Searching for Variable Events in High-Cadence Data

M. A. Thompson[a], M. D. Reed[a], and J. T. Gilker[a], and L. L. Hicks[a]

[a]Missouri State University, Department of Physics, Astronomy & Materials Science, 901 S. National, Springfield, MO 65897, U.S.A

ABSTRACT

Automated processing of data is an integral component to any autonomous survey system. We present a system that is designed to detect variability in several stages. For short period (under 20 minute) durations we employ Fourier analysis, for mid-range (1 minute to 4 hours) durations we use phase dispersion minimization, and for long period variations (> 4 hours) we use magnitude differencing. We also employ a scatter ratio technique which can detect variability from a few minutes to nearly a day, depending on amplitude. Multiple observations of the same field can be used to detect longer duration variability as well as complement the short duration techniques.

Keywords: Variable stars, detection, automation

1. INTRODUCTION

Most of what we know about stars has been learned using variability as a tool. From various types of binary stars we can determine mass, size, distance, and brightness. From pulsating stars we can also discern brightness, size, and distance as well as internal structure and processes. It is even possible to discern the rate of evolution, internal rotation, subsurface magnetic fields, and possible binary companions (including substellar) through pulsations (Kawaler et al. 1999; Mukadam et al. 2002; Reed et al. 2004; Silvotti et al. 2007). Pulsating stars are now being used to test fundamental particle physics (Winget et al. 2004; O'Brien & Kawaler 2000; Bischoff-Kim et al. 2008).

Our detection system is composed of several traditional tools as well as newly created software. The framework is divided into two parts: pre- and post-light curve. Images are reduced and corrected in the pre-light curve stage, resulting in light curves which are barycentric time versus differential magnitude from each image for each star. The pre-light curve stage has been implemented by L. Hicks in a software package called PLC, which uses IRAF scripting and the Multi-Object, Multi-Frame (MOMF(Kjeldsen & Frandsen 1992)) program to produce weighted differential photometry of all stars in a set of images above a flux threshold. These light curves are analyzed for variability in the post-light curve stage. We will discuss the post-light curve processing in this paper.

We utilize data collected during the Baker Observatory Subminute Survey (BOS survey*)(Gilker et al. 2010) in the development of our software. In the survey, fields are typically observed for four hours using 5-20 s integrations. This generates anywhere from a few hundred to a few thousand images, with $\sim$ 80 stars, on average. Each field is observed at least twice on separate nights. 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.

2. DETECTION

We employ several techniques to detect short period variability in light curves. 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 use a scatter ratio technique which can detect variability from a few minutes to $\sim 8\times$ the run length.

*http://sdbv.missouristate.edu/boss

2.1 Fast Fourier Transform

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 remove a third order polynomial and only search for peaks from 800 to 15000 μHz.

Figure 1. Light curve and 3^{rd} order fit of USNO-B1.0 0970-0310502.

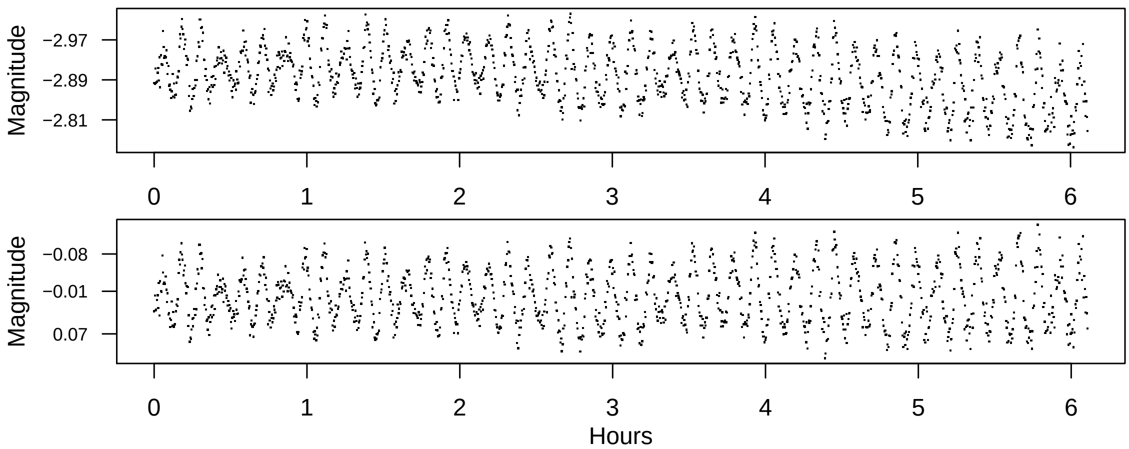
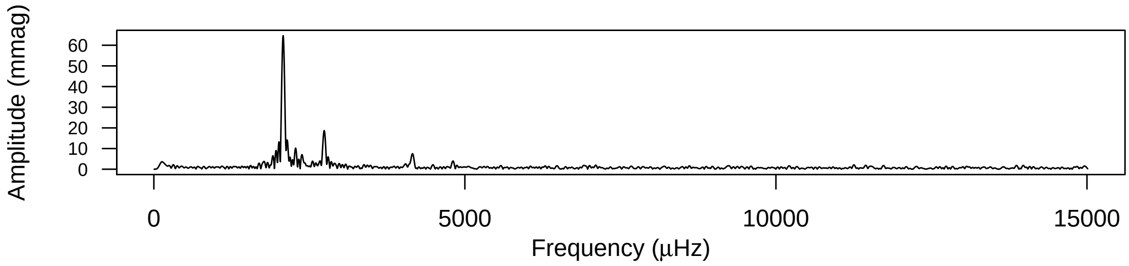


Figure 1 depicts the light curve of a short period pulsator from the BOS survey. The top panel shows the differential magnitudes straight from MOMF while the bottom panel shows the same data after removing a polynomial fit up to third order. Periodicities are then discovered by applying the FFT and analyzing the result. A plot of the Fourier transform of this light curve can be seen in Figure 2. A peak in the temporal spectrum 4σ above the noise level is considered to be significant and the frequency is recorded. σ is the mean of the FFT spectrum.

Figure 2. Fourier transform of light curve in Figure 1.



Comparing two observations of the same field allows us to search for correlated peaks from the individual Fourier analyses. Two peaks that are within a resolution element (66 μHz for a four hour run) of one another are considered a match. Peaks correlated from two separate observations of the star in Figure 1 are given in Table 1.

Table 1. Detected frequencies in Figure 2.

Frequency (μHz)	Amplitude (mmag)
2074	61
2122	37
2180	12
2240	11
2745	24
4149	8
4821	6

2.2 Phase Dispersion Minimization

Phase Dispersion Minimization(Stellingwerf 1978) or PDM is a 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. A statistic, θ, is used to determine fit for each trial fold.

PDM is similar to the FFT in that it detects periods in a light curve, but it has several advantages. Most importantly, it can detect periodicity that exists in 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 for us) 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. 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 all θ have been calculated, minima are chosen as candidate periods. 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. A complete set of θ for the light curve in the top left panel of Figure 3 can be seen in Figure 4.

Figure 3. Plot of USNO-B1.0 1359-0323772.

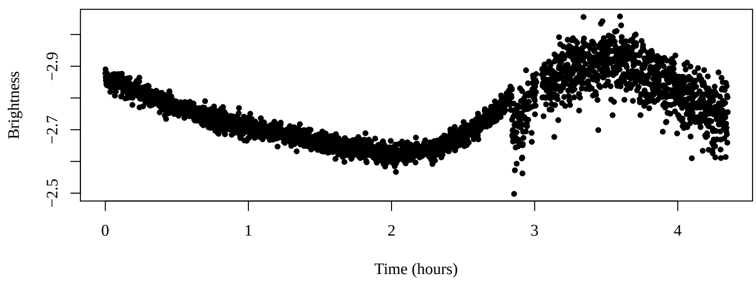
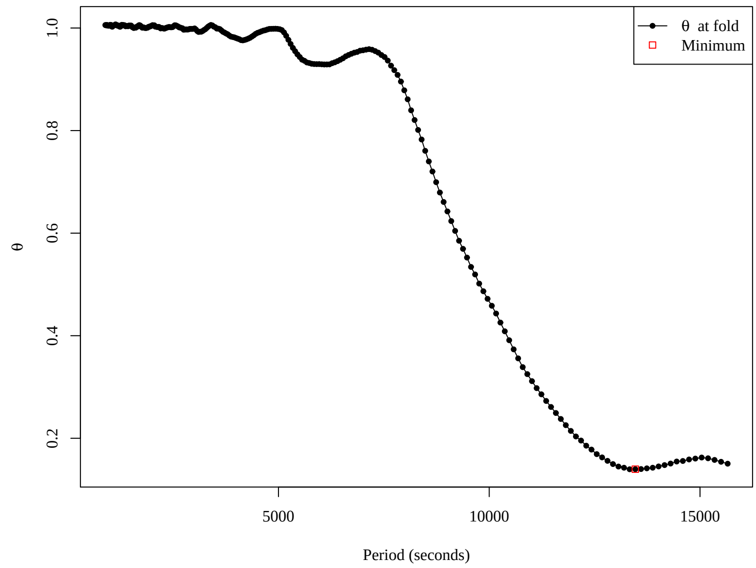


Figure 4. Plot of PDM θ for 1359-0323772.



With a low θ value, the minimum in Figure 4 is the "best" fold for this star. To visualize the bins for this fold, a plot with errorbars is used. Such a plot can be found in Figure 5. Each point represents a bin (doubled to more easily see trends). The errorbar for a given point is the variance in that bin. PDM folds with lower θ values will have smaller error bars. A bad fold is depicted in Figure 6. The variance in each bin is high - the data does not correlate well when folded over this period. As such, the error bars are wider and a clear signal can not be seen in the graph.

Figure 5. PDM fit with a period of 13465 seconds and a θ of 0.140.

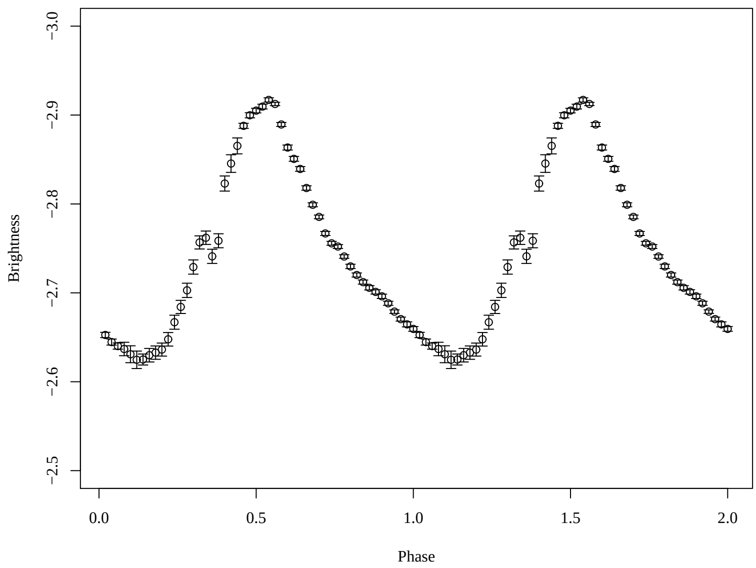
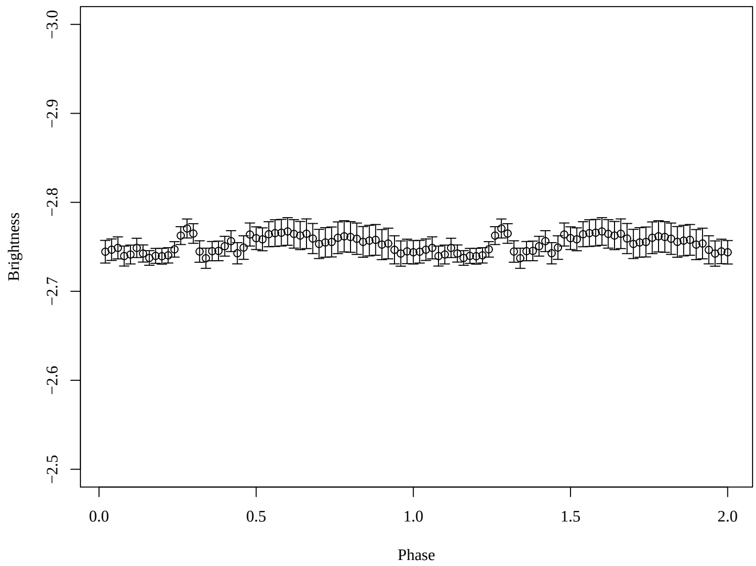


Figure 6. PDM fit with a period of 4978 seconds and a θ of 0.998.



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.

2.3 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.

2.4 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:

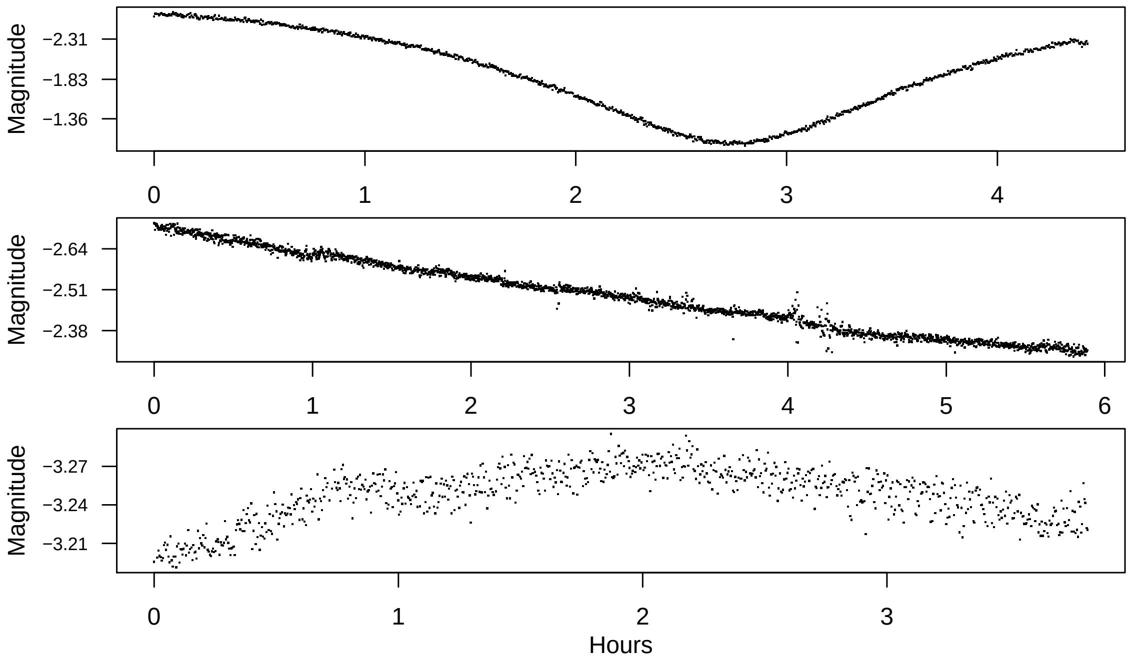
$$d = \frac{\sigma_{total}}{\sigma_{internal}} \tag{2}$$

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:

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

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. Examples of lightcurves which are well-suited to be detected by the scatter ratio are shown in Figure 7. These stars have ratios (from top to bottom) of 2.95, 34.17, 12.05, and 5.62.

Figure 7. Light curves of USNO-B1.0 0989-0202129, 1050-0293332, and 1640-0108422.



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. The scatter ratio for the lightcurve in Figure 1 is nearly four. A disadvantage is that is only indicates that a star is variable, but not what the period might be or if a star is multiperiodic.

3. CONCLUSION

The techniques discussed in this paper are currently employed in the Baker Observatory Subminute Survey, forming an important component in the autonomous data reduction and analysis pipeline. They have reliably recovered known variable stars with period from a few minutes up to several hours and have discovered many new candidate variable stars.

We are continuing to work on refining the tools and how they are employed. We currently do not make use of correlations between FT and PDM for periods where they overlap. We also only use FT and PDM positive results when they match between two or more observing runs. As such, they do not recover events such as eclipsing binaries where only one eclipse is observed or flare star if no flare occurs during one run. We are also accumluting statistics concerning which periods and amplitudes are being recovered, and which tools detect them the best. Since the scatter ratio and magnitude differencing techniques do not supply pulsation periods, variability detections using those methods require follow-up data.

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

Bischoff-Kim, A., Montgomery, M. H., & Winget, D. E. 2008, ApJ, 675, 1512

Gilker, J., Reed, M., Doennig, A., Quint, A., Hicks, L., & Thompson, M. 2010, Bulletin of the American Astronomical Society, 42, #418.05

Kawaler, S. D., Sekii, T., & Gough, D. 1999, ApJ, 516, 349

Kjeldsen, H., & Frandsen, S. 1992, PASP, 104, 413

Mukadam, A. S., Kepler, S. O., Winget, D. E., & Bergeron, P. 2002, ApJ, 580, 429

O'Brien, M. S., & Kawaler, S. D. 2000, ApJ, 539, 372

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Cambridge: University Press, —c1992, 2nd ed.,

Reed, M. D., et al. 2004, MNRAS, 348, 1164

Silvotti, R., et al. 2007, Nature, 449, 189

Stellingwerf, R. F. 1978, ApJ, 224, 953

Winget, D. E., Sullivan, D. J., Metcalfe, T. S., Kawaler, S. D., & Montgomery, M. H. 2004, ApJ, 602, L109