Mon. Not. R. Astron. Soc. **302,** 582–586 (1999)

Astronomical time-series analysis – II. A search for periodicity using the Shannon entropy

Pablo M. Cincotta,[1⋆] Amina Helmi,[2] Mariano Méndez,[1,3] Josué A. Núñez[1] and Héctor Vucetich[1]

[1] *Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, and Programa de Fotometría y Estructura Galáctica, Consejo Nacional de Investigaciones Científicas y Técnicas, Paseo del Bosque S/N, 1900 La Plata, Argentina*
[2] *Leiden Observatory, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[3] *Astronomical Institute 'Anton Pannekoek', University of Amsterdam, and Center for High-Energy Astrophysics, Kruislaan 403, NL-1098 SJ Amsterdam, the Netherlands*

Accepted 1998 September 21. Received 1998 August 10; in original form 1998 February 13

ABSTRACT

We have recently introduced a new method of searching a time series for periodic variability. The method uses the Shannon entropy to measure the amount of information provided by a set of observations that may contain an underlying periodic signal, as a function of the assumed period of this hypothetical periodic signal. Here we present the analytical arguments that support this algorithm within the broader frame of information theory. We also show that, in the absence of a periodic signal, the entropies follow a Gaussian distribution, which then provides an easy way of assessing the significance of a positive detection. We test this method using simulated data with non-sinusoidal variability, and we show that it is more sensitive than the classical periodograms or those variations adapted to deal with cases where harmonics are involved. Finally, we show that this method is capable of resolving two, almost identical, frequencies present in a given time series, even in cases where the classical periodograms fail to do so.

Key words: methods: analytical – methods: data analysis.

1 INTRODUCTION

The methods used to search for periodicity in astronomical time series may be divided into two groups: Fourier techniques and phase-diagram analyses. The first group includes the classical Fourier transform, and its variations introduced to deal with unevenly sampled data (Lomb 1976; Scargle 1982; Ferraz-Mello 1981). The second one is based on the analysis of the dispersion of the *light curve*, observed data folded over a trial period as a function of phase, for a set of trial periods (Jurkevic 1971; Stellingwerf 1978; Marraco & Muzzio 1980). Although intuitively simple, the methods based on the analysis of the phase diagram have not been studied analytically, and rely on numerical results.

In general, each method has certain advantages and disadvantages in its application, and one has to decide which one is more appropriate to the type of data available. For instance, while Fourier analysis is faster and extremely efficient, it is much less sensitive to non-sinusoidal functions than the phase-diagram analysis. Also, phase-diagram methods are less affected by randomly occurring gaps in the data, provided that the coverage of the light curve is reasonably uniform in phase.

In a previous paper, Cincotta, Méndez & Núñez (1995, hereafter CMN95) introduced a new algorithm based on the analysis of the entropy of the light curve. It relies on the concept that, from the set of all possible light curves, that which is folded using the *correct period* is the most *ordered* and therefore the one that contains more *information* about the signal. As the entropy measures the lack of information about a system (the light curve in this case), the entropy will be minimum when the trial period is equal to the actual period of the time series.

In a broad sense, this algorithm belongs to the second of the above-mentioned groups. However, unlike all the other algorithms based on the analysis of the phase diagram, the entropy method rests on an analytical theory.

In this paper we show how our technique is contained within the broader frame of information theory and we use the methods of this discipline to prove its validity. We also work out here the statistical properties of the entropy of a signal consisting only of white noise, and we use these properties to derive a procedure to test any time series against the null hypothesis that it contains no periodicity. We include an example of a simulated signal to study its sensitivity in

⋆Present address (until August 1999): Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain. E-mail: pablo@zeus.maia.ub.es, pmc@fcaglp.fcaglp.unlp.edu.ar (PMC)

comparison to other techniques. Finally, we also study the resolving power of the entropy for a time series consisting of two periodic functions with almost identical frequencies, and compare these results to those obtained using Fourier analysis (for further comparisons with other methods see CMN95).

2 ANALYTICAL ASPECTS

2.1 Definitions

Let us consider a real function u, continuous, bounded and T-periodic, defined on the interval $[t_0, t_f]$, where we assume that $|t_f - t_0|/T \gg 1$. Because u is bounded we can normalize it to the unit interval. From now on we will call u the normalized function, and A the two-dimensional space where it is contained ($A = [t_0, t_f] \times [0, 1]$). As u is T-periodic, we can write $u[(t - t_0)/T]$, and consider u defined in the unit interval: $u(\phi)$, where $\phi = (t - t_0)/T$ (mod 1), which allows us to consider the time as an angular variable. The points $(\phi, \eta = u(\phi))$ are coordinates on a cylinder C, which is equivalent to the unit square with opposite sides $\phi = 0, 1$ identified. When ϕ changes by one unit, the curve repeats itself. This happens m times, $m = 0, 1, ..., N_T$ where $N_T = [(t_f - t_0)/T]$ and $[\ldots]$ is the integer function.

Let (t, z) be a point of A. We can construct a multivalued map from A to C in the following way:

$$\boldsymbol{F}_p(t, z) = (\Phi_p(t), z), \tag{1}$$

where

$$\Phi_p(t) = \frac{t - t_0}{p} \mod 1. \tag{2}$$

Here, p is a positive real number such that $|t_f - t_0|/p \gg 1$. As we can readily see from equations (1) and (2), the first component of $\boldsymbol{F}_p$, $F_p^{(1)}$, is some phase φ, and the second one, $F_p^{(2)}$, is the identity, $\eta = z$.

In the case of a curve in A, like the time series u, we have $z = u(t)$ and we look for $\boldsymbol{F}_p(t, u(t))$, the image of the curve in C. The first component of the map is given by equation (2), $F_p^{(1)}(t, u(t)) = \varphi$, while the second component is

$$F_p^{(2)}(t, u(t)) = \eta_p = g_p(\varphi) = u\big((\varphi + n)p\big), \tag{3}$$

where $n = 0, 1, ..., N_p = [(t_f - t_0)/p]$ and $N_p \approx [N_T(T/p)]$. The number n increases by one unit each time the trial phase, $\varphi = \Phi_p(t)$, increases by 1, while the number m increases by one unit each time the true phase, $\phi = \Phi_T(t)$, increases by 1. This map has a very simple interpretation. For instance, if $p < T$ (similar considerations apply when $p > T$), we see from equation (2) that the true phase and the trial phase are related by $\varphi = (T/p)\phi$. Then two points separated by $\Delta\phi \leq 1$ are separated in the φ-axis by $(T/p)\Delta\phi$. Therefore, the map stretches the interval $\Delta\phi = 1$ by a factor $T/p > 1$ without changing the ordinates. However, since the phase is taken mod 1, the points in the (φ, η) plane do not follow a single curve like in the (ϕ, η) plane.

If we consider the time series set $\mathcal{U} = \{(t, z)\colon z = u(t)\}$, we can say that $\boldsymbol{F}_p$ maps $\mathcal{U}$ into the set $\mathcal{G}_p = \{(\varphi, \eta_p), \eta_p = g_p(\varphi)\}$; i.e. $\boldsymbol{F}_p(\mathcal{U}) = \mathcal{G}_p$. The latter is usually called the (normalized) light curve for the trial period p.

2.2 Entropy

The function u has been assumed to be T-periodic, so one expects all sets $\mathcal{G}_p$ for which $p \approx T$ to be more ordered than the rest. This degree of order can be measured by the Shannon entropy (for a theoretical background see e.g. Shannon & Weaver 1949; Katz 1967; Wehrl 1978).

Let us take the function Z, defined in the unit interval, given by

$$Z(x) = -x \ln x, \qquad 0 < x \leq 1, \qquad Z(0) = 0. \tag{4}$$

Let $\alpha = \{a_i, i = 1, ..., q\}$ be a partition on C, that is, q disjoint two-dimensional boxes covering the whole unit square. If we are considering a discrete series then $\mathcal{G}_p = \{(\varphi_i, \eta_{p_i}),\ \eta_{p_i} = g_p(\varphi_i),\ i = 1, ..., N\}$. Therefore one can define on C a probability density $\rho_p(\boldsymbol{x})$ as

$$\rho_p(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}_{p_i}), \tag{5}$$

where $\boldsymbol{x}_{p_i} = (\varphi_i, \eta_{p_i})$ is a point of $\mathcal{G}_p$, and δ is the delta function, from which it follows that $\int_C \rho_p(\boldsymbol{x})\, \mathrm{d}^2x = 1$, where $\mathrm{d}^2x = \mathrm{d}\varphi\, \mathrm{d}\eta$. The probability for each box a_i, $\mu_p(a_i)$, is then

$$\mu_p(a_i) = \int_{a_i} \rho_p(\boldsymbol{x})\, \mathrm{d}^2x, \tag{6}$$

and the entropy, for the partition α and trial period p, is defined as

$$S_p(\alpha) = \sum_{i=1}^{q} Z[\mu_p(a_i)] = -\sum_{i=1}^{q} \mu_p(a_i) \ln \mu_p(a_i) \geq 0. \tag{7}$$

For a given partition, the entropy is a bounded quantity: $0 \leq S_p \leq \ln q$, where q is, as before, the number of elements of the partition. The entropy is zero when $\mu_p(a_i) = 1$ for some i and $\mu_p(a_j) = 0$, for all $j \neq i$ (all data points fall in the same element of the partition), and it is maximum when $\mu_p(a_i) = \mu = \text{constant}$, for all i. From the normalization condition: $\sum_i \mu_p(a_i) = 1$, and therefore $\mu = 1/q$. In this case all the data points are uniformly distributed over the unit square. In what follows S_p denotes the normalized entropy (i.e., divided by $\ln q$) for the partition α.

In CMN95 we found numerical evidence that the normalized entropy is constant ($\sim$1) for almost all values of p, is minimum for $p \approx T$ and has local minima for some rational multiples of T. In order to justify these numerical results, let us consider the map $\boldsymbol{F}_p$ introduced in Section 2.1. As we showed, the map shifts points with the same true phase by an amount T/p while the ordinate remains fixed. Then, for p/T rational the light curve closes itself after some n, while for p/T irrational the curve does not close, and for large n ($N_p, N_T \gg 1$), it covers the whole unit square. Therefore, when p/T is irrational the map will be dense on C.

In every case, the values of φ are distributed uniformly over the unit interval, while the η values are distributed according to the function u. Indeed for each n, the probability that η lies within the interval $(\eta, \eta + \mathrm{d}\eta)$ is $\mathrm{d}P(\eta) = \mathrm{d}\eta/g_p'(g_p^{-1}(\eta))$, where $g_p' \equiv \mathrm{d}g_p/\mathrm{d}\varphi$ and g_p^{-1} is the inverse of g_p (defined in a domain where it exists). If u is a smooth function, since it is continuous, bounded and periodic, there are values of η for which $g_p' = 0$. In these points the probability density diverges, but the integrated probability is finite. But, since $\delta\eta = g_p'(\varphi)\delta\varphi$, in those phase intervals where $g_p' \approx 0$, $\eta \approx$ constant. Then, in the neighbourhood of a critical point, the probability for η is larger than in the rest of the unit interval. However one can prove that in the case of a dense light curve, that is for p an irrational multiple of T, even though the data points do not uniformly fill the unit square, one can always choose a partition (in η) such that the entropy is arbitrarily close to its maximum value. Thus, for any time series of period T, and for irrational p/T, the entropy is $S_p \approx 1$.

On the other hand, for p/T rational, i.e. $p/T = r/s$, r and s co-prime numbers, we showed above that $\boldsymbol{F}_p$ is not dense on C: the light curve closes itself when $n = s$. Nevertheless, for s large the map is similar to that for the irrational case. Indeed, for large s the light curve fills almost the whole unit square, and can be

considered dense for all practical purposes. For a given partition, and for s small, there are empty and non-empty boxes. The probability of occupation of the non-empty boxes is determined by the shape of the function u (see Section 2.3). However, because the distribution is not dense on C, S_p is always less than its maximum value ($S_{\max} = 1$), since the empty boxes do not contribute to the sum (7). This qualitative result shows that the entropy takes smaller values for certain rational multiples of p/T (in fact, those for which $s \ll N_p$). In the next section we show this in a quantitative manner.

2.3 Minima of the entropy

In this section we shall estimate the amplitude of the minima. Let us consider first the case of $p = T$.

Let $\eta = g(\phi)$ be a smooth light curve, that is for all ϕ its derivative is bounded: $|g'| < H$, where H is some positive real number. Let α be the partition

$$\{a_{lk} = \Delta\phi_l \times \Delta\eta_k \colon \phi_l = l/L, \eta_k = k/K\},$$

with $l = 0, ..., L, k = 0, ..., K$ and $L, K \geq 2$. Let us consider an element of the partition, a_{lk}, that is crossed by the curve. For this element it holds that $\phi_{l-1} \leq \phi < \phi_l$ and $\eta_{k-1} \leq g(\phi) < \eta_k$. Let L be large enough so that in each $\Delta\phi_l$ the inverse of g exists, except in those domains where $g' = 0$. To calculate the entropy, we must first calculate the probability $\mu(a_{lk})$, for that element. From equations (5) and (6) it is not difficult to show that $\mu(a_{lk}) = |\Delta\phi_{lk}|$, where $|\Delta\phi_{lk}| = |\phi_{lk} - \phi_{l(k-1)}|$ is the distance between the phases at η_k and η_{k-1}: $\phi_{lk} = g^{-1}(\eta_k)$ for g locally defined in a domain that includes $[\phi_{l-1}, \phi_l)$. There are two possibilities: (i) If the curve does not cover all the phase interval in a_{lk}, i.e. if $|\Delta\phi_{lk}| < |\Delta\phi_l|$, we may write $|\Delta\phi_{lk}| \approx |\Delta\eta_k|/|g'_l|$ where $g'_l \neq 0$ is the mean derivative of g within $\Delta\phi_l$. This means that $|g'_l| > L/K$. (ii) If the curve occupies the whole phase interval of the element a_{lk}, then $|\Delta\phi_{lk}| \geq |\Delta\phi_l|$, $|g'_l| \leq L/K$, and $\mu(a_{lk}) \approx |\Delta\phi_l|$. Clearly, in any case, if $g'_l = 0$, $\mu(a_{lk}) = |\Delta\phi_l|$. Since by assumption $|g'| < H$ we can choose the partition such that $L/K \sim H$, and then $\mu(a_{lk}) \approx |\Delta\phi_l| = 1/L$, i.e. constant for every element occupied by the curve. Then the entropy reduces to $S \approx \ln L/\ln(KL)$, with the derivative constraint $L/K \sim H$.

In general, $K \geq 2$ so $L \geq 2H$, and in order to satisfy the derivative constraint ($L/K \sim H$), one may be forced to choose a relatively large value of L. But for real time series, noisy and with a finite number of data points, large values of L (and/or K) produce a large amount of statistical noise; then relatively low values of L and K are desirable. [Indeed, counting fluctuations are the dominant effective noise. For example, when the light curve is dense, the fluctuations are of the order of $1/\sqrt{\lambda}$, where $\lambda = N/(KL)$ (see Section 3, equation 9). So, to avoid large fluctuations in the maximum value of the entropy, which may hide the periodicity, we require that $\lambda \gg 1$]. Therefore, in practical applications the usual situation is that L and K satisfy the derivative constraint only for some of the elements of the partition, and the previous result, $\mu(a_{lk}) = 1/L$, represents an upper bound for the actual value of the probability. In order to give a quantitative estimate of the amplitude of the minimum for real (smooth) time series and L, K both not too large, we may assume that all the non-empty elements of the partition have nearly the same probability, but $\mu(a_{lk}) \leq 1/L$. This is based on the previous result, and on the fact that in any real time series the observed values of η have errors that lead to a broad distribution of data within each box occupied by the light curve. Hence, we may approximate the probabilities by $\mu(a_i) \approx 1/q_0,\ i = 1, ..., q_0$, where q_0 ($L \leq q_0 < q$) is the number of non-empty elements of the partition, and the entropy is simply $S \approx \ln q_0/\ln q$ (similar results can be obtained in the case of a narrow pulse, i.e. non-smooth light curve). The exact value of the entropy at the minimum strongly depends on the shape of the light curve. However, in general $q_0 \geq L$, and so $S_{\min} \geq \ln L/\ln(LK)$. For example, in the case of a square wave $S_{\min} \approx \ln(L + 2K - 4)/\ln(LK)$, while for a narrow pulse $S_{\min} \approx \ln(L + K - 1)/\ln(LK)$.

From the previous analysis it also follows that when the data are evenly spaced in time and when p takes the value of the sampling period, the corresponding light curve is ordered. In this case the entropy has a minimum with an expected amplitude $\ln(2K)/\ln(LK)$.

Now let us consider the case where p is a rational multiple of T, i.e. $p/T = r/s$. As we have already mentioned, the curve repeats itself when $n = s$. If we consider that the former assumption of equal probabilities is still valid in this case, then the minimum number of non-empty boxes, q_0, occurs for $s = 1$, i.e. for p an integer multiple of the true period. The same is true for r: q_0 is minimum for $r = 1$. In general, if $r, s > 1$, q_0 increases by a factor $\sim r, s$ compared to the case where $r, s = 1$. Then, for large r, s: $q_0 \approx q$ and $S \approx 1$. Hence we conclude that the absolute minimum of S_p occurs for $p = T$.

3 STATISTICAL ANALYSIS

For the evaluation of the detection significance one needs the probability distribution of the entropy for a non-periodic signal. From the results of the previous section, this is equivalent to calculating the probability distribution of the entropies of a periodic signal at irrational multiples of the period (as long as the noise in the data is uncorrelated).

Let n_i be the number of data points falling within the element $a_i, i = 1, ..., q$, then S may be written as

$$S = \frac{N \ln N + S}{N \ln q}, \qquad S = \sum_{i=1}^{q} s_i \tag{8}$$

where, as usual, N is the total number of data points and $s_i = -n_i \ln n_i$. If N is large, then n_i will follow a Poisson distribution ($n_i \equiv n$):

$$P_\lambda(n) = \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda}, \tag{9}$$

where $\lambda = N/q$ is the mean number of data points per element of the partition. For $\lambda \gg 1$, the distribution will be strongly peaked at $n \approx \lambda$, and we can restrict our analysis to $n = \lambda(1 + \xi)$, with $\xi \ll 1$ (in what follows we assume n to be a real variable). Then, up to first order in ξ, s ($\equiv s_i$) reduces to

$$s \approx -\lambda \ln \lambda - \lambda(1 + \ln \lambda)\xi. \tag{10}$$

In order to get information about the distribution of s, we first compute the distribution for ξ. For simplicity, we write equation (9) as

$$P_\lambda(n) = \mathrm{e}^{\psi_\lambda(n)} \tag{11}$$

where $\psi_\lambda(n) = -\lambda + n \ln \lambda - \ln n!$. Using Stirling's formula [$\ln n! \approx \ln \sqrt{2\pi} + \left(\frac{1}{2} + n\right) \ln n - n$], replacing $n = \lambda(1 + \xi)$ and keeping terms up to $1/\lambda$ and up to second order in ξ, we obtain

$$\psi_\lambda(n) \approx -\ln \sqrt{2\pi\lambda} - \frac{\lambda}{2}\xi^2. \tag{12}$$

Recalling that $P_\lambda(n)\,\mathrm{d}n = \bar{f}_\lambda(\xi)\,\mathrm{d}\xi = f_\lambda(s)\,\mathrm{d}s$, where $s(\xi)$ is given by equation (10), we obtain the distribution of s, which is given by

$$f_\lambda(s) \approx \frac{1}{\sqrt{2\pi\lambda}\ln(\mathrm{e}\lambda)} \exp\left[-(s + \lambda \ln \lambda)^2/2\lambda \ln^2(\mathrm{e}\lambda)\right]. \tag{13}$$

Then, using the fact that $\langle S\rangle = q\langle s\rangle$, $\mathrm{Var}(S) = q\mathrm{Var}(s)$, from equations (8) and (13) we conclude that the entropy values follow a Gaussian distribution with the following parameters:

$$\langle S\rangle \approx 1, \qquad \sigma^2 \approx \frac{1}{N}\left(\frac{\ln N + 1}{\ln q} - 1\right)^2, \tag{14}$$

and for large N, the dispersion of the entropy goes as $\sim \ln N/\sqrt{N}$.

However, it is possible to check this hypothesis using the actual data, as long as, in the absence of a periodic signal, the noise present in the data is white [for example, this does not apply in the case of the red noise observed in several X-ray sources (Hasinger & van der Klis 1989)]. In practice, a non-periodic time series can be constructed using the same data by, for instance, randomizing the observed quantities. Once randomized, this new time series can be analysed in the same way as the original one, $\langle S\rangle$ and σ can be determined empirically, and a χ^2 test can be performed to test the validity of the Gaussian distribution. We show this in the next section.

4 EXAMPLES

4.1 Statistical significance

The classical methods of searching for periodicity (phase-diagram and power-spectrum methods) are not well sustained from the statistical point of view, and Schwarzenberg-Czerny (1989, 1996, hereafter SC96) showed that the detection significance is reduced in those cases. Instead, he proposed a different tool using periodic orthogonal polynomials to fit the data and the analysis of the variance statistic to evaluate the fit. The latter technique is a generalization of the Lomb–Scargle periodogram (Lomb 1976; Scargle 1982). SC96 showed that his method, when applied to a non-sinusoidal signal, is more sensitive than the Lomb–Scargle power spectrum (see SC96, fig. 1). Therefore, a good test of the entropy method is to compare its sensitivity for a non-sinusoidal signal against the sensitivity of the method proposed by SC96. For that purpose we studied the same synthetic signal (see SC96 for more details): 1000 times observations drawn from the standard normal distribution and a signal consisting of a narrow pulse of height S/N and 60 cycle d^{-1} plus unit-variance Gaussian white noise, for different values of S/N ranging from 16 to 1/8. We analysed each time series using a partition of 5×2 elements of the unit square. As we already explained in the previous section, to calculate the significance of a peak in the resulting periodogram we first randomized and then reanalysed each time series. We binned the entropies obtained from each randomized time series into a histogram and fitted it to a Gaussian distribution. In all cases the reduced χ^2 of these fits was $\lesssim 1$, indicating that the distribution of the entropies of the randomized time series does not deviate significantly from a Gaussian. Now we use the results of these fits to calculate the significance of the known period in σ units, corrected for the estimated number of independent frequencies (Horne & Baliunas 1986), which we show in Fig. 1 as a function of S/N. We include in the same figure the confidence level at the input frequency obtained from the Lomb–Scargle and the analysis-of-variance periodograms (SC96). This figure shows that for $S/N \geq 1$ the entropy is more sensitive, and that the significance of detection increases more rapidly with S/N compared to the other two methods, while for $S/N < 1$ all three methods are equally unable to detect the underlying periodic signal.

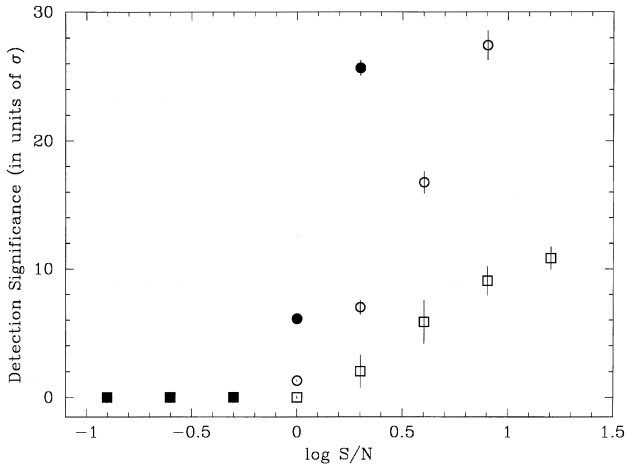

Figure 1. Significance of the detection of the input period of a simulated signal (see text) as a function of the signal-to-noise ratio S/N. Values are given in units of σ and are corrected for the estimated number of independent frequencies. Filled circles, open circles and squares represent the results obtained using our method, the analysis-of-variance statistic (Schwarzenberg-Czerny 1996) and the Lomb–Scargle periodogram, respectively. We include the error bars from Monte Carlo simulations. The significances of detection using our method for S/N = 4, 8 and 16 (not shown in this figure) were 82.3 ± 1.1 176.8 ± 1.1 and 225.7 ± 1.8, respectively, larger than those obtained using the other two methods.

4.2 Resolving power

In this section we compare the resolving power of this method to that of the Fourier periodogram. For this purpose we simulated a set of time series consisting of the sum of two periodic functions:

$$u_j(t) = \sum_{n=1}^{3} A_n \sin\left(\frac{2n\pi t}{T_j} + \phi_{nj}\right), \qquad j = 1, 2, \tag{15}$$

plus Gaussian white noise. In all our simulations the two functions had the same amplitudes, with the second and third harmonics contributing 30 and 10 per cent of the fundamental, respectively. The six phases were chosen at random: 0.773, 4.335 and 4.404 rad for the first function, and 1.47, 5.73 and 0.214 rad for the second. We fixed the period of the first signal to 3.4 s, while the other one was set to different values on different trials such that the frequency difference among both functions changed from 0.000 01 to 0.02 Hz.

We use the above function to simulate an observation of 150 s, 10 000 bins of 0.015 s, which can be compared to a typical satellite observation of a variable X-ray source. We produced three sets of simulations with signal-to-noise ratios of 4, 8 and 16. We analysed these simulated data using the entropy method for a 3×3 partition, and the classical Fourier periodogram. We searched the same period range and number of periods in both cases.

Loumos & Deeming (1978) and Kovacs (1981) showed that Fourier techniques are able to separate two frequencies only if $\Delta\nu > 1/\tau$ ($\nu = 1/T$ and $\tau = t_\mathrm{f} - t_0$), which for our simulations means $\Delta\nu \gtrsim 0.006$ Hz. From Section 2 it is not difficult to show that for the entropy method $\Delta\nu \gtrsim 1/(\tau L)$, which in this example means $\Delta\nu \gtrsim 0.002$ Hz.

In Fig. 2 we plot the $\Delta\nu$ measured from both periodograms as a function of the $\Delta\nu$ used for the simulations with the Monte Carlo error bars, and for a signal-to-noise ratio of 16. The arrow indicates the value of $1/\tau$. For comparison, we also included in this figure a line with slope = 1.

Under similar conditions, the Fourier periodogram is able to resolve both frequencies only when their separation is greater than

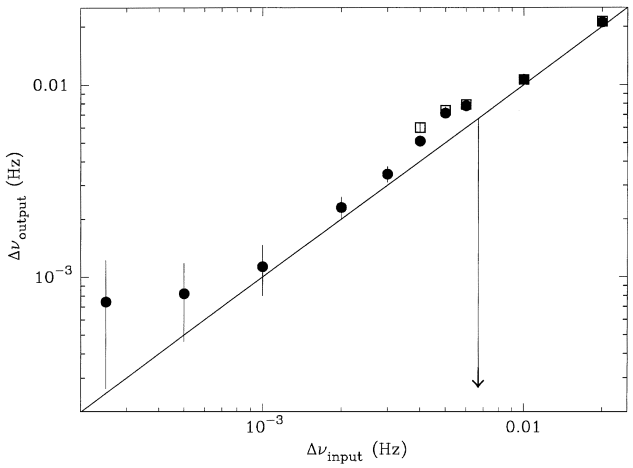

Figure 2. Measured frequency difference of a signal consisting of two periodic functions (see text) as a function of the difference used for the simulations with the Monte Carlo error bars. Circles and squares represent the results obtained using the method described here and the classical Fourier periodogram, respectively. The arrow indicates the maximum resolution expected for the Fourier periodogram given the length of the time series. The line is the $\Delta\nu_{\rm output} = \Delta\nu_{\rm input}$ relation. Both methods give the same results for $\Delta\nu_{\rm input} \geq 0.006$ Hz, while the Fourier periodogram cannot resolve the two frequencies when $\Delta\nu_{\rm input} \lesssim 0.004$ Hz.

$\sim$0.004 Hz, in accordance with the results of Loumos & Deeming (1978) and Kovacs (1981), while the entropy method is still able to resolve both frequencies when the separation is smaller.

We got similar results for the experiments with a signal-to-noise ratio of 4 and 8. In those two cases the entropy method still performed better than the Fourier periodogram, although the minimum separation that we were able to resolve increased as the signal-to-noise ratio decreased.

We have to add, however, that the frequencies measured with this method, as well as their difference, were not exactly the same as those in the original signal, and in some cases the difference was as big as 30 per cent of the expected values. As Loumos & Deeming (1978) showed, this is also true when Fourier analysis is used. However, even if the measured frequencies are not equal to the real ones, our simulations show that the method presented here can indicate their existence while Fourier analysis cannot.

5 CONCLUSIONS

We have presented the framework in which our method of searching for periodicities is contained. We proved its validity from the analytical point of view and we have confirmed the numerical results found in CMN95. Moreover, we gave an easy rule to compute the amplitude of the different minima in the periodogram and some hints about the way in which the partition has to be chosen to optimize the method. However, it should be clear from the discussion in Section 2 that the concept of *ordered* light curve depends strongly on the partition. It is therefore very important to keep in mind that a good performance of this technique depends on an appropriate selection of the partition.

For a time series of a given signal-to-noise ratio, this method is more sensitive to the presence of a period than the more often used Fourier power spectrum. Unlike the discrete Fourier transform, it does not assume any shape for the underlying periodic function, therefore making it more suitable for the study of non-sinusoidal signals.

Furthermore, the entropy method has a very good frequency resolution, and is able to separate two, almost identical, frequencies even in cases where the Fourier periodogram cannot.

It has one disadvantage, already mentioned in CMN95 and in Section 2.3, when dealing with evenly sampled data: the method is very sensitive to the pseudo-aliasing introduced by the sampling interval. In the classical astronomical case, the 1-d pseudo-alias introduced by the way the data are generally taken might in some cases be the dominant peak in the periodogram. Nevertheless, there are ways to reduce the amplitude of the entropy at these frequencies by an appropriate selection of the partition of the unit square. Furthermore, as it is possible to compute its amplitude, this alias can be easily identified.

ACKNOWLEDGMENTS

We are very grateful to Professor Carles Simó for his useful suggestions and criticism. Also we thank Daniel Carpintero for his valuable comments. We are specially grateful to two anonymous referees for carefully reading the manuscript and for their many suggestions that helped us to improve it. PMC is grateful to the Research Foundation of the State of São Paulo and the hospitality of the Instituto Astronômico e Geofísico, Universidade de São Paulo, where part of this work was done. AH wishes to thank Zonta International for the Amelia Earhart fellowship award. PMC and MM are fellows of the Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina.

REFERENCES

Cincotta P. M., Méndez M., Núñez J. A., 1995, ApJ, 449, 231 (CMN95)
Ferraz-Mello S., 1981, AJ, 86, 619
Hasinger G., van der Klis M., 1989, A&A, 225, 79
Horne J. H., Baliunas S., 1986, ApJ, 302, 757
Jurkevic I., 1971, Ap&SS, 13, 154
Katz A., 1967, Principles of Statistical Mechanics, The Information Theory Approach. W. H. Freeman, San Francisco
Kovacs G., 1981, Ap&SS, 69, 485
Lomb N. R., 1976, Ap&SS, 39, 447
Loumos G. L., Deeming T. J., 1978, Ap&SS, 56, 285
Marraco H. G., Muzzio J. C., 1980, PASP, 92, 700
Scargle J. D., 1982, ApJ, 263, 835
Schwarzenberg-Czerny A., 1989, MNRAS, 241, 153
Schwarzenberg-Czerny A., 1996, ApJ, 460, L107 (SC96)
Shannon C., Weaver W., 1949, The Mathematical Theory of Communication. Illinois Univ. Press, Urbana
Stellingwerf R. F., 1978, ApJ, 224, 953
Wehrl A., 1978, Rev. Mod. Phys., 50 (2), 221

This paper has been typeset from a TEX/LATEX file prepared by the author.