An Information Theoretic Algorithm for Finding Periodicities in Stellar Light Curves

Pablo Huijse, *Student Member, IEEE*, Pablo A. Estévez, *Senior Member, IEEE*, Pavlos Protopapas, Pablo Zegers, *Senior Member, IEEE*, and José C. Príncipe, *Fellow, IEEE*

Abstract—We propose a new information theoretic metric for finding periodicities in stellar light curves. Light curves are astronomical time series of brightness over time, and are characterized as being noisy and unevenly sampled. The proposed metric combines correntropy (generalized correlation) with a periodic kernel to measure similarity among samples separated by a given period. The new metric provides a periodogram, called Correntropy Kernelized Periodogram (CKP), whose peaks are associated with the fundamental frequencies present in the data. The CKP does not require any resampling, slotting or folding scheme as it is computed directly from the available samples. CKP is the main part of a fully-automated pipeline for periodic light curve discrimination to be used in astronomical survey databases. We show that the CKP method outperformed the slotted correntropy, and conventional methods used in astronomy for periodicity discrimination and period estimation tasks, using a set of light curves drawn from the MACHO survey. The proposed metric achieved 97.2% of true positives with 0% of false positives at the confidence level of 99% for the periodicity discrimination task; and 88% of hits with 11.6% of multiples and 0.4% of misses in the period estimation task.

Index Terms—Correntropy, information theory, period detection, period estimation, time series analysis, variable stars.

I. INTRODUCTION

A light curve represents the brightness of a celestial object as a function of time (usually the magnitude of the star in the visible part of the electromagnetic radiation). Light curve analysis is an important tool in astrophysics used for estimation of stellar masses and distances to astronomical objects. By analyzing the light curves derived from the sky surveys, astronomers can perform tasks such as transient event detection, variable star detection and classification.

There are a certain types of variable stars [1] whose brightness varies following regular cycles. Examples of this kind of stars are the pulsating variables and eclipsing binary stars. Pulsating stars, such as Cepheids and RR Lyrae, expand and contract periodically effectively changing their size, temperature and brightness. Eclipsing binaries, are systems of two stars with a common center of mass whose orbital plane is aligned to Earth. Periodic drops in brightness are observed due to the mutual eclipses between the components of the system. Although most stars have at least some variation in luminosity, current ground based survey estimations indicate that 3% of the stars varying more than the sensitivity of the instruments and $\sim 1\%$ are periodic [2].

Detecting periodicity and estimating the period of stars is of high importance in astronomy. The period is a key feature for classifying variable stars [3], [4]; and estimating other parameters such as mass and distance to Earth [5]. Period finding in light curves is also used as a means to find extrasolar planets [6]. Light curve analysis is a particularly challenging task. Astronomical time series are unevenly sampled due to constraints in the observation schedules, the day-night cycle, bad weather conditions, equipment positioning, calibration and maintenance. Light curves are also affected by several noise sources such as light contamination from other astronomical sources near the line of sight (sky background), the atmosphere, the instruments and particularly the CCD detectors, among others. Moreover, spurious periods of one day, one month and one year are usually present in the data due to changes in the atmosphere and moon brightness. Another challenge of light curve analysis is related to the number and size of the databases being build by astronomical surveys. Each observation phase of astronomical surveys such as MACHO [7], OGLE [8], SDSS [9] and Pan-STARRS [10] have captured tens of millions of light curves. Soon to arrive survey projects such as the LSST [11] will collect approximately 30 terabytes of data per night which translates into databases of 10 billion stars.

Currently, most periodicity finding schemes used in astronomy are interactive and/or rely somehow on human visual inspection. This calls for automated and efficient computational methods capable of performing robust light curve analysis for large astronomical databases.

In this paper we propose the Correntropy Kernelized Periodogram (CKP), a new metric for finding periodicities. The

CKP combines the information theoretic learning (ITL) concept of correntropy [12] with a periodic kernel. The proposed metric yields a periodogram whose peaks are associated with the fundamental frequencies present in the data. A statistical criterion, based on the CKP, is used for periodicity discrimination. We demonstrate the advantages of using the CKP by comparing it with conventional spectrum estimators and related methods in databases of astronomical light curves.

II. RELATED METHODS

Several methods have been developed to cope with the characteristics of light curves. The most widely used are the Lomb-Scargle (LS) periodogram [13], [14], epoch folding, analysis of variance (AoV) [15], string length (SL) methods [16], [17], and the discrete or slotted autocorrelation [18], [19]. For the LS and AoV periodograms statistical confidence measures have been developed to assess periodicity detection besides estimating the period.

The LS periodogram is an extension of the conventional periodogram for unevenly sampled time series. A sample power spectrum is obtained by fitting a trigonometric model in a least squares sense over the available randomly sampled data points. The maximum of the LS power spectrum corresponds to the angular frequency whose model best fits the time series. In epoch folding a trial period P_t is used to obtain a phase diagram of the light curve by applying the modulus (mod) transformation of the time axis:

$$\phi_i(P_t) = \frac{t_i \mathrm{mod}\ P_t}{P_t},$$

where t_i are the time instants of the light curve. The trial period P_t is found by an *ad-hoc* method, or simply corresponds to a sweep among a range of values. This transformation is equivalent to dividing the light curve in segments of length P_t and then plotting the segments one on top of another, hence folding the light curve. If the true period is used to fold the light curve, the periodic shape will be clearly seen in the phase diagram. If a wrong period is used instead, the phase diagram won't show a clear structure and it will look like noise.

In AoV [15] the folded light curve is binned and the ratio of the within-bins variance and the between-bins variance is computed. If the light curve is folded using its true period, the AoV statistic is expected to reach a minimum value. In SL methods, the light curve is folded using a trial period and the sum of distances between consecutive points (string) in the folded curve is computed. The true period is estimated by minimizing the string length on a range of trial periods. The true period is expected to yield the most ordered folded curve and hence the minimum total distance between points. In slotted autocorrelation [18], [19], time lags are defined as intervals or slots instead of single values. The slotted autocorrelation function at a certain time lag slot is computed by averaging the cross product between samples whose time differences fall in the given slot.

All the related methods described above are based on second-order statistic analysis. Information theoretic based criteria extract information from the probability density function, therefore it includes higher-order statistical moments present in the data. This usually implies a better modelling of the underlying process and robustness to noise and outliers.

The slotting technique was extended to the information theoretic concept of correntropy in [20]. The slotted correntropy estimator was compared with the other mentioned techniques on period estimation of light curves from the MACHO survey, performing equally well on Cepheid/RR Lyrae and much better in eclipsing binaries period estimation. However, the slotted technique has the drawback that is highly dependant on the slot size.

III. BACKGROUND ON ITL METHODS AND PERIODIC KERNELS

A. Generalized Correlation Function: Correntropy

In [12], [21] an information theoretic functional capable of measuring the statistical magnitude distribution and the time structure of random processes was introduced. The generalized correlation function (GCF) or correntropy measures similarities between feature vectors separated by a certain time delay in input space. The similarities are measured in terms of inner products in a high-dimensional kernel space. For a random process $\{X_t, t \in T\}$ with T being an index set, the correntropy function is defined as

$$V(t_1, t_2) = \mathbb{E}_{x_{t_1} x_{t_2}}[\kappa(x_{t_1}, x_{t_2})], \qquad (1)$$

and the centered correntropy is defined as

$$U(t_1, t_2) = \mathbb{E}_{x_{t_1} x_{t_2}}[\kappa(x_{t_1}, x_{t_2})] - \mathbb{E}_{x_{t_1}} \mathbb{E}_{x_{t_2}}[\kappa(x_{t_1}, x_{t_2})] \quad (2)$$

where $\mathbb{E}[\cdot]$ denotes the expectation value and $\kappa(\cdot, \cdot)$ is any positive definite kernel [12]. A kernel can be viewed as a similarity measure for the data [22]. The Gaussian kernel which is translation-invariant, is defined as follows:

$$G_\sigma(x - z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \qquad (3)$$

where σ is the kernel size or bandwidth. The kernel size can be interpreted as the resolution in which the correntropy function search for similarities in the high-dimensional kernel feature space [12], [21]. The kernel size gives the user the ability to control the emphasis given to the higher-order moments with respect to second-order moments. For large values of the kernel size, the second-order moments have more relevance and the correntropy function approximates the conventional correlation. On the other hand if the kernel size is set too small, the correntropy function will not be able to discriminate between signal and noise and approximates the Dirac delta function.

The name correntropy was coined due to its close relation to Renyi's quadratic entropy, which can be estimated through Parzen windows [12] as follows:

$$\hat{H}_{R2}(X) = -\log(IP_\sigma(X)) \qquad (4)$$

where

$$IP_\sigma(X) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} G_\sigma(x_i - x_j) \qquad (5)$$

and N is the number of samples of the random variable X and σ is the kernel size of the Gaussian kernel function. Equation (5) is the argument of the logarithm in (4) and is called the Information Potential (IP). The mean value of the correntropy function over the lags is a biased estimator of the IP [12].

For a discrete strictly stationary random process $\{X_n\}$, the univariate correntropy function or autocorrentropy is defined as

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

which can be estimated through the sample mean

$$\widehat{V}_\sigma[m] = \frac{1}{N-m+1} \sum_{n=m}^{N} G_\sigma(x_n - x_{n-m}). \qquad (6)$$

Likewise, the estimator of the univariate centered correntropy function (2) is

$$\widehat{U}_\sigma[m] = \frac{1}{N-m+1} \sum_{n=m}^{N} G_\sigma(x_n - x_{n-m})$$
$$- \frac{1}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} G_\sigma(x_n - x_m), \quad (7)$$

where the Gaussian kernel with kernel size σ is used, N is the number of samples of $\{X_n\}$ and $m \in [1, N]$ is the discrete time lag. In practice, the maximum lag should be chosen so that there are enough samples to estimate correntropy at each lag. Notice that the second term in (7) corresponds to the IP (5), which is the mean of the autocorrentropy function over the lags.

The Fourier transform of the centered autocorrentropy function is called correntropy spectral density (CSD) and is defined as [12], [21], [23]:

$$P_\sigma[f] = \sum_{m=-\infty}^{\infty} \widehat{U}_\sigma[m] \cdot \exp\left(-j2\pi f \frac{m}{F_s}\right) \qquad (8)$$

where $\widehat{U}_\sigma[m]$ is the univariate centered correntropy function, and F_s is the sampling frequency. The CSD can be considered as a generalized power spectral density (PSD) function, although it is a function of the kernel size and it does not measure power. As with correntropy, the kernel size controls the influence of the higher-order moments versus the second-order statistical descriptors. Particularly, for large values of the kernel size, the CSD approximates the conventional PSD.

In [23] the correntropy function and the CSD were used to solve the problem of detecting the fundamental frequency in speech signals. Correntropy outperformed conventional correlation, showing better discriminatory and robustness to noise. In [24] correntropy was used to solve the blind source separation (BSS) problem, successfully separating signals coming from independent and identically distributed sources and also Gaussian sources. Correntropy outperformed methods that also make use of higher-order statistics such as Independent Component Analysis (ICA). In [25], correntropy was used as a discriminatory metric for the detection of nonlinearities in time series, outperforming traditional methods such as the Lyapunov exponents.

B. Periodic Kernel Functions

Kernels can also be viewed as covariance functions for correlated observations at different points of the input domain [22]. In our research we are interested in measuring similarities among samples separated by a given period. A kernel function is periodic with period P if it repeats itself for input vectors separated by P. Periodic kernel functions are appropriate for nonparametric estimation, modelling and regression of periodic time series [26]. Periodic kernel functions have also been proposed in the Gaussian processes literature [27]–[29].

A periodic kernel function can be obtained by applying a nonlinear mapping (or warping) $u(z)$ to the input vector z. In [28] a periodic kernel function was constructed by mapping a unidimensional input variable z using a periodic two-dimensional warping function defined as

$$u_P(z) = \left(\cos\left(\frac{2\pi}{P}z\right), \sin\left(\frac{2\pi}{P}z\right)\right). \qquad (9)$$

The periodic kernel function $G_{P;\sigma}(z-y)$ with period P, is obtained by applying the warping function (9) to the inputs of the Gaussian kernel function (3). The periodic kernel function is defined as

$$G_{\sigma;P}(z-y) = G_\sigma(u_P(z) - u_P(y))$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\sin^2\left(\frac{\pi}{P}(z-y)\right)}{0.5\sigma^2}\right) \quad (10)$$

where the following expression is used

$$\|u_P(z) - u_P(y)\|^2 = 4\sin^2\left(\frac{\pi(z-y)}{P}\right).$$

The periodic kernel function (10) is related to the von Mises probability density function [30].

IV. CORRENTROPY KERNELIZED PERIODOGRAM

In this paper we propose an ITL based method for finding periodicities in unevenly sampled time series. The proposed method does not require any resampling, slotting or folding scheme, as it is computed directly from the available samples and detects periodicity using the actual magnitudes and time instants of the samples. The new metric combines the centered correntropy function and the periodic kernel function. For a discrete unidimensional random process $\{X_n\}$ with $n = 1, \ldots, N$, kernel sizes σ_t and σ_m, and a trial period P_t, the proposed metric is computed as

$$\widehat{V}_{P\{\sigma_t,\sigma_m\}}(P_t) = \frac{\sigma_m}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (G_{\sigma_m}(\Delta x_{ij}) - IP)$$
$$\cdot G_{\sigma_t;P_t}(\Delta t_{ij}) \cdot w(\Delta t_{ij}) \quad (11)$$

where $\Delta x_{ij} = x_i - x_j$, $\Delta t_{ij} = t_i - t_j$, $G_{\sigma_m}(\cdot)$ is the Gaussian kernel function (3), IP is the information potential (5), $G_{\sigma_t;P_t}(\cdot)$ is the periodic kernel function (10) and $w(\Delta t_{ij})$ is a Hamming window.

In (11) magnitude differences are evaluated using the Gaussian kernel function, while time differences are evaluated using the periodic kernel function. The new metric performs the pointwise multiplication between the centered Gaussian kernel coefficient $(G_{\sigma_m}(\Delta x_{ij}) - IP)$ and the periodic kernel coefficients $G_{\sigma_t; P_t}(\Delta t_{ij})$. Notice that (11) is a function of the trial period P_t and it has two free parameters: the magnitude kernel size σ_m and the time kernel size σ_t. By analogy with the conventional periodogram (square of the discrete Fourier transform of the data) we call (11) the Correntropy Kernelized Periodogram (CKP). Another extension of the conventional theory is the wavelet periodogram [31]. A Hamming window defined as

$$w(\Delta t_{ij}) = 0.54 + 0.46 \cdot \cos\left(\frac{\pi \Delta t_{ij}}{T}\right), \qquad (12)$$

where T is the total time span of the light curve, is used in (11) to have a smoother estimation of the periodogram [32].

The periodic kernel gives larger weights to the samples whose time difference are multiples of the trial period. In other words, the periodic kernel emphasizes the magnitude differences that are separated by the trial period and its multiples. Intuitively, for a periodic time series, the magnitude differences between samples separated by the underlying period in time are expected to be minimum (most similar). If the trial period of the periodic kernel is set to the true period, the metric is expected to reach a maximum value. By using the maximum value of (11) over a set of trial periods, the best estimation of the underlying period is obtained.

In order to compare periodograms from different time series, the CKP has to be invariant to the data scale. In [12] a scale-invariant criterion based on Renyi's quadratic entropy is proposed. The condition for Gaussian variables is that the magnitude's kernel size should be directly proportional to the spread of the data. Herein, we use this condition as an approximate method to make the representation scale invariant. In (11) we set σ_m as follows,

$$\sigma_m(X_n) = k \cdot \min\left\{0.7413 \, \mathrm{iqr}(X_n), \mathrm{std}(X_n)\right\}, \qquad (13)$$

where k is a constant, iqr is the interquartile range[1] and std is the standard deviation [12]. The iqr is a measure of statistical dispersion, being equal to the difference between the third and first quartiles of the data. The first and third quartiles are the medians of the first half and second half of the rank-ordered data, respectively. For a normal distribution the standard deviation is equal to $0.7413 \cdot \mathrm{iqr}$. The selection of constant k is discussed in Section VI-A.

Fig. 1 shows an example using a synthetic time series to illustrate the effect of the proposed metric. Fig. 1(a) shows a synthetic time series $y_i = \sin\left(\frac{2\pi t_i}{P}\right) + 0.8 \cdot \epsilon_i$, with $t_i = \frac{T_{max}}{N}(i + 0.5 \cdot \varepsilon_i)$, where ϵ_i and ε_i are normally distributed random variables with zero mean and unit standard deviation. The noise in time simulates uneven sampling. In this example $N = 400$, $T_{max} = 25$, and the underlying period is $P = 2.456$ seconds. Fig. 1(b) shows the kernel coefficients
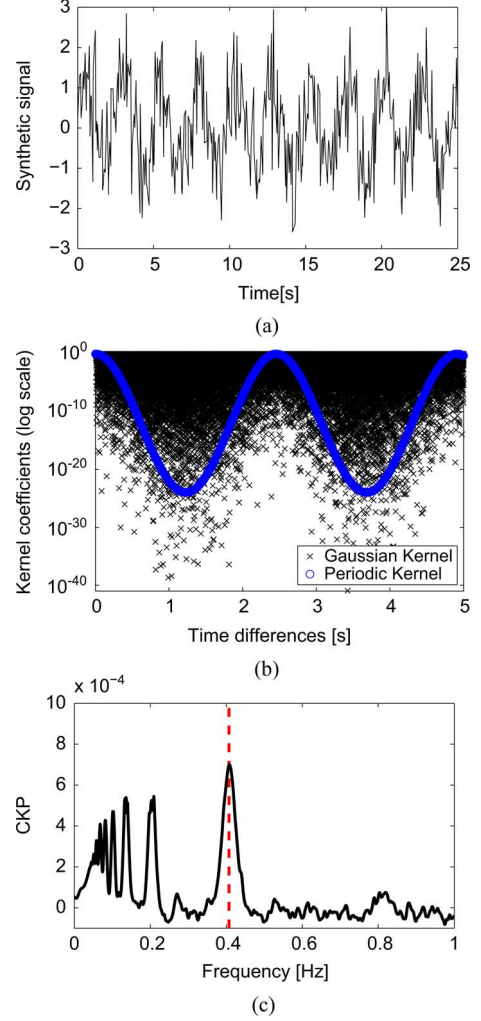

Fig. 1. (a) Synthetic time series $\sin\left(\frac{2\pi t}{P}\right)$ with period $P = 2.456$ seconds plus Gaussian noise. The time instants have been randomly perturbed to simulate uneven sampling. (b) Kernel coefficients, $(G_{\sigma_m}(\Delta x_{ij}(\Delta t_{ij})) - IP)$ and $G_{\sigma_t; P_t}(\Delta t_{ij})$, as a function of the time differences. The CKP is the pointwise product between the centered Gaussian kernel coefficients and the periodic kernel coefficients. (c) CKP as a function of the trial frequency, the dotted line marks the location of the true period. The global maximum of the CKP corresponds to the underlying period.

$(G_{\sigma_m}(\Delta x_{ij}(\Delta t_{ij})) - IP)$ and $G_{\sigma_t; P_t}(\Delta t_{ij})$ as a function of the time differences collected from the time series. The magnitude kernel size is set using (13) with $k = 0.3$. The time kernel size and the trial period are set to $\sigma_t = 0.1$, $P_t = 2.456$, respectively. Fig. 1(c) shows the CKP for a range of periods, the location of the underlying period in the periodogram is marked with a dotted line. The CKP reaches a global maximum at the corresponding underlying period $P = 2.456$.

A. A Statistical Test Based on the CKP For Periodicity Discrimination

For a periodic time series with an oscillation frequency f, its periodogram will exhibit a peak at that frequency with high probability. But the inverse is not necessarily true, a peak in the periodogram does not imply that the time series is periodic. Spurious peaks may be produced by measurement errors, random fluctuations, aliasing or noise.

[1]If the data does not follow a normal distribution or it contains outliers, the interquartile range will provide a better spread estimation than the standard deviation because it uses the middle 50% of the data

In this section, a statistical test for periodicity is introduced, using the global maximum of the CKP as test statistic. The null hypothesis is that there are no significant periodic components in the time series. The alternative hypothesis is that the CKP maximum corresponds to a true periodicity. The distribution of the maximum value of the CKP is obtained through Monte Carlo simulations. Surrogate time series [33], [34] are used to test the null hypothesis. The surrogate generation algorithm has to be consistent with the null hypothesis. To achieve this, the block bootstrap method [35], which breaks periodicities preserving the noise characteristic and time correlations of the light curve, is used. The procedure used to construct an unevenly sampled surrogate using the block bootstrap method is as follows

1) Obtain a data block from the light curve by randomly selecting a block of length L and a starting point $j \in [1, N - L]$.
2) Subtract the first time instant of the block, so that it starts at 0 days.
3) Add the value of the last time instant of the previous block to the time instants of the current block.
4) Parse the current block to the surrogate time series.
5) Repeat steps 1–4 until the surrogate time series have the same amount of samples of the original light curve.

For a given significance level α and kernel sizes σ_t and σ_m, the null hypothesis is rejected if

$$\max_f \widehat{V}_{p\{\sigma_m, \sigma_t\}}(f) > \widehat{V}^\alpha_{p\{\sigma_m, \sigma_t\}},$$

where for N light curves $\widehat{V}^\alpha_{p\{\sigma_m, \sigma_t\}}$ is pre-computed as follows:
1) Generate M surrogates from each light curve using block bootstrap.
2) Save the maximum CKP ordinate value of each surrogate.
3) Find P_α such that a $(1 - \alpha)\%$ of the ordinate values saved from the surrogates are below this threshold (one-tailed distribution).
4) Compute $\widehat{V}^\alpha_{p\{\sigma_m, \sigma_t\}}$ as the mean P_α and its corresponding error bars as the standard deviation of P_α for the N light curves ($N \cdot M$ surrogates).

V. PERIOD DETECTION METHOD

A. Description of the MACHO Database

The MACHO project [7] was designed to search for gravitational microlensing events in the Magellanic Clouds and the galactic bulge. The project started in 1992 and concluded in 1999. More than 20 million stars were surveyed. The MACHO project has been an important source for finding variable stars. The complete light curve database is available through the MACHO project's website[2]. There are two light curves per stellar object: channels blue and red. Only the blue channel light curves are used here. Each light-curve has approximately 1000 samples and contains 3 data columns: time, magnitude and an error estimation for the magnitude.

Astronomers from the Harvard Time Series Center (TSC) have a catalog of variable stars from the MACHO survey. The underlying periods of the periodic variable stars were estimated

using epoch folding, AoV, and visual inspection. In this paper, we consider the TSC periods to be the gold standard.

A subset of 1500 periodic light curves (500 Cepheids, 500 RR Lyrae, and 500 eclipsing binaries) and 3500 non-periodic light curves was drawn from the MACHO survey. The subset was divided into a training set for parameter adjustment and a testing set. The training set consisted of 2500 light curves (750 periodic and 1750 non periodic) randomly selected from the available classes. The remaining 2500 light curves were used for testing purposes.

There is a natural imbalance between periodic and aperiodic classes of stars. Only 3% of the surveyed stars are expected to be variables and $\sim 1\%$ to be periodic. Due to this, when detecting periodic behaviour, we have to achieve a false positive rate less than 0.1%.

B. Description of the Procedure for Periodicity Detection

In what follows, the steps of the periodicity detection algorithm, for a given time kernel size σ_t and magnitude kernel size σ_m, are described.
1) **Cleaning**: The light curve's blue channel is imported. The mean $\bar{e}$ and the standard deviation σ_e of the photometric error are computed. Samples that do not comply with $e_i < \bar{e} + 3 \cdot \sigma_e$, where e_i is the photometric error of sample i, are discarded.
2) **Computing the periodogram**: The CKP (11) is computed on 20000 logarithmically spaced periods between 0.4 days and 300 days. The periods associated to the ten highest local maxima at the periodogram are saved as trial periods for the next step.
3) **Fine-tuning of trial periods**: The CKP is used to fine tune the ten trial periods. Each trial period is fine-tuned around a 0.5% of its value ($[1.0025 \cdot f_{trial}, 0.9975 \cdot f_{trial}]$), using a step size of $df = \frac{0.01}{T}$ in frequency, where T is the total time span of the light curve.
4) **Selection of the best trial period**: The trial periods are sorted in descending order following its CKP ordinate value. The best trial period P_{best} is selected as the one with the highest value of $\widehat{V}_P$, that is not a multiple of a spurious period, as described below.
5) Finally, if the best period comply with $\widehat{V}_{P\{\sigma_m, \sigma_t\}}(P_{best}) > \theta$ then the light curve is labeled as periodic, where θ is the periodogram threshold for periodicity. The confidence associated to θ is obtained using the procedure described in Section IV-A.

Table I gives a summary of the parameters of the proposed method. The kernel sizes and the periodicity threshold do not appear in Table I, because they need to be calibrated using a procedure described in the following sections.

To obtain the spurious periods the following spectral window function is used

$$W(f) = \frac{1}{N} \left| \sum_{i=1}^{N} \exp\left(j2\pi f t_i\right) \right|^2, \quad (14)$$

where t_i with $i = 1, \ldots, N$ are the time instants of the light curve. Equation (14) is the periodogram of the sampling pattern of the light curve. The frequencies associated to the peaks of the

TABLE I
SUMMARY OF PARAMETERS OF THE PERIOD DETECTION PIPELINE

Description	Value
Minimum period	0.4 days
Maximum period	300 days
Periodogram resolution	20.000 periods[3]
Number of fine-tuned trial periods	10
Fine-tune resolution	$\frac{0.01}{T}$ periods[4]

[3]Logarithmically spaced.

[4]Linearly spaced, T is the total time span of the light curve.

spectral window are related to spurious periodicities caused by the sampling. In most cases, the spurious periods obtained from the spectral window are multiples of the sidereal day (0.99727 days) and yearly periodicities (365.25 days). The moon phase period (29.53 days) is also added to the list of spurious periods. The moon phase has no relation to the sampling but it is intrinsic to the data. The daily sampling produces alias peaks of the true periods (P) in the CKP at $\left(\frac{1}{P+k} \right) \frac{1}{\text{days}}$, where $k \in \mathbb{Z}$. The CKP values of the aliases are very low and therefore they do not need to be filtered out.

C. Computational Time

The algorithm[3] was programmed in C/CUDA and implemented in a Graphical Processing Unit (GPU) NVIDIA Tesla C2050 with 448 cuda cores. The computational time required to process one light curve (1000 samples) is 1.812 seconds. The total time to process the 5000 light curves is 2.5 hours.

D. Performance Criteria

In what follows we define performance measures for the problems of period detection and period estimation. The former consists of discriminating between periodic and non periodic light curves. The latter consists in estimating the true periods of periodic light curves. For the period estimation problem the classification is done by using the TSC periods as golden standard. An estimated period P_{est} is classified as either a Hit, a Multiple or a Miss with respect to the reference period P_{ref} according to the following criteria:

- Hit if $|P_{\text{ref}} - P_{est}| < \varepsilon \cdot P_{\text{ref}}$
- Multiple if $P_{est} > P_{\text{ref}}$ and

$$\left| \frac{P_{est}}{P_{\text{ref}}} - \left\lfloor \frac{P_{est}}{P_{\text{ref}}} \right\rfloor \right| < \varepsilon,$$

  or if $P_{est} < P_{\text{ref}}$ and

$$\left| \frac{P_{\text{ref}}}{P_{est}} - \left\lfloor \frac{P_{\text{ref}}}{P_{est}} \right\rfloor \right| < \varepsilon,$$

  where $\lfloor x \rceil$ is the largest integer less than or equal x.
- Miss if it does not belong to any of the other categories.

The tolerance value ε controls the accepted relative error between the estimated period and the reference period. A value of $\varepsilon = 0.005$, i.e., a relative error of 0.5% will be considered, small enough to obtain a clean folded curve from the estimated period.

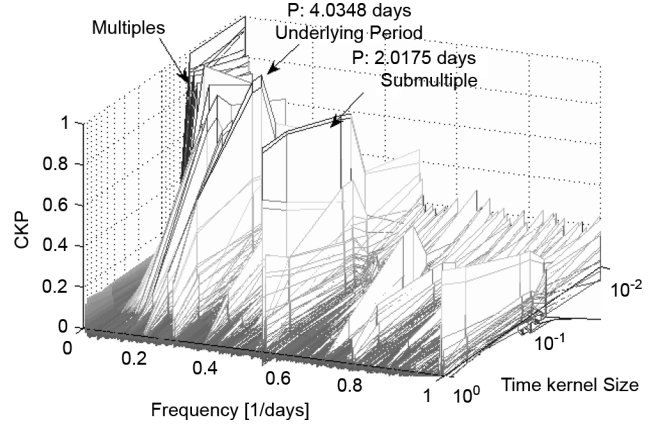[3]The C-language implementation is available by request to the authors.

Fig. 2. CKP versus frequency and time kernel size σ_t for light curve 1.3449.27 from the MACHO catalog. The underlying period of 4.0348 days is found as the maximum of the CKP with $\sigma_t \in [0.075, 0.125]$.

The problem of period detection in light curves can be treated as a binary classification problem. The classes are periodic light curves $\{+1\}$ and non-periodic light curves $\{-1\}$. Confusion matrix and Receiver Operating Characteristic (ROC) curves are used to evaluate the period detection method. An ROC curve is a plot of the true positive rate (TPR) as a function of the false positive rate (FPR). Different points in the ROC curve are obtained by changing the threshold value at the output of the classifier. In this case,

- true positive (TP) is a periodic light curve classified as periodic
- false positive (FP) is a non-periodic light curve classified as periodic
- true negative (TN) is a non-periodic light curve classified as non-periodic
- false negative (FN) is a periodic light curve classified as non-periodic

The TPR represents the proportion of periodic light curves that are correctly identified as such. The FPR represents the proportion of non-periodic light curves that are incorrectly classified as periodic.

VI. EXPERIMENTS

A. Parameter Calibration

The CKP is a function of the kernel sizes σ_m and σ_t. These parameters are adjusted using the 2500 light curves in the training set. The value of σ_m is set using (13), so we need to choose the value of the constant k. Fig. 2 shows the CKP as function of the frequency and σ_t for light curve 1.3449.27 from the MACHO catalog. In this example the underlying period is picked as the global maximum of the CKP for a time kernel size $\sigma_t \in [0.075, 0.125]$. After extensive experiments we identified that this particular range of kernel sizes values gives the best results for the MACHO light curves. There is no clear rule for choosing the time kernel size, although intuitively, it should depend on the sampling pattern.

The period estimation results for different combinations of k and σ_t on the 750 periodic light curves of the training set are

TABLE II
PERIOD ESTIMATION PERFORMANCE OF THE CKP USING DIFFERENT
COMBINATIONS OF KERNEL SIZES FOR THE TRAINING DATABASE

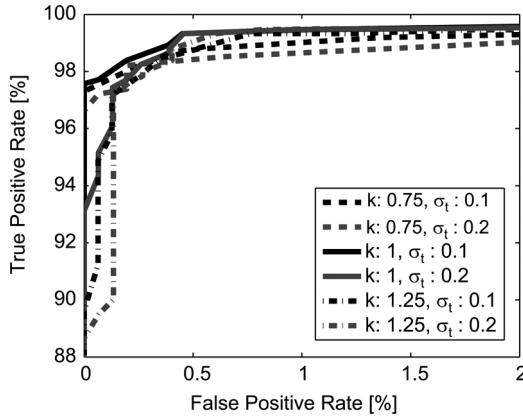| k	σ_t	Hits[%]	Multiples[%]	Misses[%]
0.75	0.05	75.47	23.47	1.07
0.75	0.1	86.13	13.60	0.27
0.75	0.15	85.33	13.87	0.80
1	0.05	82.27	16.93	0.80
1	**0.1**	**88.67**	**11.20**	**0.13**
1	0.15	86.93	12.67	0.40
1.25	0.05	76.40	22.13	1.47
1.25	0.1	85.73	14.13	0.13
1.25	0.15	85.47	14.00	0.53

Fig. 3. Receiver operating characteristic curves for the period detection problem using different combinations of kernel sizes in the training database.

TABLE III
STATISTICAL SIGNIFICANCE THRESHOLDS OF THE CKP WITH
$\sigma_t = 0.1$ AND $k = 1$

$1 - \alpha$	$\widehat{V}_P^\alpha$
0.99	$3.59e - 4$
0.95	$3.12e - 4$
0.90	$2.80e - 4$

shown in Table II. The best performance is obtained with $k = 1$ and $\sigma_t = 0.1$.

Fig. 3 compares the ROC curves obtained using different combination of both kernel sizes for the period detection problem. As mentioned before, due to the imbalance between periodic and non-periodic light curve classes, false positive rates below 0.1% are desired. Looking at the ROC curves it is clear that the best kernel size combination, in the area below 1% FPR, is $k = 1$ and $\sigma_t = 0.1$. These values are fixed for the following experiments.

B. Statistical Significance

Using the procedure described in Section IV-A statistical significance thresholds for the CKP were computed. Table III shows the significance thresholds and their corresponding CKP ordinate values for the best combination of kernel sizes ($\sigma_t = 0.1$ and $k = 1$). The thresholds were computed using the light curve training set ($N = 2500$) and five hundred surrogates per light curve ($M = 500$). Fig. 4 shows the location of these thresholds in the ROC curve of the testing dataset. FPR rates below 1% are associated with confidence levels between 95% and 99%. Fig. 5 shows three light curves in which the CKP ordinate value associated with the fundamental period has a
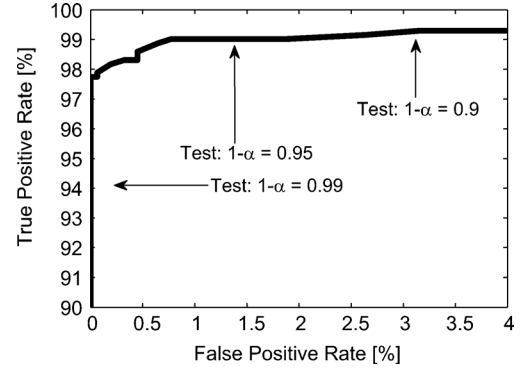

Fig. 4. ROC curve of the CKP with $\sigma_t = 0.1$ and $k = 1$ in the test subset. The significance thresholds of the CKP are shown in the ROC curve.
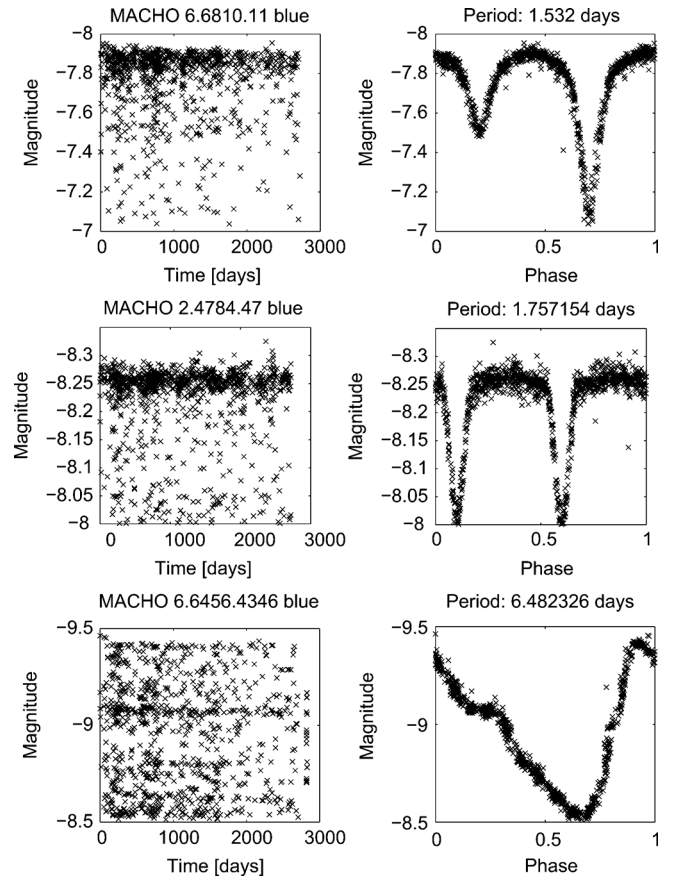

Fig. 5. Examples of periodic light curves detected by the CKP method with a level of confidence greater than 99%. The original light curves are shown on the left column. The right column shows the same light curves folded with the estimated period.

confidence level higher than 99%. In the folded light curves (Fig. 5 right column) the periodic nature of the light curve can be clearly observed. In period detection schemes based on visual inspection these light curves would be undoubtedly labeled as periodic. Fig. 6 shows three light curves in which the CKP ordinate values associated with the fundamental period have a statistical confidence between 90% and 95%. These light curves are indeed periodic although compared to the previous three (shown in Fig. 5), their periodicity is less clear as their signal-to-noise ratio is smaller. By associating a
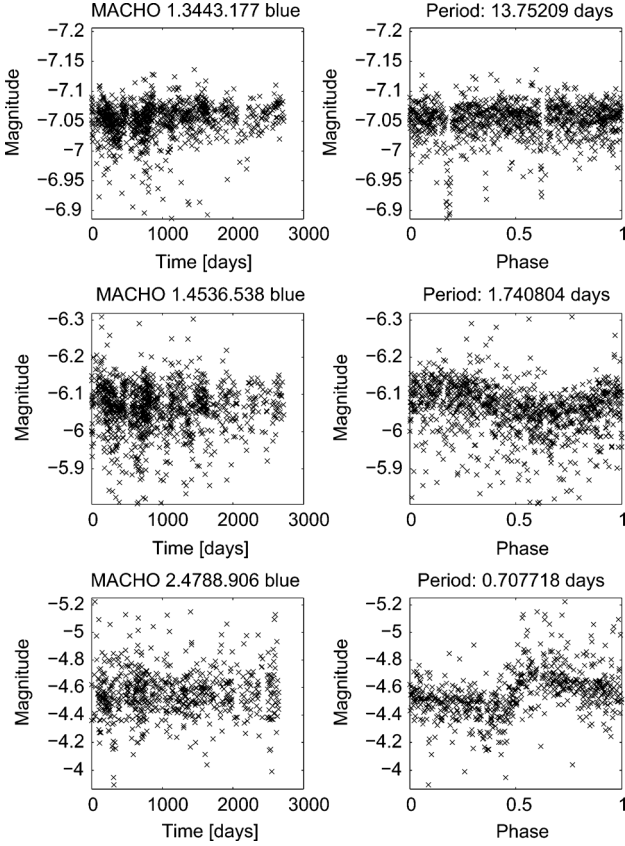
Fig. 6. Examples of periodic light curves detected by the CKP method with a level of confidence between 90% and 95%. The original light curves are shown on the left column. The right column shows the same light curves folded with the estimated period.
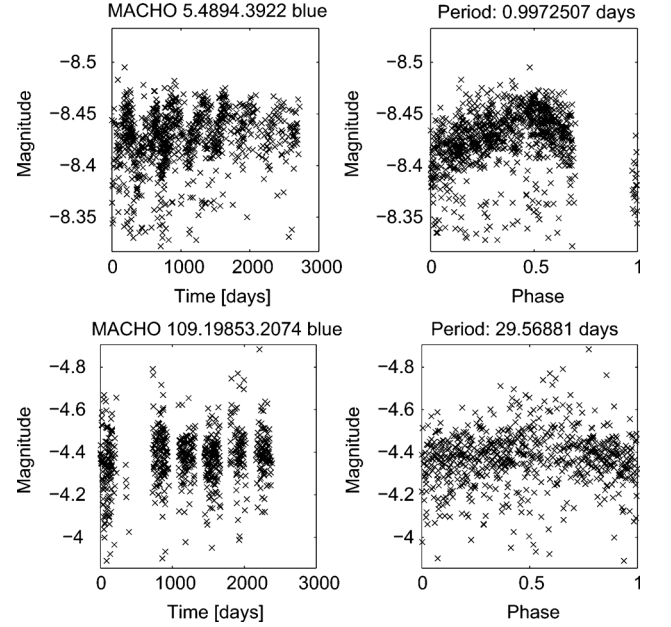

Fig. 7. Examples of spurious periods found among the top peaks of the CKP. The original light curves are shown on the left column. The right column shows the folded light curves with the spurious period. These spurious periods are related to periodic daily changes in the atmosphere and periodic monthly changes in the brightness of the moon. These periods are discarded as spurious by our method.
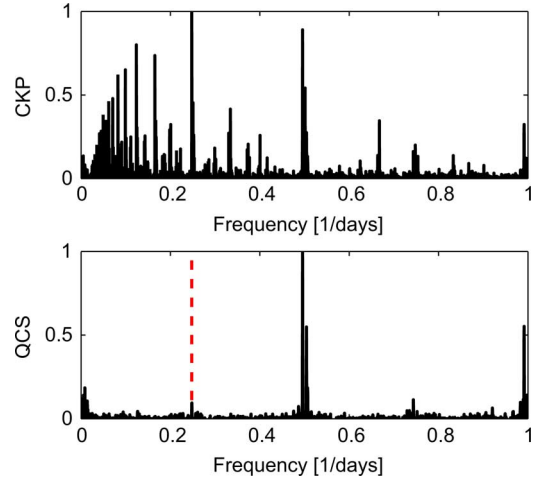

Fig. 8. Comparison between the periodograms obtained using the proposed CKP metric and the QCS. The dotted line marks the true period. The underlying period corresponds to the global maximum of the CKP. On the contrary, the QCS value of the true period is very small.

statistical level of confidence to the detected periods, we have obtained a way to set a period detection threshold and also a better interpretation of period quality. The level of confidence on the detected period could be used in later (post-processing) stages of the period detection pipeline. For example, periods with lower confidence levels may be selected for additional analysis stages in which finer resolution or different parameter combinations may be used. Fig. 7 shows examples of periods associated to the sidereal day (0.99727 days) and moon phase (29.53 days). These periods are discarded as spurious as described in Section V-B.

C. Influence of the Periodic Kernel

In this experiment we demonstrate the advantages of the kernelized periodogram with respect to the conventional spectrum estimator. The quadratic correntropy spectrum (QCS) is defined as follows:

$$\widehat{P}_{\sigma_m}(f) = \sum_{i=1}^{N} \sum_{j=1}^{N} (G_{\sigma_m}(\Delta x_{ij}) - IP)$$
$$\cdot \exp(-j2\pi f \Delta t_{ij}) \cdot w(\Delta t_{ij}) \quad (15)$$

where $\Delta x_{ij} = x_i - x_j$, $\Delta t_{ij} = t_i - t_j$, and IP is the information potential estimator (5). Notice that (15) differs from the CSD definition (8) in two aspects. First the CSD definition uses integer lags, while the QCS uses directly the differ-

ence between samples. This change is useful to deal with unevenly sampled light curves. Secondly, we include in (15) a Hamming window (12) for enhancing the spectrum estimation. The constructed quadratic correntropy spectrum (QCS) is similar to the CKP (11) except that the periodic kernel has been replaced by $\exp(-j2\pi f \Delta t)$ to compute the Discrete Fourier Transform. First, we use an example to illustrate the differences between periodograms using a single light curve, from object 1.3449.27 of the MACHO survey. This light curve corresponds to an eclipsing binary with fundamental period 4.03488 days. Fig. 8 shows the CKP and QCS estimators for light curve 1.3449.27. The CKP was computed using $k = 1$ and $\sigma_t = 0.1$.

TABLE IV
PERIOD ESTIMATION PERFORMANCE OF THE CKP VERSUS THE QCS FOR THE
TESTING DATABASE

Method	Hits[%]	Multiples[%]	Misses[%]
CKP	**88.00**	**11.60**	**0.4**
QCS	77.87	20.80	1.33

The same kernel size was used for the QCS. The dotted line marks the location of the underlying period. The global maximum of the CKP is associated with 4.0349 days, which corresponds to the underlying period of the light curve. Other local maxima appear at multiples and sub-multiples of the underlying period. On the other hand, the global maximum of the QCS is associated to 2.0175 days, the closest integer sub-multiple of the underlying period. The QCS value at the frequency corresponding to the true period is very small, in fact smaller than many other spurious peaks. The procedure described in Section V-B is used to evaluate the performance of QCS estimator. Table IV shows the results obtained by the QCS and CKP estimators on the testing subset. The CKP obtains 12% more hits and 70% less misses than the QCS estimator, which shows clearly the advantages of the kernelized periodogram.

D. Comparison With Other Methods

The performance of the CKP method was compared with the slotted correntropy and other widely used techniques in astronomy. The software VarTools [36], [37] was used to perform a Lomb-Scargle periodogram and Analysis of Variance analysis. The regularized Lafler-Kinman string length (SLLK) statistic and the slotted autocorrelation were also considered. For Vartools LS, the period associated with the highest peak of the LS periodogram, that is not a multiple of the known spurious periods (sidereal day, moon phase), gives the estimated period. A periodogram resolution of $\frac{0.1}{T}$ and a fine tune resolution of $\frac{0.01}{T}$, where T is the total time span of the light curve, were used. For Vartools AoV and SLLK, the corresponding statistics are minimized in an array of periods ranging from 0.4 to 300 days with a step size of 1 e-4. For AoV the default value of 8 bins is used. For AoV and SLLK, the period that minimizes the corresponding metrics, that is not a multiple of the known spurious periods, is selected as the estimated period. For the slotted autocorrelation/autocorrentropy the highest peak of the PSD/CSD estimator function, that is not associated to the known spurious periods, delivers the estimated period. For the slotted autocorrentropy/autocorrelation a slot size $\Delta\tau = 0.25$ was considered [20]. For the CKP method the best combination of kernel sizes ($\sigma_t = 0.1$ and $k = 1$) obtained with the training dataset is used. The influence of the higher-order statistical moments is assessed by comparing the CKP with a linear version of the proposed metric. In this linear version, the Gaussian kernel used to compare magnitude values is replaced by a linear kernel. The periodic kernel remains unchanged. The procedure to detect a period is the same as explained in Section V-B, with an additional pre-processing step where the data vector is zero-mean normalized. The results for period estimation on the testing subset are shown in Table V. The CKP method obtained the highest hit rate (88%), followed by the linear version of the CKP (80%), the slotted correntropy (78%) and the AoV periodogram (75%). The
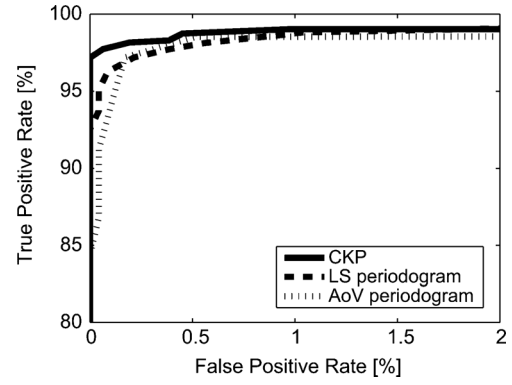

Fig. 9. Receiver operating characteristic curves of the CKP ($k = 1, \sigma_t = 0.1$), the LS periodogram and the AoV periodogram.

TABLE V
PERIOD ESTIMATION PERFORMANCE OF THE CKP METHOD VERSUS OTHER
TECHNIQUES FOR THE TESTING DATABASE

Method	Hits[%]	Multiples[%]	Misses[%]
CKP	**88.00**	**11.60**	**0.4**
CKP (linear kernel)	80.40	18.67	0.93
Slotted correntropy	78.80	19.73	1.47
Slotted correlation	70.00	28.80	1.20
VarTools LS	61.73	36.00	2.27
VarTools AoV	75.33	23.60	1.07
SLLK	71.47	26.27	2.27

CKP obtained 8.6% more hits and 57% less misses than its linear version. This is because the Gaussian kernel incorporates all the even-order moments of the process and gives robustness to outlier data samples which are common in astronomical time series. The CKP obtained 10.5% more hits and 72.7% less misses than the slotted correntropy. This is because in the slotted correntropy, kernel coefficients are averaged on time slots, therefore the actual time differences between samples are not used. Out of the cases where the correct period is found by the CKP but not by the AoV periodogram, an 84% corresponds to eclipsing binaries. This is expected as conventional methods perform well on pulsating variables [20]. Eclipsing binaries light curves are typically more difficult to analyze as their variations are non-sinusoidal and due to their morphology/shape characteristics most methods tend to return harmonics or sub-harmonics rather than the true period. The proposed method obtained the lowest miss rate (0.4%). In all these missed cases[4], the true period was correctly found by the proposed metric but they were filtered out for being too near to the sidereal day. More accurate ways of filtering out spurious periods, using the data samples instead of a straightforward comparison of the detected periods, could be implemented to recover such missed cases. Fig. 9 shows ROC curves for the task of periodic versus non-periodic discrimination in the 2500 light curve testing subset. The CKP is compared with the LS and AoV periodograms. The proposed method clearly outperforms its competitors in the FPR range of interest (below 1%). It is worth noting that even if a harmonic of the true period is found, periodicity can still be detected. This is true for all the methods as periodicity detection rates are comparatively better than Hit rates obtained for period estimation.

[4]Light curves 1.4539.37, 3.6605.124 and 6.5726.1276, with periods 2.9955 (three times sidereal day), 3.9813 (four times the sidereal day) and 0.99676, respectively.

VII. CONCLUSION

We have proposed a new metric for periodicity finding based on information theoretic concepts. The CKP metric yields a kernelized periodogram. It has been shown that the proposed method has several advantages over the correntropy spectral density and other conventional methods of period detection and estimation. The CKP is computed directly from the actual magnitudes and time-instants of samples. It does not require resampling, slotting nor folding schemes as other methods. The CKP metric is used as the main part of a fully-automated pipeline for period detection and estimation in astronomical time series. The CKP metric is used as test statistic to estimate the confidence level of the period detection.

Results on a subset of the MACHO survey shows that the CKP metric clearly outperforms the LS and AoV periodograms in the period detection of light curves. Moreover, the CKP method clearly outperforms slotted correntropy, slotted correlation, LS-periodogram, AoV and SLLK methods on the period estimation of periodic light curves. This is because the CKP method incorporates higher-order moments when computing the periodogram, and it emphasizes the trial period and its multiples. Future work will focus on enhancing the spurious rejection criterion, developing an adaptive kernel size adjusting rule, and discriminating quasi-periodic and multi-periodic light curves.

REFERENCES

[1] M. Petit, *Variable Stars. Reading*. New York: Wiley, 1987.

[2] L. Eyer, "First thoughts about variable star analysis," *Baltic Astron.*, vol. 8, pp. 321–324, 1999.

[3] J. Debosscher, L. Sarro, C. Aerts, J. Cuypers, B. Vandenbussche, R. Garrido, and E. Solano, "Automated supervised classification of variable stars. I methodology," *Astron. Astrophys.*, vol. 475, pp. 1159–1183, 2007.

[4] G. Wachman, R. Khardon, P. Protopapas, and C. Alcock, "Kernels for periodic time series arising in astronomy," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discov. Databases: Part II*, Bled, Slovenia, 2009, pp. 489–505.

[5] D. Popper, "Stellar masses," *Ann. Rev. Astron. Astrophys.*, vol. 18, pp. 115–164, 1980.

[6] P. Protopapas, R. Jimenez, and C. Alcock, "Fast identification of transits from light-curves," *Monthly Notices Roy. Astronom. Soc.*, vol. 362, pp. 460–468, Sep. 2005 [Online]. Available: arXiv:astro-ph/0502301

[7] C. Alcock *et al.*, "The MACHO project: Microlensing results from 5.7 years of LMC observations," *Astrophys. J.*, vol. 542, pp. 281–307, 2000.

[8] A. Udalski, M. Kubiak, and M. Szymanski, "Optical gravitational lensing experiment. OGLE-II—the second phase of the OGLE project," *Acta Astronomica*, vol. 47, pp. 319–344, 1997.

[9] D. G. York *et al.*, "The Sloan digital sky survey: Technical summary," *Astronom. J.*, vol. 120, no. 3, pp. 1579–1579, 2000.

[10] N. Kaiser *et al.*, "Pan-starrs: A large synoptic survey telescope array," in *Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE) Conf. Ser.*, 2002, vol. 4836, pp. 154–164.

[11] Z. Ivezic *et al.*, LSST: From Science Drivers to Reference Design and Anticipated Data Products, Jun. 2011 [Online]. Available: http://www.lsst.org/lsst/overview/

[12] J. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York: Springer-Verlag, 2010.

[13] N. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science*, vol. 39, pp. 447–462, 1976.

[14] J. Scargle, "Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, vol. 263, pp. 835–853, 1982.

[15] A. Schwarzenberg-Czerny, "On the advantage of using analysis of variance for period search," *Month. Notices of the Roy. Astronom. Soc.*, vol. 241, pp. 153–165, 1989.

[16] M. Dworetsky, "A period finding method for sparse randonmly spaced observations," *Month. Notices of the Roy. Astronom. Soc.*, vol. 203, pp. 917–923, 1983.

[17] D. Clarke, "String/rope length methods using the Lafler-Kinman statistic," *Astron. Astrophys.*, vol. 386, pp. 763–774, 2002.

[18] W. Mayo, "Spectrum measurements with laser velocimeters," in *Proc. Dyn. Flow Conf., DISA Electronik A/S DK-2740*, Skoolunder, Denmark, 1978, pp. 851–868.

[19] R. A. Edelson and J. Krolik, "The discrete correlation function: A new method for analyzing unevenly sampled variability data," *Astrophys. J.*, vol. 333, pp. 646–659, 1988.

[20] P. Huijse, P. Estevez, P. Zegers, J. C. Principe, and P. Protopapas, "Period estimation in astronomical time series using slotted correntropy," *IEEE Signal Process. Lett.*, vol. 18, no. 6, pp. 371–374, 2011.

[21] I. Santamaria, P. Pokharel, and J. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2187–2197, 2006.

[22] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.

[23] J. Xu and J. Principe, "A pitch detector based on a generalized correlation function," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1420–1432, 2008.

[24] R. Li, W. Liu, and J. C. Principe, "A unifying criterion for instantaneous blind source separation based on correntropy," *IEEE Trans. Signal Process.*, vol. 87, no. 8, pp. 1872–1881, 2007.

[25] A. Gunduz and J. C. Principe, "Correntropy as a novel measure for nonlinearity tests," *IEEE Trans. Signal Process.*, vol. 89, no. 1, pp. 14–23, 2009.

[26] M. Michalak, "Time series prediction using periodic kernels," in *Computer Recognition Systems 4*. Berlin, Germany: Springer-Verlag, 2010, pp. 136–146.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[28] D. Mackay, *Introduction to Gaussian Processes*. Berlin, Germany: Springer, 1998, vol. 168, pp. 133–165.

[29] Y. Wang, R. Khardon, and P. Protopapas, "Nonparametric Bayesian estimation of periodic functions," *ArXiv e-Prints*, 2011 [Online]. Available: arxiv.org/abs/1111.1315

[30] M. Evans, N. Hastings, and B. Peacock, "Von Mises Distribution," in *Statistical Distributions*, 3rd ed. New York: Wiley, 2000, pp. 191–192.

[31] F. Abramovich, T. Bailey, and T. Sapatinas, "Wavelet analysis and its statistical applications," *The Statistician Part 1*, vol. 49, pp. 1–29, 2000.

[32] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. New York: Holden-Day, 1968.

[33] A. Schmitz and T. Schreiber, "Testing for nonlinearity in unevenly sampled time series," *Phys. Rev. E*, vol. 59, no. 4, pp. 4044–4047, 1999.

[34] T. Schreiber and A. Schmitz, "Surrogate time series," *Phys. D: Nonlin. Phenom.*, vol. 142, pp. 346–382, 1999.

[35] P. Buhlmann, "Bootstraps for time series," *Statist. Sci.*, vol. 17, no. 1, pp. 52–72, 1999.

[36] J. D. Hartman *et al.*, "Deep MMT transit survey of the open cluster M37 . II variable stars," *Astrophys. J.*, vol. 675, pp. 1254–1277, 2008.

[37] J. Devor, "Solutions for 10,000 eclipsing binaries in the bulge fields of OGLE II using DEBiL," *Astrophys. J.*, vol. 628, pp. 411–425, 2005 [Online]. Available: arXiv:astro-ph/0504399

Pablo Huijse (S'10) was born in Valdivia, Chile, in 1985. He received the B.S. and P.E. degrees in electrical engineering from the University of Chile in 2009.

He is currently pursuing the Ph.D. degree in electrical engineering at the University of Chile. He was a Visiting Student with Harvard University, Cambridge, MA, in 2012. His research interests are primarily on astronomical time series analysis using information theoretic concepts.

Pablo A. Estévez (M'98–SM'04) received the B.S. and the P.E. degrees in electrical engineering (EE) from the Universidad de Chile, Santiago, in 1978 and 1981, respectively, and the M.Eng. and Dr.Eng. degrees from the University of Tokyo, Japan, in 1992 and 1995, respectively.

He is currently the Vice-Chairman and an Associate Professor of the EE Department, Universidad de Chile. He was an Invited Researcher with the NTT Communication Science Laboratory, Kyoto, Japan; the Ecole Normale Supérieure, Lyon, France, and a Visiting Professor with the University of Tokyo.

Dr. Estévez is currently the Vice President for Members Activities of the IEEE Computational Intelligence Society (CIS). He has served as a Distinguished Lecturer and a Member-At-Large of the ADCOM of the IEEE CIS. He is an Associate Editor of the IEEE Transactions on Neural Networks and Learning Systems.

Pavlos Protopapas received the B.Sc. degree from Imperial College London. U.K., and the Ph.D. degree from the University of Pennsylvania, Philadelphia, in theoretical physics.

He is a Lecturer with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, and a researcher with the Harvard Smithsonian Center for Astrophysics. He is the PI of the Time Series Center a multidisciplinary effort in the analysis of time series. His research interests are primarily on time domain astronomy and large data analysis.

Pablo Zegers (SM'04) received the B.S. and P.E. degrees in engineering from the Pontificia Universidad Católica, Chile, in 1992, the M.Sc. degree from The University of Arizona, Tucson, in 1998, and the Ph.D. degree, also from The University of Arizona, in 2002.

He is currently an Associate Professor of the College of Engineering and Applied Sciences, Universidad de los Andes, Chile. His interests are artificial intelligence, machine learning, neural networks, and information theory. From 2006 to 2010, he was the Academic Director of this College, and for a brief period at the end of 2010, the Interim Dean.

Dr. Zegers is currently the Secretary of the Chilean IEEE section.

José C. Príncipe (M'83–SM'90–F'00) received the B.Sc. degree in electrical engineering from the University of Porto, Portugal, in 1972, and the M.Sc. and Ph.D. degrees from the University of Florida, Gainesville, in 1974 and 1979, respectively.

He is currently a Distinguished Professor of Electrical and Biomedical Engineering, University of Florida, Gainesville. He is BellSouth Professor and Founder and Director of the University of Florida Computational Neuro-Engineering Laboratory (CNEL). He is involved in biomedical signal processing, in particular, the electroencephalogram (EEG) and the modeling and applications of adaptive systems.

Dr. Príncipe is the past Editor-in-Chief of the IEEE Transactions on Biomedical Engineering, past President of the International Neural Network Society, former Secretary of the Technical Committee on Neural Networks of the IEEE Signal Processing Society, and a former member of the Scientific Board of the Food and Drug Administration. He was awarded the 2011 IEEE CIS Neural Networks Pioneer Award.