these sources independently of our tedious methods. This program could then be applied to other data sets, such as the LSST data, once they are available.

3. Procedure

The first step in our project was to reduce the data from the five fields in the data set, beginning by using the program TopCat to identify the fields. In particular, we began by looking at the data from the Landolt photometric calibration field SA 110 (4, see Fig. 2), choosing this to be our benchmark for our new data analytical program based on its location close to the galactic plane with a corresponding a large number of stars. From here, we used SourceExtractor [5] to extract aperture photometry for all sources in all individual frames. Subsequently we cross identified all sources to group photometry for individual stars in the field over the decade of data. Based on this we could extract light curves, i.e. apparent magnitude as a function of time, for each of our sources (see Fig. 3 for an example). Using these light curves, we determined the absolute variation in brightness across the entire observing period and compared this to the median photometric uncertainties. Sources for which the variability amplitude significantly exceeds typical measurement uncertainties are then identified as true variable sources.

Using this method, we identified 81 variable sources in the SA 110 field from a total of 2783 stars with at least 600 individual observations each. From the light curves of these sources a periodogram or power spectrum can be produced using the Lomb-Scargle algorithm [6], showing the power of variability for a range of different periods (see top panel of Fig. 4). Based on this we identify the most likely periodicity of each sources from the period of maximum power. Using this period we can generate a "phase-folded light curve," by folding the time of observations with the variability period (see bottom panel of Fig. 4).

Using the program AstroQuery, the coordinates of each of these variable sources are used to search the SIMBAD Astronomical Database to discover what the true identification of these sources are and whether they are already known or whether we have discovered a new variable source. At the

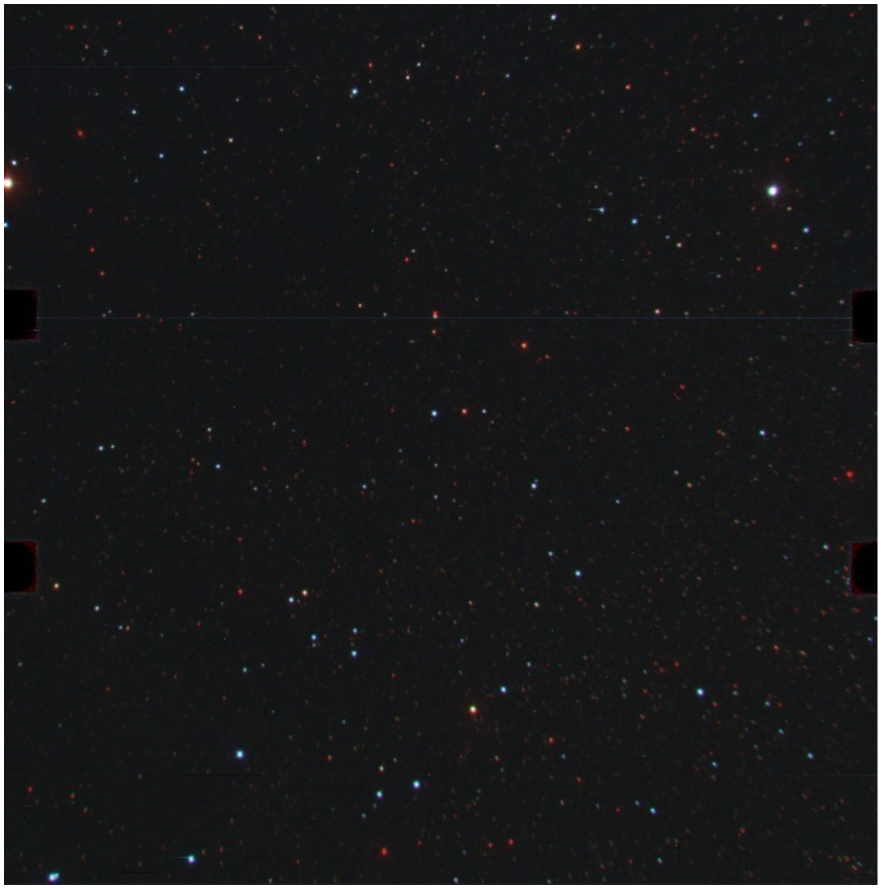

Figure 2: Standard star field 110 from SDSS data, shown as color-composite image combining data in the g^0, r^0, and i^0 bands. Field-of -view is approximately 1 degree on a side.

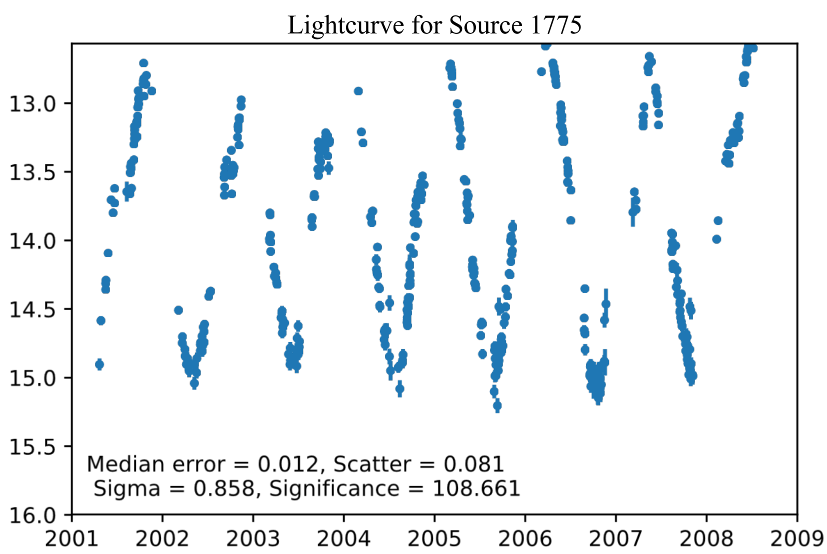

Figure 3: Light curve for Source 1775 in SA 110. Periodogram for Source 1775

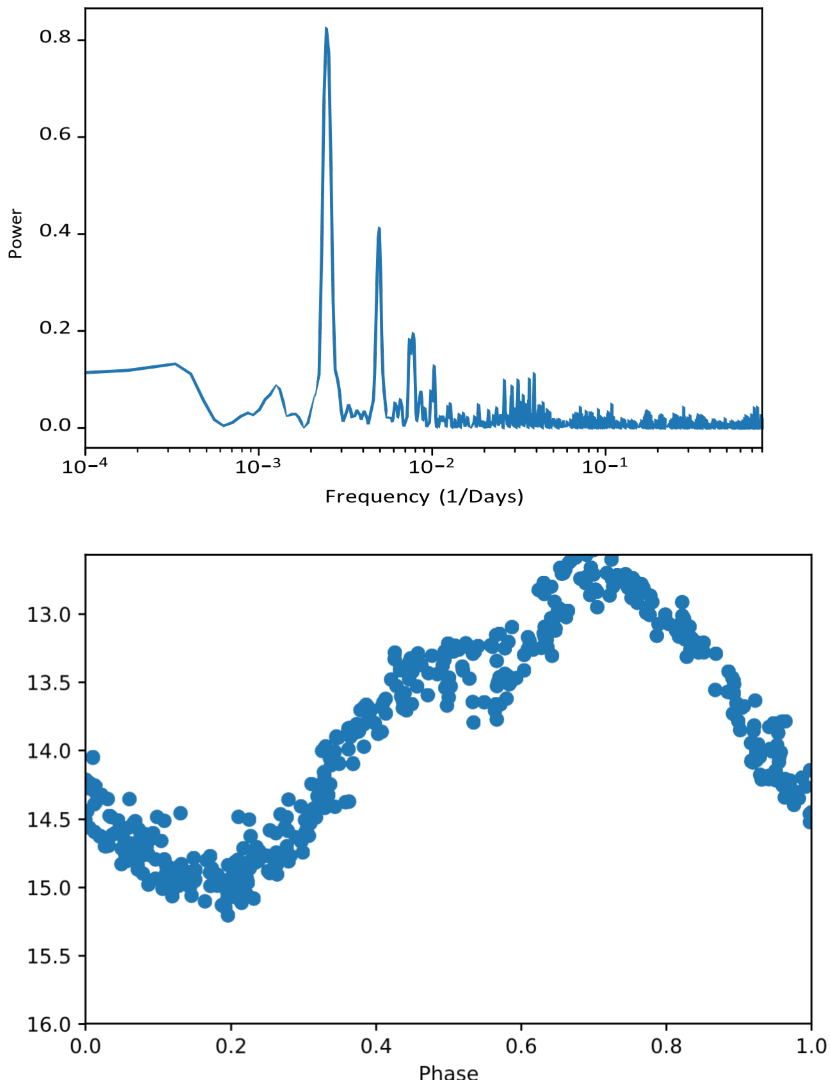

Figure 4: *Top panel:* Power-spectrum or periodogram for source-id 1775, generated from the light curve shown in Fig. 3. The obvious peak corresponds to a period of 275 days. *Bottom panel:* Phase-folded light curve for source-id 1775.

very least, even if the source is already well known the data from this telescope will add to our knowledge of the source. From the 81 variable sources we found in our pilot field only 8 stars are found in SIMBAD. The example shown in Figures 3 and 4 was identified as IRAS 18422+0019, and was previously identified as a long-period variable star in [7].

4. Summary and future work

Thus far we have written two important codes: the first to extract data for each source over all time and create a light curve (Figure 3), then a power spectrum (Figure 4, Top panel) and finally, using the Lomb-Scargle relation, a phase folded light curve (Figure 4, Bottom panel). The second code was able to use the database SIMBAD to identify whether the sources in our pilot field were known variable sources or not, and of the 81 variables found in SA 110 only 8 were identified by SIMBAD.

The final step will be to, with the use of machine learning, find out what the cause of the variability in each source is.

This first step before implementing machine learning will be to use a fitted spline curve on each phase-folded light curve. A spline curve will fit a function to the period seen in the folded light curve, which can then be over-plotted onto the full light curve and subtracted out. This will allow us to see if there are any longer variations to the sources that were not initially clear to us. This is useful for our later step of understanding why the source is varying as a source varying on two different scales would suggest two causes to the variation, whereas the computer may try to assign a single cause if the small and large variations are not separated out.

The next step will be to process the data from the other four SA fields in order to have a complete list of all the sources we have observed in this data set. Once we have done the same analyses on this new data and created a more specific set of boundary conditions, we will be able to identify variable stars in the data set with greater accuracy. Ultimately, the goal of this process will be to learn about the physical reasons behind why the star is variable, meaning we will try to identify whether the variability is due to intrinsic or extrinsic properties. This will be accomplished by observing the brightness variations and light curves of our sample and comparing them with the samples of light curves found and classified from other surveys. This process will be done via machine learning, where we will train a computer code to learn how to identify the source of the variability from the properties of the light curves in our data set from the previously discovered surveys (such as those found in SIMBAD earlier). This will then allow the code to be applied to our light curves and find a classification for each star.

Acknowledgments

NN is extremely grateful to Professor Jay S. Gallagher III for his support through this project. NN also gratefully acknowledges support from WGSC via a Student Fellowship NASA Training Grant #NNX15AJ12H. RK gratefully acknowledges partial funding support from the National Science Foundation under project AST-1664362 and from the National Aeronautics and Space Administration under project 80NSSC18K1498. Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

References

[1] Ejnar Hertzsprung. Uber die Sterne der Unterabteilungenc undac nach der Spektralklassifikation von Antonia C.¨ Maury. *Astronomische Nachrichten*, 179(24):373–380, 1 1908.

[2] Henry Norris Russell. Relations Between the Spectra and Other Characteristics of the Stars. *Popular Astronomy*, 22:275, 5 1914.

[3] Dorota M Skowron, Jan Skowron, Przemek Mroz, Andrzej Udalski, Paweł Pietrukowicz, Igor Soszy´ nski,´ Michał K Szymanski, Radosław Poleski, Szymon Kozłowski, Krzysztof Ulaczyk, Krzysztof Rybicki, and Pa-´ tryk Iwanek. A three-dimensional map of the Milky Way using classical Cepheid variable stars. *Science (New York, N.Y.)*, 365(6452):478–482, 8 2019.

[4] A.U. Arlo U. Landolt and A.K. Alan K. Uomoto. Optical Multicolor Photometry of Spectrophotometric Standard Stars. *The Astronomical Journal*, 133(1961):768–790, 3 2007.

[5] E Bertin and S Arnouts. SExtractor: Software for source extraction. *\aaps*, 117:393–404, 6 1996.

[6] Jacob T. VanderPlas. Understanding the Lomb-Scargle Periodogram. 3 2017.

[7] S. J. Wozniak, P. R. Williams, W. T. Vestrand, and V. Gupta. Identifying Red Variables in the Northern Sky´ Variability Survey. *The Astronomical Journal*, 128(6):2965–2976, 12 2004.