Second Workshop on Robotic Autonomous Observatories
ASI Conference Series, 2012, Vol. 7, pp 149 – 155
Edited by Sergey Guziy, Shashi B. Pandey, Juan C. Tello &
Alberto J. Castro-Tirado

Processing and variability detection from high-time cadence data; lessons from the Baker observatory sub-minute survey*

M. D. Reed[1†], L. L. Hicks[1,2], M. A. Thompson[1,2], J. T. Gilker[1] and L. A. Smith[2]

[1]*Baker Observatory and the Department of Physics, Astronomy, and Materials Science, Missouri State University, 901 S. National, Springfield, MO, 65897, USA*
[2]*Dept. of Computer Science, Missouri State Univ., 901 S. National, Springfield, MO, 65897, USA*

Abstract. Automated observatories are wonderful tools for collecting data. But once you have the data, you have to do something with it. Over the past two years, we have been conducting the Baker Observatory Sub-minute (BOS) Survey to search for short period phenomena. The BOS Survey uses a 0.4 m telescope with a Roper Scientific RS1340b CCD and no filter, located at Baker Observatory in southwest Missouri. We observe a 15' square field using exposure times of 2-20 seconds for four hour durations during two separate nights. We have developed software to process these data in a largely automated fashion. We have also implemented automated variability detection over a large range of timescales. For the shortest periods, we have found Fourier analysis to work well; for medium periods, we have tried several methods, including phase dispersion minimization; and we are even attempting to detect variability with timescales of days. However the biggest lesson we have learned is that data consistency is the most important component in recovering variability data.

Keywords : Stars; survey

1. Introduction

These proceedings contain many examples of autonomous or robotic telescopes which have obtained many thousands of images. The less human interaction required, the

*http://sdbv.missouristate.edu/boss
†email: `MikeReed@missouristate.edu`

easier it is for a telescope to accumulate data. However, processing data which telescopes can readily accumulate can be a daunting task. Automated scripting and pipelines are necessary to keep up, yet most scripted reductions are not generic. Rather, they are searching for one specific property, such as the measurement of a gamma-ray burst afterglow or binary star ephemerides where targeting one star or object is the goal.

We do not claim that our pipeline is any different. However, as we are accumulating survey data, we are interested in every star in the image. This does not make our pipeline general purpose but our bias is in time rather than target. We are searching for variable stars, of any type, but particularly with short periods ($<$ 30 minutes). In this paper, we briefly describe our image reduction pipeline and analysis programs. In the conclusions section, we discuss lessons we have learned along the way.

2. The Baker Observatory sub-minute survey

The survey is designed to look for objects which are variable over short timescales– on the order of a minute to a few hours. As our main scientific goal is the asteroseismology of compact pulsating stars, which have periods shorter than $\approx$ 15 minutes, we realized that the sky is largely unexplored in this time domain. High time cadence observations are usually targeted at specific objects. Yet many classes of variable stars were discovered serendipitously (for example, see Kilkenny et al. (1997)). When we began our survey, we were unaware of any other work to observe the sky at high temporal resolution. Now, we are aware of the RApid Temporal (RAT) Survey (Barclay et al. 2011), which shares our aims, though uses different methodology. The RAT survey had 1 to 2 minute cycle times and observed each field only once for 2 hours.

In the Baker Observatory Sub-minute (BOS) survey, fields are typically observed for four hours using 5-20 s integrations. Our CCD only has one second of dead time, which provides high temporal resolution, sensitive to variability just a few times the exposure time. This generates anywhere from several hundred to a few thousand images per run, with $\sim$ 80 stars, on average, per field. Each field is observed at least twice on separate nights. The two runs can be separated by as few as two days or as much as one year. Of course how we sample the sky affects how we search for variability. Data collected from this survey is optimized for the detection of short-period variable stars with periods less than four hours. However, our goal is to detect all variable stars within the field which the data will allow. Over the first two years of the survey, we observed fields which contained at least one known variable star. These covered all types of variability from eclipsing binaries, to classical Cepheids, to pulsating white dwarfs, with the exception of novae and Miras.

3. Data processing

Our data reduction pipeline is a program called PLC (pre-light curve) written by L. Hicks. This program joins several small programs, IRAF tasks, and MOMF (Multi-

Object, Multi-Frame) photometry (Kjeldsen & Frandsen 1992). It begins with raw images (and calibration frames) and finishes with barycentric corrected times and differential magnitudes for every star above a threshold. MOMF uses a weighted differential magnitude scheme minimizing point-to-point scatter while extracting fluxes using both point-spread-function fitting and an aperture correction. All processing occurs within the PLC environment, so the user does not need to switch between any programs or environments. The software currently requires some human interaction, but is significantly reduced from performing these tasks in the traditional manner.

Bias and flat field corrections are accomplished using standard IRAF routines, field WCS coordinates are set using astrometry.net (our telescope does not have encoders), and UT image header times are converted to barycentric times. MOMF then analyzes the images and custom programs convert MOMF output of differential magnitudes to lightcurves with each point assigned a quality rating. Variability detection routines then examine the lightcurves, often using only the best two quality ratings ($< 3\sigma$).

4. Variability detection

Our variability detection system is composed of several traditional tools and a few less-often used routines all combined within newly generated software. The framework is divided into several analysis paths to detect variability on different timescales. Fourier analysis is employed for periods from 1 to 20 minutes, phase dispersion minimization (PDM) is used to search for periodicities from 1 minute up to 90% of the run length (typically about 3.5 hours), and magnitude differencing is checked to detect variability on scales longer than the run length. We also make use of a scatter ratio technique which is part of the MOMF software and can detect variability from a few minutes to $\sim 8\times$ the run length. The Fourier and PDM periods are cross-checked for consistency, should they overlap.

Fast Fourier Transforms: Detection of short periods (1 to 20 minutes) is accomplished with a Fourier analysis. A Fast Fourier Transform (Press et al. 1992) is employed to analyze a light curve in the frequency domain. Fourier techniques can detect coherent signals (represented as a combination of sinusoids) buried in non-coherent noise. Real signals will produce peaks in the spectrum while noise will produce low level values across all frequencies. As this technique is only useful if several pulsation cycles are observed during a run, we flatten the lightcurve, removing trends longer than ~ 1 h with a third order polynomial and only search for peaks from 800 to $15000\,\mu$Hz.

A peak in the temporal spectrum 4σ above the mean noise level is considered to be significant (Breger et al. 1994) and the frequency is recorded. The quantity σ is the mean of the FT spectrum and we employ a σ-clipping routine so that high-amplitude pulsations do not unduly influence the noise level. Comparing two observations of the same field allows us to search for correlated peaks from the individual Fourier analysis. Two peaks that are within a resolution element (66 μHz for a four hour run) of one another are considered a match.

Phase Dispersion Minimization: Phase Dispersion Minimization (Stellingwerf 1978, PDM) is a well-established technique to find periodicity in a signal. It folds a light curve over a set of trial periods to find the one that correlates best (contains small point-to-point scatter). A statistic, θ (given below), is used to determine the fit for each trial fold.

PDM has similarities to the FT analysis but contains several differences. PDM does not require more than one periodicity within the length of a data run and readily detects periodicities with non-sinusoidal shapes, such as those present in most binary stars. Further, PDM can process data with large gaps in the sampling rate and still obtain periods. Gaps can appear in light curves as a result of bad weather or other unexpected events.

A low and high period bounding is chosen (one minute to 90% of the run length) and then PDM tests trial periods at discrete steps between these bounds. The data are binned and a θ statistic is calculated for each fold. The θ statistic is calculated as follows:

$$\theta = \frac{\sigma^2(n-1)}{n - bins} \tag{1}$$

where σ^2 is the variance of a single bin ($\frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$) and *bins* is the total number of bins (Press et al. 1992). We use 50 bins for our data, but it can be changed to adjust the resolution of the PDM fold. Since the data is folded in phase, these bins represent values across a range of 0 to 1.

Lower θ values indicate a good fit for that period. After calculating all the θ values, candidate periods are chosen. A smoothing algorithm is applied to the PDM output as true minima tend to be at the bottom of a gradual change in θ, so this smoothing pass works well to eliminate spurious periods. Minima below 0.85 after a smoothing pass are recorded as possible periodicities.

PDM can succeed only if at least one pulsation cycle is observed during any given run. This is why we only examine periods shorter than 90% of the run length. If PDM detects the same periodicity to within a resolution element, between two runs of any star, then that star is considered to be a candidate variable.

Magnitude Differencing: To search for variability on timescales longer than the length of a run, such as Cepheid variables, we search stars which appear in multiple runs for a change in brightness. For stars in each field, which do not appear as variable from other methods, we determine an average instrumental magnitude and standard deviation for each observing run. As we do not observe any standard stars or apply any brightness calibration on any runs, we correct one run to another for a given field. The difference in average magnitudes for the same stars between different runs is determined and averaged to be used as a brightness offset. Then the second run is corrected by the offset, so that on average, their magnitudes match those from the first run. A standard deviation in the offset is calculated and stars which have a brightness change greater than 3σ are considered candidate variable stars.

Since each approximately four hour run is treated as a single brightness point, this method is only sensitive to periodicities at least several times longer than the run length. Also, this method does not provide the period and requires follow-up observations of all candidate variables to confirm variability and determine the period.

Scatter Ratio: The MOMF (Kjeldsen & Frandsen 1992) program calculates a scatter ratio statistic during the course of its magnitude extractions. This statistic is the ratio of total root-mean-square (rms) noise to point-to-point rms noise. It is defined as follows:

$$d = \frac{\sigma_{total}}{\sigma_{internal}} \tag{2}$$

where σ_{total} is the total noise, or $\sqrt{\frac{1}{n}\sum x^2}$ and $\sigma_{internal}$ is the internal noise, defined as follows:

$$\sigma^2_{internal} = \frac{\sum_{i=1}^{n-1}(m_i - m_{i+1})^2 \sum_{i=1}^{n-1}(m_i - m_{i+1})^2}{2(n-1)} \tag{3}$$

This scatter measurement is a ratio of the variance of the entire light curve against another variance calculated on a point-to-point basis. A high ratio implies that the light curve is changing over timescales that are long compared to the integration time. This ratio can effectively detect variations for periods of a few minutes up to $\sim 8\times$ the run length. We count any star with a ratio above 2.0 as a possible variable star.

This method is best at detecting stars where less than one pulsation cycle is completed during a run. However, it can detect variability with periods as short as ~ 10 integrations, if the amplitude is high enough. A disadvantage of this method, is that it only indicates that a star is variable, but not the periodicity.

5. Results

We have completed our initial observing phase where we observed 40 different fields on two separate nights. We have reduced 22 of these fields which included 1965 stars, of which 84 are candidate variable stars. This represents 4% of observed stars, which is the same percentage that the RAT Survey (Barclay et al. 2011) found. From the 22 fields, we therefore had 22 known variable stars within these fields, of which 12 were recovered. The 10 known variables which were not recovered were due to several reasons including misidentification with an incorrect USNO star, non-matching periods between the two runs, typically caused by eclipsing binaries where the eclipse was only observed during one of the two runs, and pulsation amplitudes below our detection limit (three stars).

However, all of the software is in place and now it is just a matter of tweaking the software to optimize the results. The pointing accuracy of our images, determined by comparing astrometry.net coordinates to USNO ones is 1.3 arcseconds on average.

6. Conclusions

We have successfully observed 40 stellar fields over two separate nights to search for variability and are in the process of completing image reductions and analysis. Software has been produced which incorporates standard techniques for reducing images and searching for periodicities and is now in an advanced debugging and optimization stage. We have successfully recovered many of the known variable stars within our observed fields and discovered many new candidate variable stars. Our detection rate is about as expected.

However, we have certainly learnt an important lesson, which is: Insure data quality. The RAT Survey (Barclay et al. 2011) used short runs on many fields but over relatively few nights of photometric to near-photometric conditions using a 1.5 m telescope at CTIO with a wide ($0.28\,\mathrm{deg}^2$) field of view. We are using a 0.4 m telescope in Missouri, with a $0.0625\,\mathrm{deg}^2$ field of view, covering many more nights with longer-duration observations. As such, we have stumbled into several data quality issues. These include condensation on the CCD chip from high humidity (now fixed by flushing with dry nitrogen), a shutter problem, tracking issues, and, of course, foul weather. We have attempted to use these data despite their flaws, but if we only used our data obtained on photometric nights, reductions and detections would be much easier. We are now testing a robotic 0.4 m telescope at our same site to take over BOS Survey operations.

Acknowledgments

This research is sponsored by the National Science Foundation grant 0913175 and The Missouri Space Grant Consortium, funded by NASA. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect views of the National Science Foundation or NASA.

References

Barclay T., Ramsay G., Hakala P., et al., 2011, MNRAS, 413, 2696

Breger M., Ostermann W. M., Jiang S., et al., 1994, A&A, 289, 162
Kilkenny D., Koen C., OĎonoghue D., et al., 1997, MNRAS, 285, 640
Kjeldsen H., Frandsen S., 1992, PASP, 104, 413
Press W. H., Teukolsky S. A., Vetterling, W. T., et al., 1992, Cambridge: University Press, c1992, 2nd edition
Stellingwerf R. F., 1978, ApJ, 224, 953