Variability Detection by Change-Point Analysis

Seo-Won Chang and Yong-Ik Byun

Department of Astronomy, Yonsei University, Seoul, Korea

E-mail : seowony@galaxy.yonsei.ac.kr

Variability Detection in Time Domain Astronomy

The detection and characterization of variability is often the first step to understand the nature of various cosmic objects.

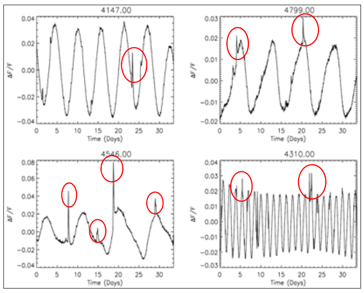


Figure. 1. *Periodic light curves of flaring K dwarfs (Walkowicz et al. 2011). The detection of significant outlying features (red circles) is not effective when the quiescent stellar variability remains in the raw light curve.*

- Most variability detection methods require conventional models that are mainly focused on the strictly periodic signals, and are not suitable for the study of arbitrary-shaped, non-periodic, and sporadically occurring variations, especially those of short time scales.
- Also, in many cases, signal estimation is equated with smoothing of data for de-noising. This sometimes discards vital information in time series data.
- We introduce a non-parametric method to extract all significant features based on the change-point analysis (CPA) with filtering algorithm using local statistics.

Change-Point Detection Algorithm

Using a combination of cumulative sum (CUSUM) scheme and bootstrap rank statistics (Taylor 2000), our method produces a series of estimated change points which correspond to the moments of apparent systematic changes, and then determines the optimal solution to minimize the false positives through backward elimination procedure.

Given dataset $x_1, x_2, \ldots, x_n$, the estimated change point location $\hat{p}$ is,

$$\hat{p} = \arg_{p_k}\left(\max_{1\le p_k\le n}\left|S_{p_k}\right|\right),$$

where $S_{p_0}=0$, $S_{p_k}=S_{p_{k-1}}+(X_{p_k}-\bar{X})$ and $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

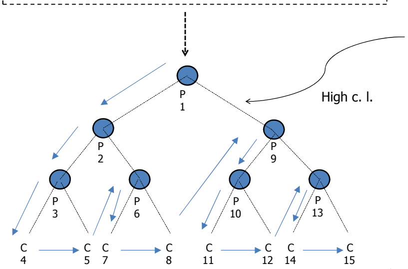
Based on N bootstrap samples that are randomly re-ordered original values, we calculate the maximum, minimum, and difference of the bootstrap CUSUM,

$$X_n = S^0_{\mathrm{diff}} < S_{\mathrm{diff}} = S_{\max} - S_{\min}$$

where $S_{\max}=\left(\max_{1\le p_k\le n} S_{p_k}\right)$, $S_{\min}=\left(\min_{1\le p_k\le n} S_{p_k}\right)$.

It is straightforward to derive estimates of confidence level as follows:

$$\text{Confidence Level (c. l.)} = 100\times\frac{X_n}{N}\,\%$$



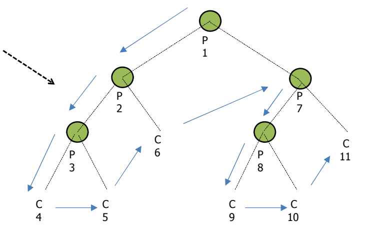
The sub-region is successfully segmented into exactly two segments by the CPA algorithm,

$$\bar{X}_1=\cdots=\bar{X}_{p_k}\neq\bar{X}_{p_{k+1}}=\cdots=\bar{X}_n$$

If no change-points could be found at all, the adjacent sub-region will be considered.

$$\bar{X}_1=\bar{X}_2=\cdots=\bar{X}_n=\bar{X}$$

iterative χ^2 goodness-of-fit test



Using backward elimination with different selection levels, we remove some candidates that are not statistically significant anymore.

Outlying feature detection

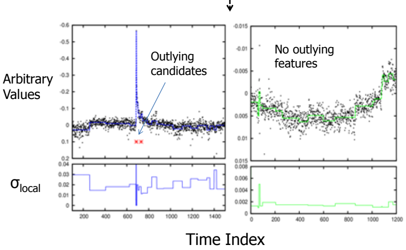


Figure. 2. *Examples of variability detection by change-point analysis*

We define a simple criteria similar to Micro-lensing Alert system (Glicenstein 2001) in the presence of hetero-scedastic measurement errors (W_i):

$$\frac{\left(X_i-\bar{X}_{p_k \text{ or } global}\pm W_i\right)}{\sigma_{local}}\geq N,\ \mathrm{ConM}\geq M$$

where $\sigma_{local}=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(X_i-\bar{X}_{p_k}\right)^2}$ and

ConM = consecutive measurements of X_i

The values of N, M are selected by maximizing the detection efficiency of *significant* outlying features.

Application to MMT Transit Survey Data

By applying this method to over 30,000 light curves from the MMT transit survey of M37 open cluster (Hartman et al. 2008), we efficiently identified several hundred instances of abrupt brightness changes **without any smoothing or interpolation of the raw data**.

FINDFlare : Flare-like event detection

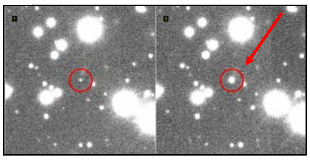
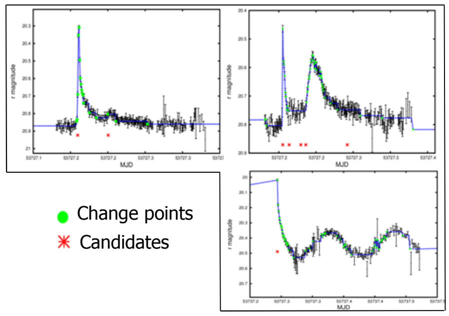
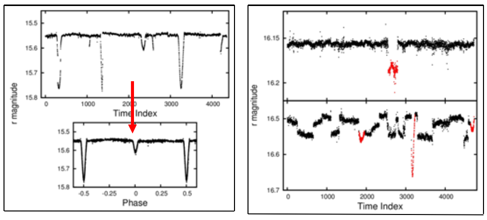


Figure. 3. *(Left) Images show the sudden optical brightening of the object. This phenomenon shows up as a series of outlier points in the light curve, (Right) Example light curves of discovered flare-like events.*



- This method is optimized to detect multiple outlying features (upper right panels) embedded in real astronomical datasets that are unevenly spaced in time and show statistically non-stationary noise behavior.
- Without any filtering (e.g., moving average) of periodic variability (lower right panel), our method also reveals the inherent stochastic nature of variability.

FINDEclipse : Eclipsing-like event detection



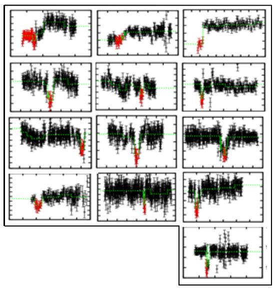


Figure. 4. *(Left) Typical light curve of detached eclipsing binary, (Middle) Non-periodic event and complex eclipsing system, (Right) Various light curves of discovered eclipsing-like variable candidates with small amplitude and short duration.*

- Because time series observations often suffer from the incomplete coverage of the entire eclipse periods, eclipsing variables are usually observed repeatedly to form a complete phased light curve (left panel).
- Our method is useful to detect the moments of eclipse ingress, center, or egress in cases where the eclipsing pattern is not repeated or the coverage is not sufficient for the detection through conventional period analysis (middle and right panels).

False positives

- Most false detections are easily recognized because they are caused by certain systematics such as artifacts on the images, seeing-correlated variations accompanied by image blending, diffraction spikes from nearby bright stars, and crossing moving objects.

Conclusion

- We confirm that the CPA is a powerful method to determine whether a change has taken place in time series dataset. From our re-analysis of MMT transit survey data, we found previously unknown evidences about stellar variability, including a total of 606 flare events, 18 eclipsing-like features, and 3 transit-like features.
- Our CPA approach is particularly effective in detecting non-periodic events from data with varying noise as well as short duration events from either non-varying or varying light curves.

References

Glicenstein, J. –F. 2001, ASPC, 239, 28

Hartman, J. D., et al. 2008, ApJ, 675, 1233

Taylor, W. 2000, Change-Point Analyzer 2.0 shareware program, Taylor Enterprises, Libertyville, Illinois. Web: http://www.variation.com/cpa

Walkowicz, L. M., et al. 2011, AJ, 141, 50