Transiting Planets
Proceedings IAU Symposium No. 253, 2008
Frédéric Pont, Dimitar Sasselov & Matthews Holman, eds.

doi:10.1017/S1743921308026677

De-Trending Time Series Data for Variability Surveys

Dae-Won Kim[1,2,3], Pavlos Protopapas[2,3] and Rahul Dave[3]

[1]Department of Astronomy, Yonsei University, Seoul, Korea
email: dakim@cfa.harvard.edu

[2]Harvard Smithsonian Center for Astrophysics, Cambridge, MA, USA

[3]Initiative in Innovative Computing at Harvard, Cambridge, MA, USA

Abstract. We present an algorithm for the removal of trends in time series data. The trends could be caused by various systematic and random noise sources such as cloud passages, change of airmass or CCD noise. In order to determine the trends, we select template stars based on a hierarchical clustering algorithm. The hierarchy tree is constructed using the similarity matrix of light curves of stars whose elements are the Pearson correlation values. A new bottom-up merging algorithm is developed to extract clusters of template stars that are highly correlated among themselves, and may thus be used to identify the trends. We then use the multiple linear regression method to de-trend all individual light curves based on these determined trends. Experimental results with simulated light curves which contain artificial trends and events are presented. We also applied our algorithm to TAOS (Taiwan-American Occultation Survey) wide field data observed with a 0.5m f/1.9 telescope equipped with 2k by 2k CCD. With our approach, we successfully removed trends and increased signal to noise in TAOS light curves.

1. Introduction

The conventional approaches to remove trends is differential photometry with a reasonable selection of template stars near the star of interest (Young *et al.* 1991, Everett & Howell 2001). With the help of modern CCDs, it is not hard to select enough bright stars as a template set for the determination of trends. However, the de-trending results are then sensitive to the selection of template stars. If the template stars contain intrinsic variables, the determined trends will be different from the true trends. Therefore excluding such intrinsically variable stars from template stars is essential for the trends determination.

In this proceeding, we present a new de-trending algorithm which has a better and systematic template selection method that can solve the problems mentioned above, showing better de-trending results. Experiments with simulated light curves show that our algorithm selects reasonable template stars and re-generates the original signals successfully. We also used the TAOS (Alcock *et al.* 2003) dataset to test our de-trending algorithm. The main goal of the TAOS project is the detection of Kuiper Belt Object (Luu & Jewitt 2002) by means of stellar occulations. TAOS is a wide field survey with four 0.5m telescopes, each equipped with a 2k x 2k CCD camera which has a 1.8 deg^2 field of view. TAOS light curves have low signal to noise ratio (S/N) due to the short exposure time, less than 1s and also shows strong trends due to unstable weather conditions (Lehner *et al.* 2008, Zhang *et al.* 2008).

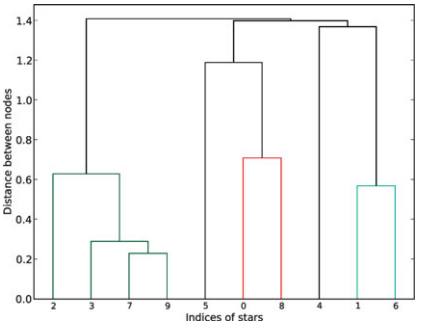

Figure 1. A dendrogram.

2. Algorithm

Our de-trending algorithm chooses the template stars using the similarity matrix, C, of the correlation between light curves of all the stars. Its elements are the Pearson correlation values between each pair of light curves. If the light curve of a star manifests a trend without being intrinsically variable, then the star would be highly correlated with a lot of other stars. However if a star has both trend and intrinsic variability, it would not be highly correlated with other stars. Therefore a star which has strong correlation with many other stars is an optimal template candidate for de-trending.

We use a clustering algorithm to extract those highly correlated stars in the similarity matrix. Clustering is a useful method for grouping data according to similarity. It is used widely for color extraction or image segmentation, data mining, or document retrieval as in text search engines (Shi & Malik 2000, Grabmeier & Rudolph 2004, Hammouda & Kamel 2004). Among several clustering algorithms, we use the hierarchical tree clustering algorithm (Jain *et al.* 1999). This algorithm first defines the distance between any two stars as $1 - C$. It then links stars into nodes (as represented by a linking in the dendrogram) based on the smallest distance between the stars. The distance between two nodes is now defined as the longest distance from any star in the first node to any star in the second node. We then recursively link nodes based on the smallest distance between them, organizing stars and nodes into a hierarchical tree. The hierarchy tree is traditionally represented by a dendrogram as shown for a representative set in Fig. 1.

In order to construct a template set, we now need to extract clusters from this tree which are highly correlated only among themselves. At least one free parameter such as threshold cut of distance is needed. This is the conventional problem of hierarchical tree algorithms (Daniels & Giraud-Carrier 2006). Unfortunately, it is almost impossible to choose such a threshold cut a priori because we don't have any physical information of trends. Therefore, we developed a new agglomerative merging algorithm (Kim *et al.* 2008, hereafter Kim08). At each level in the dendrogram, we calculate the centroid of the distance measures of nodes at that level. We compare this centroid to the centroid of the nearest node on the dendrogram at the next higher level. If this difference is larger than the largest mean separation of distances amongst all stars, we do not merge the nodes, and stop. Thus we have a dynamical threshold to the level of clustering depending upon the characteristics of the trends (such as a localized effect in a region of the CCD). Fig. 1 shows each cluster identified by our algorithm in each different color.

With the identified clusters, we determine the trends in each cluster by producing a Master Light Curve (MLC) using the standard-deviation-weighted sum of normalized

https://doi.org/10.1017/S1743921308026677 Published online by Cambridge University Press

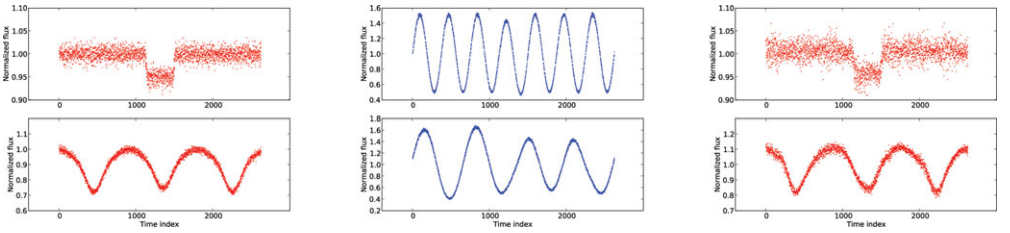

Figure 2. Left : Artificial light curves of two events. Top is a transit and bottom is an eclipsing binary. Middle : Light curves after applying trends. Right : De-trended results of two events.

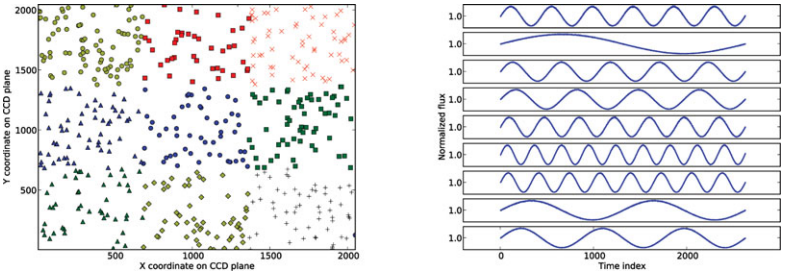

Figure 3. Left : Positions of identified clusters. Right : 9 determined master trends.

light curves (Kim08). Finally, we use the well-known multiple linear regression method to de-trend each individual light curve using the MLC (Kim08).

3. Test with Synthetic Light Curves

We tested our de-trending algorithm with a field of 512 simulated light curves. Each star has a different mean flux with poisson noise added, and the coordinates of stars are randomly distributed. We divided the entire field into 3x3 rectangular area and inserted different trend in each area. The trends are sine curves with different frequencies (representing for eg. clouds moving at different speeds). Finally, we added 2 different events into two of the bright stars. One is a transit-shaped signal (Mandel & Agol 2002) and the other is an eclipsing binary (left two panels of Fig. 2). On applying the trends, large amounts of original signals is washed out (middle two panels of Fig. 2).

Figure 3 shows the 9 identified clusters by our algorithm. Left panel of Fig. 3 shows the positions of stars in each cluster while right panel shows the determined master trends. None of the identified clusters contain the two simulated event stars; which means our clustering algorithm is capable of excluding the signals which confuse other de-trending algorithms.

With the 9 identified master trends, we de-trended the synthetic light curves. As expected, we recovered the artificially added events. The simulated intrinsic signals on these events are well re-generated (right two panels of figure 2).

4. Application to TAOS data

We selected sample data from a night's observation of a field in TAOS which had strong trends. We show in Fig. 4 the results of de-trending on a single light curve in this sample. The strong trends in the top panel are well removed as shown in the bottom panel. We also show the improvement in the S/N of all light curves after de-trending in the right

https://doi.org/10.1017/S1743921308026677 Published online by Cambridge University Press

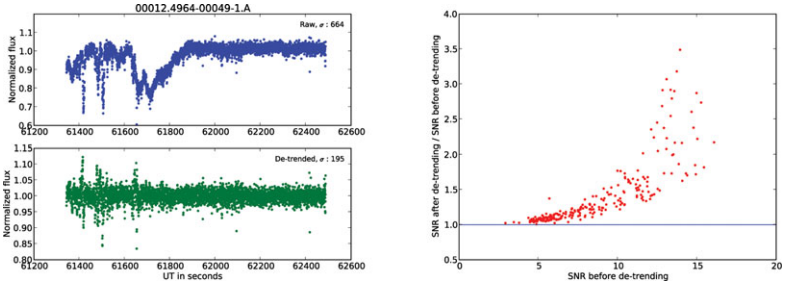

Figure 4. Left : One example of de-trended light curve of TAOS data. Right : S/N comparison between after- and before de-trending.

panel. The S/N of all stars is increased after applying our de-trending treatment. Because faint stars suffer from other noise sources such as Poisson noise more than bright stars do, the improvement of faint stars is not as big as the improvement of bright stars.

5. Conclusion

We developed a new de-trending algorithm based on the similarity matrix and the hierarchical clustering algorithm. Experiments with simulated light curves show that our algorithm selects better template stars for the trends determination; which yields better de-trending results.

References

M. E. Everett & S. B. Howell, PASP, 2001

K. M. Hammouda & M. S. Kamel, IEEE Transactions on Knowdeledge and Data Engineering, Vol. 16, No. 10, 2004

K. Daniels & C. Giraud-Carrier, Proceesing of 5th International Conference on Machine Learning and Applications, 2006

J. Grabmeier & A. Rudolph, Data Mining and Knowledge Discovery, Vol. 6, No. 4, 2004

A. K. Jain, M. N. Murty, & P. J. Flynn, ACM Computing Surveys, Vol. 31, No 3., 1999

Kim *et al.* in preparation.

M. J. Lehner, C.-Y. Wen, J.-H. Wang, S. L. Marshall, M. E. Schwamb, Z.-W. Zhang, F. B. Bianco, J. Giammarco, R. Porrata, C. Alcock, T. Axelrod, Y.-I. Byun, W. P. Chen, K. H. Cook, R. Dave, S.-K. King, T. Lee, H.-C. Lin, S.-Y. Wang, arXiv:0802.0303

K. Mandel & E. Agol, ApJL, 2002

J. Shi & J. Malik, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, 2000

A. T. Young, R. M. Genet, L. J. Boyd, W. J. Borucki, G. W. Lockwood, G. W. Henry, D. S. Hall, D. P. Smith, S. L. Baliumas, PASP, 1991

Z.-W. Zhang *et al.* in preparation.

https://doi.org/10.1017/S1743921308026677 Published online by Cambridge University Press