Period detection in light curves from astronomical objects using correntropy

Conference Paper · July 2010

DOI: 10.1109/IJCNN.2010.5596557 · Source: DBLP

5 authors, including:

Pablo Estevez
University of Chile

Pablo Andres Huijse
Universidad Austral de Chile

Pablo Zegers
University of the Andes (Chile)

Jose C Principe
University of Florida

Period Detection in Light Curves from Astronomical Objects Using Correntropy

Pablo A. Estévez, *Senior Member*, *IEEE*, Pablo Huijse, Pablo Zegers, *Senior Member*, *IEEE*, Jose C. Principe, *Fellow Member*, *IEEE*, and Pavlos Protopapas

***Abstract*— In this paper we propose a new method for determining the period in astronomical time series using correntropy, an information theoretical concept recently developed in the computational intelligence field. The time series correspond to the stellar brightness over time, so-called light curves, and are characterized as being noisy and unevenly sampled. The advantages of using correntropy instead of correlation are to escape from the constraints of linearity and Gaussianity and are clearly demonstrated. The performance of the proposed method is compared with other algorithms published in the literature on a set of light curves drawn from the MACHO survey. The results show that the correntropy-based method obtains the correct periods more frequently than the Lomb-Scargle periodogram and the Period04 program.**

I. Introduction

The increasing use of CCD technology in astronomical observation has allowed the collection of data from a large number of sources at a scale that was not thought possible before. This has allowed to produce complete surveys of all types of astronomical objects, measure all types of events, and therefore, helped to improve the understanding of the underlying physics.

One of the things that is important to detect is periodic behavior, a telltale of objects such as eclipsing binaries, RRLs (pulsating variable stars named after RR Lyrae), cepheids (intrinsically variable stars with exceptionally regular periods of light pulsation), etc [11]. From the detection of recurrent events it is possible to infer many characteristics of the observed objects. This has spun the use of data mining techniques in astronomy, thus allowing to process large numbers of data and give statistical meaning to the measurements.

The time series analyzed in this work comes from photometric surveys, i.e. collections of observations of the intensity of light (or magnitudes) of astronomical objects in multiple channels (different spectral bands, telescopes, or instruments). Due to the nature of observation schedules, objects are observed unevenly over time with short or even long gaps in between. The MACHO (MAssive Compact Halo Object) survey [1], operated with the purpose of searching for the missing dark matter in the galactic halo, like brown dwarfs or planets, by observing a light amplification due to microlensing[1]. This data set has been an excellent source for finding variable stars, and for astronomical light curves, i.e. brightness magnitude over time. In addition to have an irregular sampling, light curves are noisy, and its sampling rate may not be appropriate for detecting the event of interest. On top of that there are billions of astronomical objects; another reason why an efficient and automatic method of finding periodicity is required.

Current techniques use Lomb-Scargle (LS) periodogram [8], [9], which is an extension of the classical periodogram developed by the signal processing community. The LS periodogram is based on the idea of fitting a trigonometrical model to the time series in a mean square sense. The calculations are done in a "per point" fashion instead of the classical "per interval of time" approach. By using the LS technique it is possible to estimate the spectra of an unevenly sampled time series. The estimated period given by the LS periodogram is then used to "fold" the time series, i.e. it is plotted modulo the estimated period, such that the periodic behavior is clearly seen. Then, the estimated period is trimmed such that the scatter of the folded plot is reduced. Once this is achieved it is possible to count with a precise measure of the period. This final step is known as analysis of variance (AoV) in astronomy and is due to [16]. This method is very powerful and accurate however is very computationally expensive due to the fact that the search resolution is inversely proportional to the total time span of the light curve. The LS periodogram method can produce multiple possible periods and it is not trivial to determine which one is the correct period with the corresponding increase of computational effort needed to determine the meaningful one.

The need to automate the discovery of periodic behavior and the inherent difficulty of this problem calls in for computational intelligence techniques [2], [3], [14], [17], [19]. This work focuses on using information theoretical techniques based on the correntropy concept [10], [15], [18] to explore the possibility of designing an efficient and fully automated periodic event detector.

The objective is to determine whether the correntropy concept produces better results or not, and if it is possible to use this concept to design an automated discovery process. In order to do so we process a set of light curves from the

Pablo A. Estévez is with the Electrical Engineering Department of Universidad de Chile; email address: pestevez@cec.uchile.cl.

Pablo Huijse is with the Electrical Engineering Department of Universidad de Chile; email address: pablo.huijse@gmail.com.

Pablo Zegers is with the College of Engineering and Applied Sciences of Universidad de los Andes; email address: pzegers@miuandes.cl.

Jose C. Principe is with the Computational Neuroengineering Laboratory of University of Florida; email address: principe@cnel.ufl.edu.

Pavlos Protopapas is with the IIC, School of Engineering and Applied Sciences of Harvard University; email address: pprotopapas@cfa.harvard.edu.

[1]Heavy objects curve the space around them so when light travels by an object, the light can be magnified

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

MACHO survey. This temporal data was obtained using non uniform sampling, i.e. the number of days within samples changed all the time, and at different times during the night, which implies different levels of attenuation and noise thanks to the varying width of the atmosphere.

We first introduce the reader to the concept of correntropy and then move on processing the light curves with a mix of standard signal processing techniques and this new information theoretical tool. We finish this paper with a discussion on how to improve the proposed procedure and advance in the direction of building up a completely automated process for detecting periodicity in astronomical time series.

II. Correntropy

Correntropy is a generalized correlation function introduced in [15]. It is defined in terms of inner products of vectors, which can be computed using a positive definite kernel function, κ, satisfying Mercer's conditions

$$\kappa(x_i, x_j) =< \Phi(x_i), \Phi(x_j) >,$$

where $\Phi(x_i)$ transforms the data x_i nonlinearly from input space to high-dimensional feature space.

For a discrete-time strictly stationary univariate random process, $\{x_n\}$ with $x_n \in \Re, n \in \{1, \ldots, N\}$, the autocorrentropy is defined as

$$V[m] = E[\kappa(x_n - x_{n-m})],$$

where $E[\cdot]$ denotes mathematical expectation. The most popular Mercer kernel is the Gaussian function ,

$$G_\sigma(x_i, x_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

where σ is the kernel size or bandwidth. We will apply the Gaussian kernel throughout this paper. Note that the correntropy function include higher order statistical information about the input random process. The kernel size allows controlling the emphasis given to higher order moments versus the second moment. The second moment corresponds to the autocorrelation function,

$$R[m] = E[x_n x_{n-m}].$$

The autocorrentropy can be estimated through the sample mean

$$\hat{V}[m] = \frac{1}{N-m+1} \sum_{n=m}^{N-1} G(x_n - x_{n-m}) \quad (2)$$

for $0 < m < N-1$. The name correntropy comes from the fact that it looks like correlation but the sum over the lags is the argument of Renyi's quadratic entropy. The correntropy spectral density (CSD) is defined as

$$P[f] = \sum_{m=-\infty}^{\infty} (\hat{V}[m] - < \hat{V}[m] >) e^{-j2\pi \frac{f}{F_s} m} \quad (3)$$

where $< \hat{V}[m] >$ is the mean value of correntropy. This is tantamount to the Fourier transform of the centered correntropy function.

In [18] correntropy was applied to pitch detection, i.e. the determination of the fundamental frequency in speech. Correntropy obtained much better resolution than correlation and other methods in pitch determination. Other applications are detection of sinewaves in impulsive noise regimes [7], and blind source separation [6]. Correntropy can be employed also as a discriminative measure for detecting nonlinearity in time series [10].

III. Methods

We use light curves over time drawn from the MACHO survey available at Harvard University, Time Series Center. The light curves data come in three columns: time, flux, and error. The latter is an estimation of the error of measurement in the photometric procedure. We discarded points with errors greater than a threshold satisfying the following criterion

$$|e(i) - < e > | > 2 \times std.dev(e).$$

where $< e >$ is the mean of the error. The time series were analyzed using the following procedure:

1) Select windows having at least 100 points without gaps larger than 10 days. This results in windows of length $N > 100$, with approximately 1 sample per day.
2) Fit a cubic spline to selected windows.
3) Resample the data evenly in selected windows at $F_s =$ 20 samples per day.
4) Choose a kernel size and calculate correntropy using (2) with a cap of $N/10$ lags.
5) Apply Hamming window over centered (zero-mean) correntropy.
6) Compute correntropy spectral density using (3) for a range of frequencies in $[0, F_s/2]$.
7) Detect peaks of the square modulus of CSD versus frequency. Save the period (inverse of frequency) associated to the maximum peak as a candidate to be the right period.
8) Repeat last four steps using a different kernel size until completing the set of values searched for this parameter.
9) Find the best fitting period among the set of candidates by folding the light curves. Transform the time axis according to
$$\tau = \frac{(t \;\; mod \;\; T)}{T},$$
where T is the candidate period and mod stands for the modulo operation. In this way the light curve is folded onto the interval $[0, 1]$. This interval is partitioned in bins of size 0.05. All data within a bin is averaged to get a single value. The maximum value in the folding curve is searched for. This is repeated for each candidate period. The period that gets the maximum peak in the folding curves is chosen as the right one.

Note that the last step is a way to choose the best kernel bandwidth, because each period candidate has associated a given kernel size. The criterion that the best fit for the period produces the highest peak in the folding curve, is not necessarily general for all kind of stellar objects. But it is mostly true for eclipsing binaries, the objects of the current study.

There are two parameters involved in the window selection process, the maximum gap size allowed and the minimum window size. The values of these parameters were chosen heuristically, taking into account that the range of periods goes from less than a day to more than 100 days. In addition, the time series available have an average of 1000 samples and a mean sample density of about 0.5 samples/day. There is a trade-off when choosing the maximum gap size allowed: the larger the gap size, the lower the sample density. A sample density of 0.5 makes highly unlikely to detect a period of 1 day or less, therefore we search for windows with larger sample densities. On the other hand, if the maximum gap size allowed is too small, then it would become harder to find windows containing enough samples to detect larger periods. Correntropy was calculated using a 10% of the window size, in order to obtain smoother spectra. If the correntropy were calculated using the whole window size, then the estimations would become very noisy due to the different number of samples considered for each lag.

IV. Experimental Results

A. Analysis of a Single Light Curve

First we will illustrate our method in detail with a single time series. We study the sequence *lc* 1.3804.164 *blue*, which is illustrated in Fig. 1. The best known period for this time series, found using the AoV method and visually inspected, is **4.18753** days, which produces a peak of 17.1372 in the folding curve as shown in Fig. 2. In what follows we will use this value as the correct period for our analysis.

Following the procedure described in section III, two windows containing more than 100 points without gaps larger than 10 days were extracted from the light curve, the first

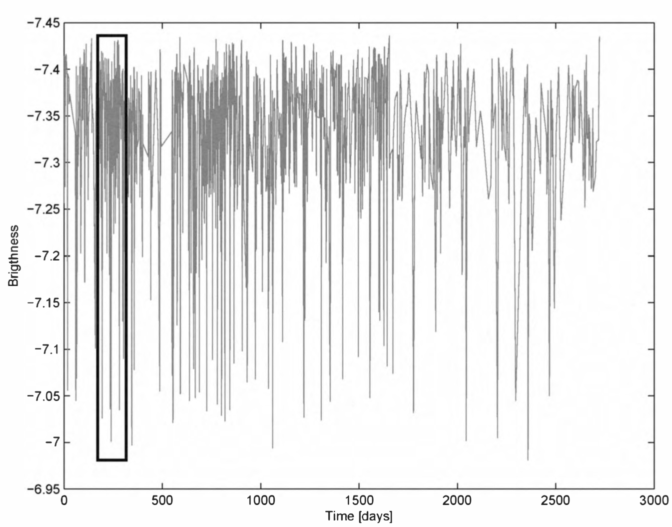

Fig. 1. The light curve *lc* 1.3804.164 *blue* and a selected window from 80 to 203 samples

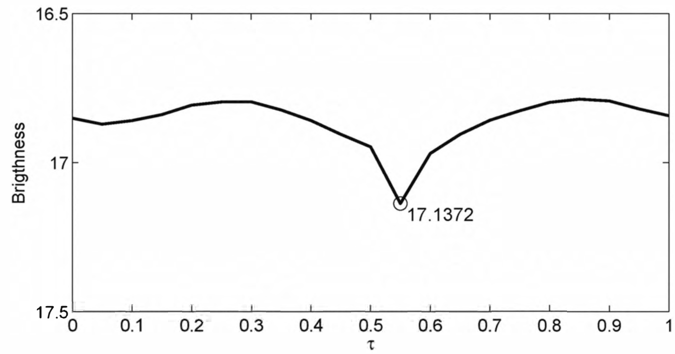

Fig. 2. Folding of the whole time series using a period of 4.18753 days. It shows a peak of 17.1372. Note that the y-axis represents the *magnitude* of the star. Magnitude measures the brightness of a celestial object, however the brighter the object appears, the lower the value of its magnitude. It is customary in astronomy to plot the magnitude scale reversed.

from the 80 to 203 samples and the second one from the 359 to 484 samples.

1) The 80-203 Window: The whole light curve and the selected window are shown in Fig. 1. The window contains 143.9 days with an average of 0.85471 days per sample. In order to erase the effects of the non-uniform sampling cubic splines are used to interpolate and then resample the window in a uniform manner. A Hamming window is applied in order to smooth out the artifacts caused by the borders.

First, to establish a reference the autocorrelation is calculated, as shown in Fig. 3, where the number of lags has been limited to $1/10$ of the data size. The power spectrum density (PSD) of the correlation is illustrated in Fig. 4. The maximum peak of the PSD corresponds to a period equal to 6.9871 days, which if used to fold the time series it produces a peak with value 16.8958, which is far away from the correct period and peak. A second peak has associated a period of 33.6352 days. In fact none of the periods associated to the peaks found in PSD, correspond to the correct period. This means that the autocorrelation is unable to find the right period for this astronomical time series.

Fig. 5 shows a waterfall plot of correntropy power spectral density versus kernel bandwidth and period. Three peaks can be clearly seen. The first peak corresponds to a period of 4.1878 days for a kernel bandwidth of 0.018, which is very near to the correct period. The second peak corresponds to a period of 6.8829 days for a kernel bandwidth of 5.0, which is very near to period found by autocorrelation. The third peak corresponds to a period of 17.8559 days for a kernel bandwidth of 10.0. These results are in agreement with the theory, which indicates that correntropy approaches correlation for large values of kernel bandwidth, i.e. the second moment becomes predominant. However for smaller values of kernel sizes, the influence of the higher-order moments allows finding the correct period. Fig. 6 shows the correntropy for a kernel size of 0.018, and Fig. 7 illustrates correntropy power spectral density, which shows a distinctive peak when the period is 4.1878 days, very close to the period found by Harvard-TSC team (4.18753).

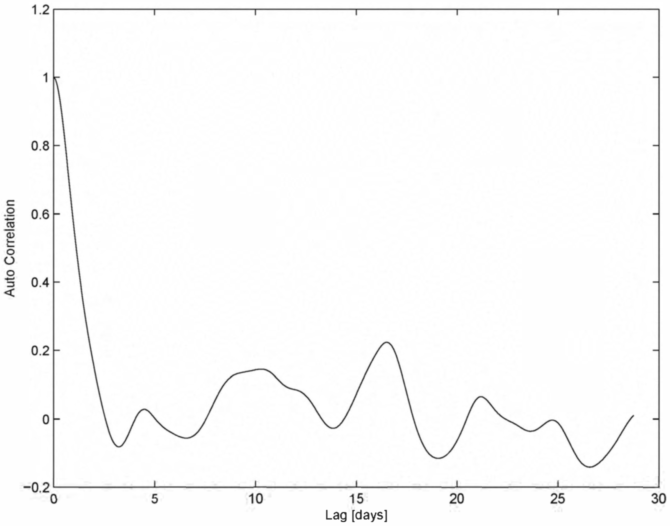

Fig. 3. Autocorrelation function obtained from the 80-203 window.

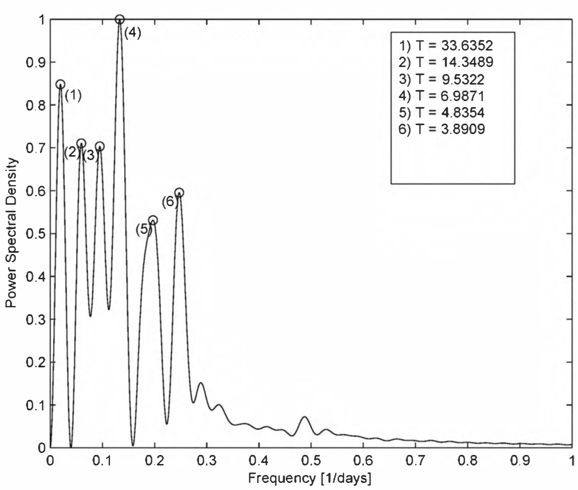

Fig. 4. Power spectral density obtained from the autocorrelation function.

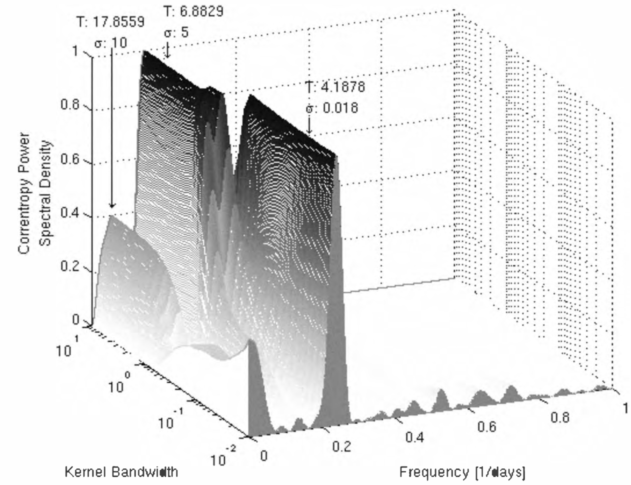

Fig. 5. Correntropy power spectral density as a function of the period and the kernel bandwidth.

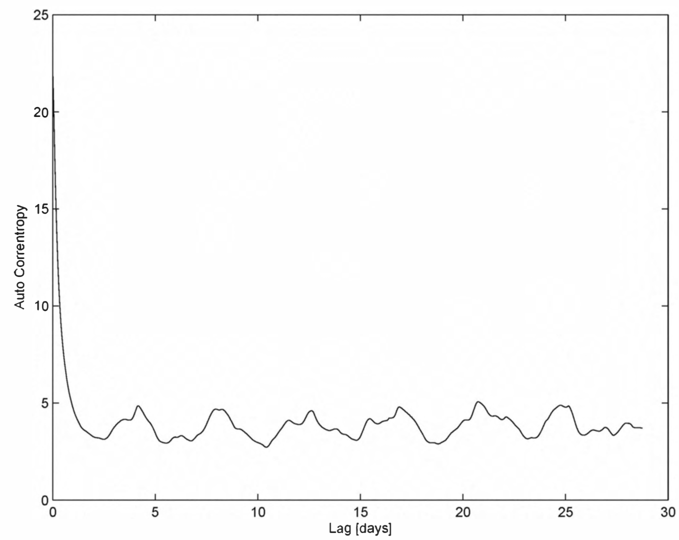

Fig. 6. Correntropy plot for a kernel bandwidth of 0.018.

Fig. 8 shows the results of folding the whole time series using the period of 4.1878. The peak found by the folded sequence is 17.1311, which is very close to the peak found by AoV method: 17.1372. The correntropy was able to produce a peak that differed from that of the AoV method a mere 0.0064%. This is an estimate good enough to start a fine tuning of the folding process and obtain a better period estimate.

2) The 359-484 Window: The same process was done with the second window of samples. As showed in figures 9 and 10, again the autocorrelation method was unable to find the correct period. It found a period of 30.0957 days, associated to a peak of 16.8851 in the folded plot.

Fig. 11 shows a waterfall plot of correntropy power spectral density versus kernel bandwidth and period. The main peak corresponds to a period of 4.1871 days for a kernel bandwidth of 0.75, which is very near to the correct period. This period generates a peak of 17.0736 in the folded data. All these values are extremely close to those obtained with the AoV method (4.18753 days and a peak of 17.1372) and the difference amounts to only 0.0103%. Fig. 12 shows the correntropy for a kernel size of 0.75, and Fig. 13 illustrates correntropy power spectral density, which shows a distinctive peak when the period is 4.1871 days.

B. Results for Sixteen Light Curves

In this section we present the results of applying our correntropy-based method to 16 light curves drawn from the MACHO survey. For comparison purposes we use Period04 and Vartools programs. Period04 is a computer program dedicated to the statistical analysis of large astronomical time series containing gaps [5]. The VarTools program provides tools for calculating variability/periodicity statistics of light curves as well as tools for modifying light curves [4]. The LS

TABLE I

ESTIMATED PERIODS USING THE CORRENTROPY-BASED METHOD, HARVARD-TSC, PERIOD04 AND VARTOOLS-LS.

Light Curve	Correntropy-based Method			Harvard-TSC	Period04	VarTools-LS
Blue Channel	Window [Samples]	Kernel Bandwidth	Period [Days]	Period [Days]	Period [Days]	Period [Days]
1.3448.153	426-523	0.40	3.2211	3.2765	3.2764	3.2764
1.3449.27	468-603	0.02	4.0239	4.0349	2.0174	2.0174
1.3449.948	394-567	0.40	6.9994	14.0064	7.0037	7.0037
1.3564.163	68-287	0.01	4.6179	4.7155	4.7158	4.7159
1.3804.164	80-203	0.018	4.1878	4.1875	2.0937	2.0937
1.3804.164	359-484	0.75	4.1871	4.1875	2.0937	2.0937
1.3809.1058	338-935	0.08	14.3489	28.9073	14.4548	14.4544
1.3810.19	1-424	0.05	44.4685	88.9406	44.4619	44.4611
1.4168.434	39-489	0.02	21.8746	43.9301	21.9593	21.9597
1.4173.1409	321-612	0.02	7.0486	14.1534	7.0763	7.0763
1.4174.104	61-649	0.08	4.2412	8.4929	4.2465	4.2465
1.4288.975	352-647	0.08	8.6845	17.6131	8.8069	8.8071
1.4411.612	1-1100	0.10	22.4889	45.1143	22.5648	22.5641
1.4538.81	56-279	0.05	5.4845	5.5343	2.7671	2.7671
1.4539.37	53-251	0.06	2.9570	2.9955	3.0093	1.4977
1.4539.778	333-457	0.25	16.2076	16.2502	8.1254	8.1253
1.4652.565	47-625	0.06	13.7102	25.5718	13.7860	13.7860

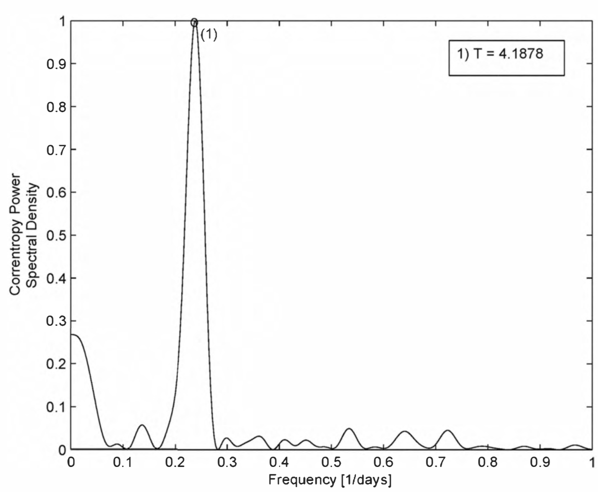

Fig. 7. Correntropy power spectral density for a kernel bandwidth of 0.018. It clearly shows a peak when the period is 4.18780 days, very close to the value found with AoV (4.18753 days).

option of VarTools performs a Lomb-Scargle period search on the light curves [8], [9], [12], [13]. As a golden standard we use the periods published at the website of Harvard University, Time Series Center. These periods have been computed using AoV and visually inspected by the Time Series Center team.

Table I shows the estimated periods for 16 stellar objects (one of them using two different windows) using the correntropy-based approach. The periods determined by Harvard-TSC, Period04 and VarTools-LS are illustrated for comparison purposes. Using Harvard-TSC periods as golden

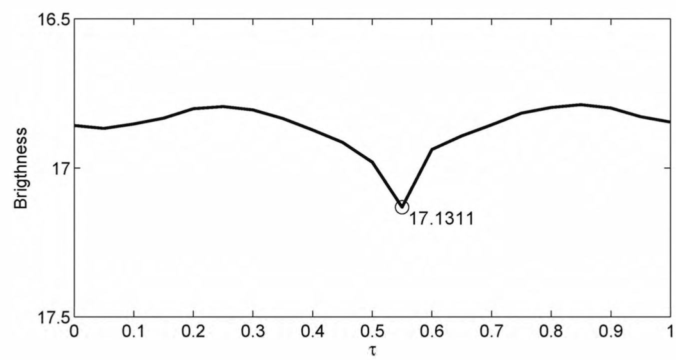

Fig. 8. Folding of the whole time series for a period of 4.18780 days. It presents a distinctive peak of 17.1311.

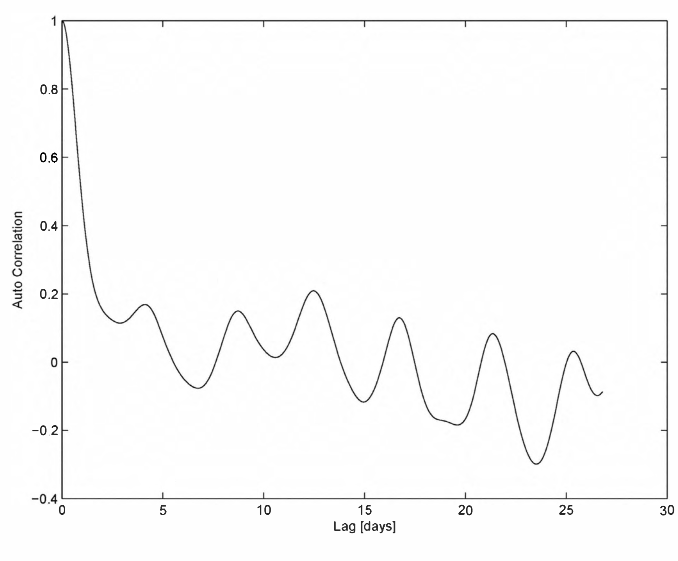

Fig. 9. Autocorrelation function obtained from the window.

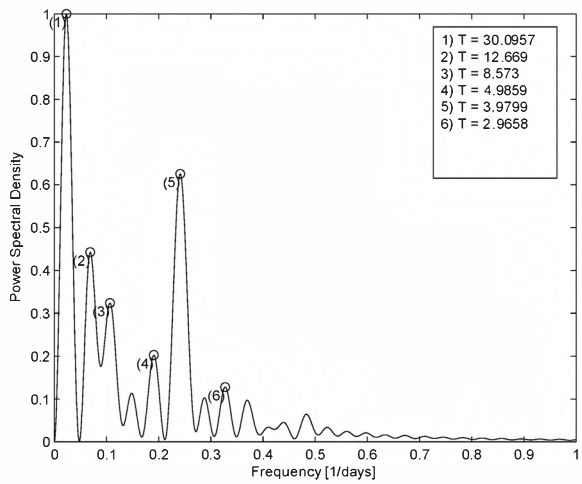

Fig. 10. Power spectrum density obtained from the autocorrelation function.

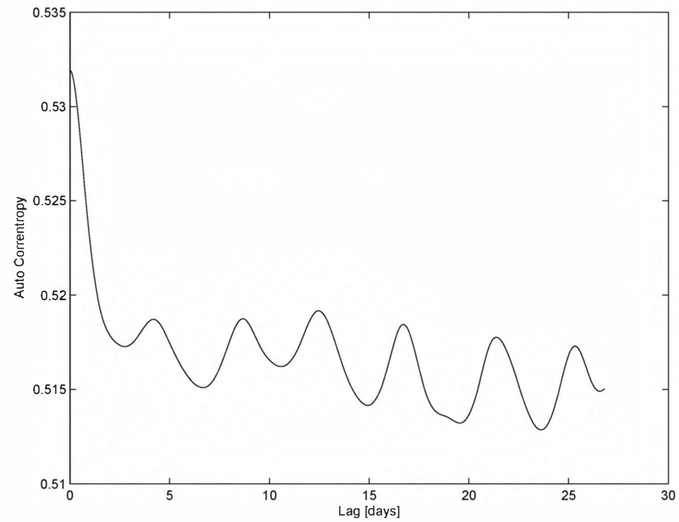

Fig. 12. Correntropy plot for a kernel bandwidth of 0.75.

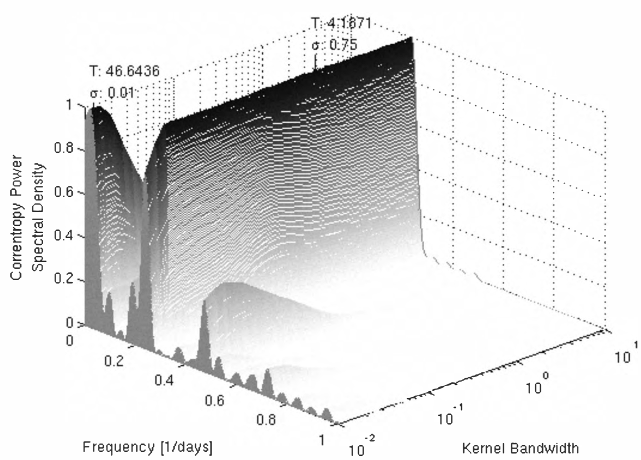

Fig. 11. Correntropy PSD surface as a function of the period and the kernel bandwidth (different point of view of the same surface).

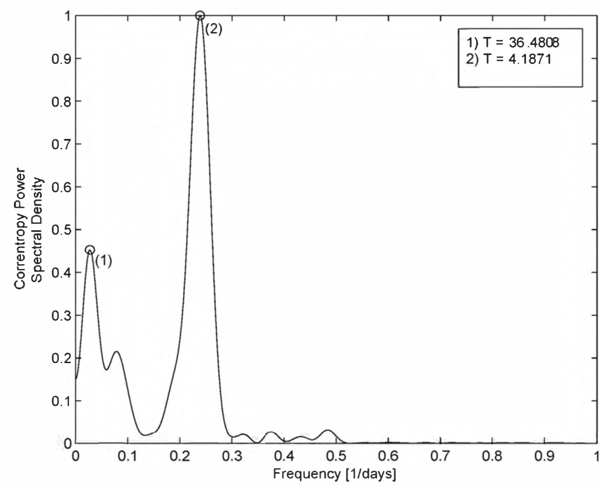

Fig. 13. Correntropy spectral density for a kernel bandwidth of 0.75. It clearly shows a peak when the period is 4.1871 days, very close to the value found with the folding process (4.18753 days).

standard, the correntropy-based method gets 7 hits out of 16, while Period04 and VarTools have 3 and 2 hits, respectively. In all mismatched cases the period determined was half of the correct period.

V. Discussion

As mentioned earlier, this study deals with light curves from eclipsing binaries. Binary stars are systems where two stars orbit around a common center of mass. An eclipsing binary is a binary star whose orbital plane is exactly aligned with the plane of the sky. The mutual eclipses of a binary system produce a particular pattern in the folded light curves as shown in Fig. 14. The biggest decrease in brightness occurs in point 1, when the brighter star of the binary system is fully blocked. Point 2 corresponds to the eclipse of the secondary star, whose brightness is very similar to the main star. The correct period corresponds to interval 1-1 in Fig. 14, but a common mistake is to obtain half of the real period (interval 1-2). As shown in Table I, all the tested methods obtained half of the real period several times.

To deal with this situation, an alternative is to analyze several peaks of CPSD instead of just the global maximum. Even if it does not correspond to the maximum peak, the real period should still appear in the spectra of the auto-correntropy. In addition, the peak of the folded curve could be replaced as the method used to discriminate candidate periods. A more robust approach should be devised, for example, comparing variances between the original signal and the folded signal, as in the ANOVA criterion used by Harvard-TSC.

VI. Conclusions

A method based on correntropy has been proposed for period detection in astronomical time series with gaps. The

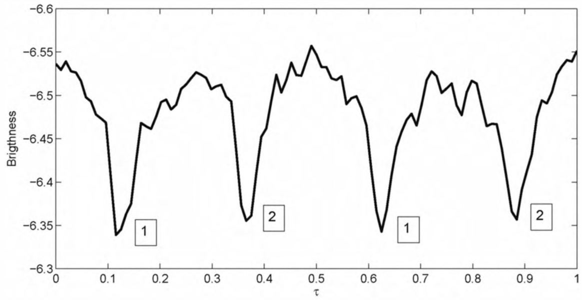

Fig. 14. Folded light curve for eclipsing binary lc 1.3809.1058 $blue$ from Macho catalog, using twice the real period. There are two eclipses of relatively equal depth.

advantages of using correntropy instead of correlation were clearly illustrated with a detailed example. Seven out of sixteen light curves' periods were correctly determined by the proposed method, which is an improvement with respect to Period04 and VarTools-LS programs that found only 3 and 2 correct periods, respectively. In the other 9 cases, the periods determined by our method corresponded to half of the correct periods. There is room for improvement in several steps of the method. A better interpolation and resampling method could be used instead of cubic splines. The problem of how to handle missing samples should be considered more deeply. Other aspects are the choice of the kernels and how to deal with the error measurements. In addition, from the practical point of view it is important to make the whole method automatic in order to process the massive databases available in astronomy. The preliminary results presented here confirm the feasibility of using correntropy for detecting periodicity in astronomical time series.

ACKNOWLEDGMENT

This research was supported by CONICYT-CHILE under grant FONDECYT 1080643.

REFERENCES

[1] C. Alcock, R.A. Allsman, D.R. Alves, T.S. Axelrod, A.C. Becker, D.P. Bennett, K.H. Cook, N. Dalal, A.J. Drake, K.C. Freeman, M. Geha, K. Griest, M.J. Lehner, S.L. Marshall, D. Minniti, C.A. Nelson, B.A. Peterson, P. Popowski, M.R. Pratt, P.J. Quinn, C.W. Stubbs, W. Sutherland, A.B. Tomaney, T. Vandehei and D. Welch, "The MACHO Project: Microlensing Results from 5.7 Years of LMC Observations," Astrophysical Journal, vol. 542, pp. 281-307, 2000.

[2] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," Proceedings of the 5th Annual ACM Workshop on COLT, pp. 144-152, Pittsburgh, USA, 1992.

[3] J. Debosscher, L. M. Sarro, C. Aerts, J. Cuypers, B. Vandenbussche, R. Garrido and E. Solano, "Automated Supervised Classification of Variable Stars. I. Methodology," Astronomy and Astrophysics, vol. 475, pp. 1159-1183, December, 2007.

[4] J. D. Hartman, B. S. Gaudi, M. J. Holman, B. A. McLeod, K. Z. Stanek, J. A. Barranco, M. H. Pinsonneault and J. S. Kalirai, "Deep MMT Transit Survey of the Open Cluster M37. II. Variable Stars," Astrophysical Journal, vol. 675, pp. 1254-1277, 2008. Software vailable online at: http://www.cfa.harvard.edu/ jhartman/vartools/.

[5] P. Lenz and M. Breger, "Period04 User Guide," Communications in Asteroseismology, vol. 146, pp. 53-136, 2005. Available online at: http://www.univie.ac.at/tops/CoAst/archive/cia146.pdf.

[6] R. Li, W. Liu and J. C. Principe, "A Unifying Criterion for Blind Source Separation Based on Correntropy," Signal Processing, vol. 87, no. 8, pp. 1872-1881, August, 2007.

[7] W. Liu, P. P. Pokharel and J. C. Principe, "Correntropy: Properties and Applications in Non-Gaussian Signal Processing," IEEE Transactions on Signal Processing, vol. 55, no. 11, pp. 5286-5298, November, 2007.

[8] N. R. Lomb, "Least-Squares Frequency Analysis of Unequally Spaced Data," Astrophysics and Space Science, vol. 39, pp. 447-462, February, 1976.

[9] J. D. Scargle, "Studies in Astronomical Time Series Analysis. II. Statistical Aspects of Spectral Analysis of Unevenly Spaced Data," The Astrophysical Journals, vol. 263, pp. 835-853, December, 1982.

[10] A. Gunduz and J. C. Principe, "Correntropy as a Novel Measure for Nonlinearity Tests," Signal Processing, vol. 89, pp. 147-23, 2009.

[11] M. Petit, Variable Stars (New York: Wiley), 1987.

[12] W. H. Press and G.B. Rybicki, "Fast algorithm for spectral analysis of unevenly sampled data," The Astrophysical Journal, vol. 338, pp. 277-280, 1989.

[13] W. H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, Numerical Recipes in C, 2nd ed. (New York: Cambridge University Press), 1992.

[14] P. Protopapas, J. M. Giammarco, L. Faccioli, M. F. Struble, R. Dave and C. Alcock, "Finding Outlier Light Curves in Catalogues of Periodic Variable Stars, " Monthly Notices of the Royal Astronomical Society, vol. 369, pp. 677-696, June, 2006.

[15] I. Santamaría, P. P. Pokharel and J. C. Principe, "Generalized Correlation Function: Definition, Properties, and Application to Blind Equalization," IEEE Transactions on Signal Processing, vol. 54, no. 6, pp. 2187-2197, June, 2006.

[16] A. Schwarzenberg-Czerny,"On the advantage of using analysis of variance for period search," Monthly Notices of the Royal Astronomical Society (MNRAS) , vol. 241, pp. 153-165, 1989.

[17] G. Wachman, R. Khardon, P. Protopapas and C. Alcock,"Kernels for Periodic Time Series Arising in Astronomy," Proceedings of the European Conference on Machine Learning, Lecture Notes in Computer Science, Vol. 5782, pp. 489-505, 2009.

[18] J.-W. Xu and J. C. Principe, "A Pitch Detector Based on a Generalized Correlation Function," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 8, pp. 1420-1432, November, 2008.

[19] T.-F. Wu, C.-J. Lin and R. C. Weng, "Probability Estimates for Multi-Class Classification by Pairwise Coupling," Journal of Machine Learning Research, vol. 5, pp. 975-1005, 2004.