THE ASTROPHYSICAL JOURNAL, 498:666–670, 1998 May 10

A SIMPLE METHOD FOR DETECTING PERIODIC SIGNALS IN SPARSE ASTRONOMICAL EVENT DATA

ANDRES CICUTTIN, ALBERTO A. COLAVITA,[1] ALBERTO CERDEIRA,[2] RADU MUTIHAC,[3] AND SILVIO TURRINI
International Centre for Theoretical Physics–Istituto Nazionale Di Fisica Nucleare, Microprocessor Laboratory, Via Beirut 31, 34100 Trieste, Italy;
cicuttin@ictp.trieste.it, colavita@ictp.trieste.it, cerdeira@ictp.trieste.it, mutihac@ictp.trieste.it, turrini@ictp.trieste.it
Received 1994 April 17; accepted 1997 December 8

ABSTRACT

We present a simple method to detect periodic signals in sparse astronomical event data. The method is particularly appropriate for gamma-ray astronomy where the number of available photons is sparse in time and Poissonian noise dominates the statistics. It is based on an autocorrelation function, which provides phase independence. We have implemented and successfully applied this method on simulated data. This paper presents some numerical results and a description of the model used to generate the synthetic data along with a formal definition of the signal-to-noise ratio in the generated time series.
Subject headings: methods: data analysis — methods: statistical

1. INTRODUCTION

The gamma-ray spectrum was the last photon energy range to be opened to astronomical observation. In this regard EGRET, part of *Compton Gamma Ray Observatory* (*CGRO*), has made great contributions to the field (Kanbach et al. 1988; Fichtel et al. 1994).

In gamma-ray astronomy, the detected photons emitted by various sources are relatively few and sparse in time. The small number of detected gamma photons is due both to the low emission of the sources and to the limited gathering capacity of present-day gamma-ray telescopes. This implies a sparse time series of photon arrival times from a given source in the sky and, consequently, poor statistics for data analysis. Under these conditions, the gathered data are mainly dominated by Poissonian noise, which makes it difficult to observe periodic behaviors of the emitting source. The detection of periodic signals with sparse data is a difficult problem when there is no indication of periodicity from a different band of the spectrum (Bignami, Caraveo, & Mereghetti 1993).

The new generation of gamma-ray telescopes, like GLAST, ARGO, and others based on large-area detectors (Colavita 1996; Sacco et al. 1993), will provide a large amount of data with better statistics and at the same time with better temporal and angular resolution. Hence, new gamma-ray sources that have no counterparts in other regions of the spectrum could be identified.

A suitable method to explore periodicity in sparse data should be able to deal with a small number of events and should provide some rejection to the constant background emission, which always contaminates data from a point source. Different methods have been developed to cope with the problem of periodicity searching (Cincotta, Mendez, & Nuñes 1995; Bai 1992; Gregory & Loredo 1992; Swanepoel & De Beer 1990; Scargle 1989; Leahy et al. 1983).

The method we propose allows the detection of periodicity or, at least, the formulation of a hypothesis as to its existence even in the case of a very small number of detected photons. The noncorrelated noise is adequately suppressed by means of an autocorrelation of the data, while the influence of background emission is limited by considering only the variations around a mean value. It is also possible to determine the period of the main component by inspecting the peaks of an estimation function, which we have defined.

2. THE METHOD

The method of finding periodicity within sparse data can be structured into a sequence of four separate steps:

1. Folding of the data into a trial period (§ 2.1),
2. Subtraction of the mean value from folded data (§ 2.2),
3. Autocorrelation of the processed data (§ 2.3), and
4. Cosine Fourier transformation of the autocorrelation function (§ 2.4).

2.1. *Folding*

Let us assume that $\{T_i\}$ are the photon arrival times. By means of the transformation

$$t_i = T_i - \left[\frac{T_i}{\tau}\right] \cdot \tau \ , \quad (1)$$

where [·] stands for the integer part function, we perform a folding of the time series within a trial period τ. The $\{t_i\}$ obtained in this way belong to the interval $[0, \tau]$. We next build a histogram by dividing this period into k bins and counting the number of t_i that fall within each bin. From now on, k will be considered an implicit variable of the functions we will define.

After folding the whole sequence into segments of a chosen period length, the last segment is generally incomplete since the period is not an exact divisor of the sequence length and will therefore form a *tail* of events. If we correlated all the segments, the *tail* would be completed with nonexistent zero events, and this would indicate false periodicity. This sort of *tail effect* is particularly important when the number of folded periods is small and the *tail*

[1] Universidad Nacional de San Luis, Ejercito de los Andes 950, 5700 San Luis, Argentina.

[2] Center for Applied Studies in Nuclear Development, 5[ta] y 30 Miramar, C. Havana, Cuba.

[3] University of Bucharest, Faculty of Physics, P.O. Box MG-11, Magurele, (76900) Bucharest, Romania.

length is approximately half of the folding period. We avoided this effect by ignoring the *tail.*

2.2. *Subtraction of the Mean Value*

Let q_i be the number of t_i that fall into the ith interval of τ; then the mean value of the histogram is given by

$$\bar{q} = \frac{1}{k} \sum_{j=1}^{k} q_j \, . \tag{2}$$

By subtracting the mean from each of the bins, we obtain a new set of data $\{p_i\}$ defined as follows:

$$p_i = q_i - \bar{q} \, , \quad i = 1, 2, \ldots, k \, . \tag{3}$$

In this way, the new histogram given by $\{p_i\}$ has its mean value equal to zero. The effect of this subtraction is to cancel the contribution of the nonperiodic background emission. The values $\{p_i\}$ will be smaller if the histogram of the $\{q_i\}$ is relatively flat. Therefore, $\{p_i\}$ should reflect in some way the fluctuations around the mean value. These fluctuations are meaningful in the case of periodicity.

2.3. *Autocorrelation*

By means of the autocorrelation function, we expect to suppress the noncorrelated noise that is mainly due to the discrete nature of data. By using the autocorrelation function we ignore the phase of the periodic component that we are trying to detect in the time series (Bai 1992).

We can define the autocorrelation function F as follows:

$$F(j) = \frac{1}{N} \sum_{i=l}^{k-j} p_i \cdot p_{i+j} \, , \quad j = 0, 1, \ldots, k - 1 \, , \tag{4}$$

where N is the number of arrival times used.

The folding procedure a priori assumes a virtual periodicity; hence, we may set periodic boundary conditions on the previous expression. With this assumption we define a new autocorrelation function in order to use all possible pairs of values (p_i, p_j) as follows:

$$F(j) = \begin{cases} \frac{1}{N} \sum_{i=l}^{k} p_i \cdot p_{i+j} \, , & \text{if } i + j \le k \\ \frac{1}{N} \sum_{i=l}^{k} p_i \cdot p_{(i+j)-k} \, , & \text{if } i + j > k \end{cases} \quad \text{for } j \le \frac{k}{2} \, . \tag{5}$$

2.4. *Cosine Fourier Transformation of the Autocorrelation Function*

The last step deals with the evaluation of the area under the autocorrelation function graph. To point out its meaning, let us imagine that the histogram of $\{p_i\}$ consists of uncorrelated noise only; then its autocorrelation function will look like a narrow peak. The autocorrelation function has its maximum when its argument is zero, whereas for nonzero arguments it falls down quickly and keeps on fluctuating around a base value, encompassing a small area. The integral of this function can be considered a *measure* of how much noise is present in the histogram given by $\{p_i\}$. To evaluate the area under the curve representing the autocorrelation function F, we propose a discrete cosine Fourier transformation of F. This corresponds to a weighted integral where the cosine is the weight function. This procedure should bring to mind the Wiener-Kinchine theorem, which is used to define the power spectral density of a stationary random process to be the Fourier transform of its autocorrelation function. Finally we define the period estimation function (PEF) as follows:

$$\text{PEF}(\tau) = \sum_{j=0}^{k/2} F(j) \cos\left(\frac{2\pi}{k} j\right) . \tag{6}$$

This function depends explicitly on the period used to fold the time series, and we expect it to reach its maximum if the folding period coincides with the actual period of the time series. This maximum should be distinguishable from the spurious peaks due to the noise fluctuation. We have defined the signal-to-noise ratio of the PEF as the ratio of its peak value referenced to its mean value and its standard deviation. We have chosen this ratio to have a measure of how distinguishable the peak we are looking for is from others peaks, i.e., the peak that would correspond to the period of the sequence.

To detect the peak we have to scan from a minimum to a maximum τ (the period interval of interest) using a period step that must be less than half of the peak width in order to avoid skipping the right peak. When we are folding the time series with a period close to the actual period present in the time series, we have to be sure that the last folded segment does not contribute negatively to the final folded data. That is, the accumulated phase difference must not be larger than half a period. To clarify this point let us consider the difference between the folding period and the actual one to be close to zero, and let us designate it by $\Delta\tau$; then the largest phase difference among the folded periods will occur between the first and the last periods. If the length of the time series is L, then the maximum phase difference will be (L/τ) times $\Delta\tau$, and this value has to be smaller than $(\tau/2)$; that is;

$$\frac{L}{\tau} \Delta\tau < \frac{\tau}{2} \, . \tag{7}$$

Then the period step $\Delta\tau$ must observe the following condition:

$$\Delta\tau < \frac{\tau^2}{2L} \, . \tag{8}$$

3. SIMULATIONS

We present hereafter preliminary simulation results to illustrate the applicability of the method, as well as the model used to generate periodic time series.

The arrival times of emitted photons from a constant source follow a Poissonian succession, and the time interval between two consecutive detected photons constitutes the outcome of a stochastic variable. Its probability density function is given by an exponential function

$$p(t) = \lambda \cdot e^{-\lambda t} \, . \tag{9}$$

Starting from equation (9), it is possible to generate a time series by accumulating the interval times obtained by means of the following expression:

$$\Delta t_i = -(1/\lambda) \ln(\text{RND}) \, , \tag{10}$$

where Δt_i is the interval between the photons i and $i + 1$ arrival times, while RND is a uniformly distributed random number between 0 and 1. In this case, the parameter λ is constant, i.e., time independent. Nevertheless, if the emitting source has a periodic behavior, we expect λ to be a periodic function of time. Let us consider the simplest case in which the time series corresponds to a uniform background plus a pure sinusoidal component; then, λ can be expressed in the

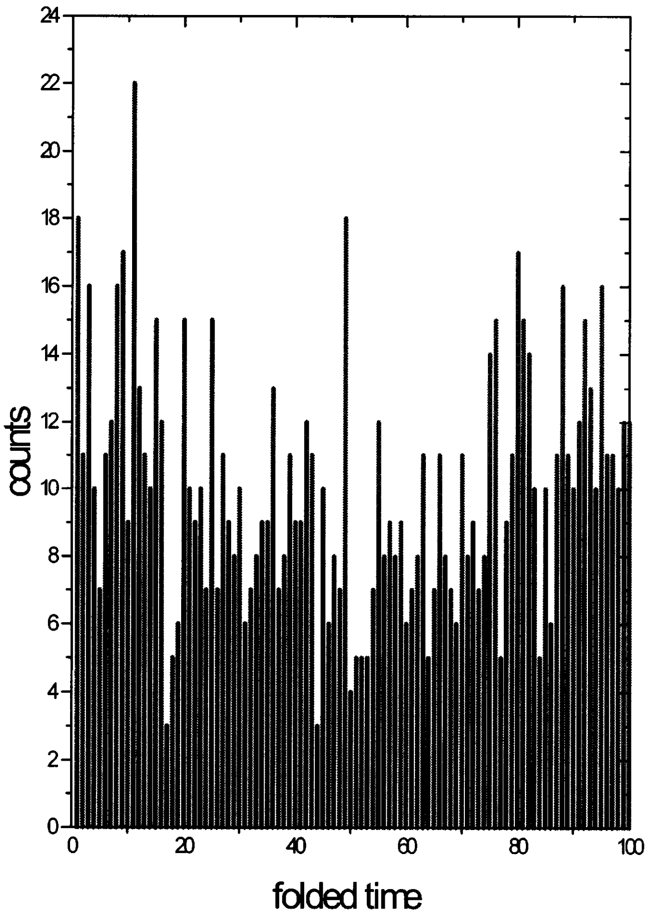

FIG. 1a

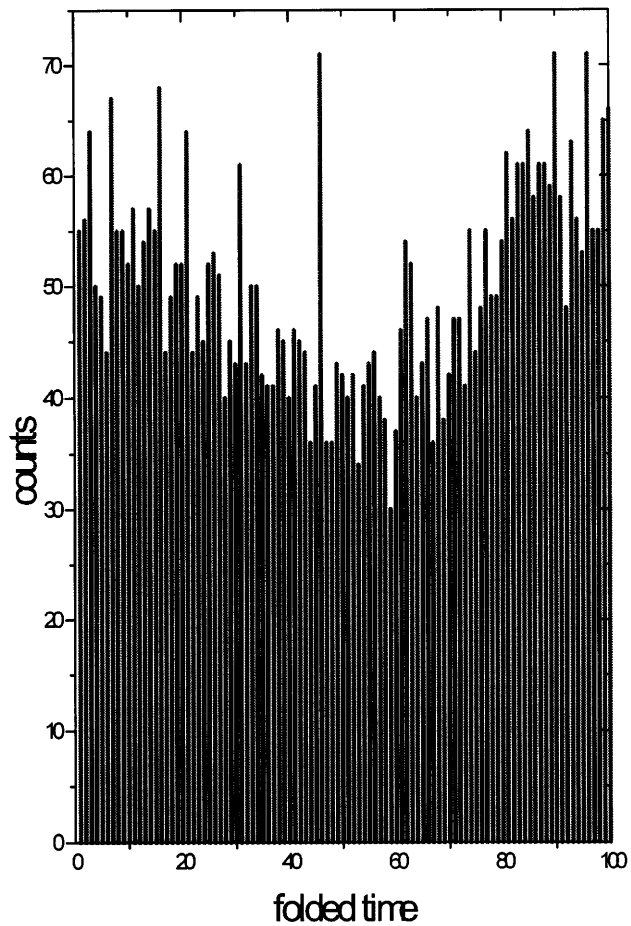

FIG. 1b

FIG. 1.—Histograms of folded time series. Signal period 0.3, folding period 0.3, arrival times: (*a*) 1000 photons and (*b*) 5000 photons.

following form:

$$\lambda(t') = \lambda_0(1 + \delta \sin \omega_0 t') , \tag{11}$$

where $\omega_0 = (2\pi/\tau_0)$ and τ_0 is the actual period.

Taking into account that $\lambda(t')$ is the mean number of counted photons in the time unit

$$\lambda = \frac{dN}{dt'} , \tag{12}$$

and that we explicitly may write

$$\frac{dN}{dt'} = \frac{dN}{dt} \frac{dt}{dt'} , \tag{13}$$

then

$$\lambda(t') = \lambda_0 \frac{dt}{dt'} . \tag{14}$$

Therefore, we can transform the *time* of a constant time series in such a way that the new time series may have a periodic component described by a sinusoidal parameter λ as in equation (11). By comparing the equations (11) and (14), it turns out that we can write

$$\frac{dt}{dt'} = (1 + \delta \sin \omega_0 t') , \tag{15}$$

and the transformation rule becomes

$$t = t' - \frac{\delta}{\omega} \cos \omega t' . \tag{16}$$

A numerical procedure was used to solve equation (16).

The above new time series described by $\lambda(t')$ has been the subject of our method to investigate whether a periodic component is present and to estimate the sensitivity of the method to detect periodicity.

If the amount of photons emitted by a source is described by the Poissonian distribution $P(\lambda, n) = (e^{-\lambda}\lambda^n/n!)$, then we can consider the single sources as two independent sources whose emission processes are also described by the Poissonian distributions $P(\lambda_1, n)$ and $P(\lambda_2, n)$ where $\lambda_1 + \lambda_2 = \lambda$, hence the following relation holds

$$P(\lambda, n) = \sum_{m=0}^{n} P(\lambda_1, m)P(\lambda_2, n - m) \quad \text{if } \lambda_1 + \lambda_2 = \lambda , \quad \text{and } n, m \in Z . \tag{17}$$

We use the previous analysis to interpret our single hypothetical source with parameter $\lambda(t) - \lambda_0(1 + \delta \sin \omega t)$ as two independent sources with parameters $\lambda_1 = \lambda_0(1 - \delta)$ and

TABLE 1

STATISTICAL RESULTS

TIME SERIES (Photons)	PERIOD ESTIMATION FUNCTION			
	Mean Value	Peak	Standard Deviation	S/N Ratio
1000........	1.136	6.25	0.88	5.8
5000........	0.931	16.27	0.68	22.5

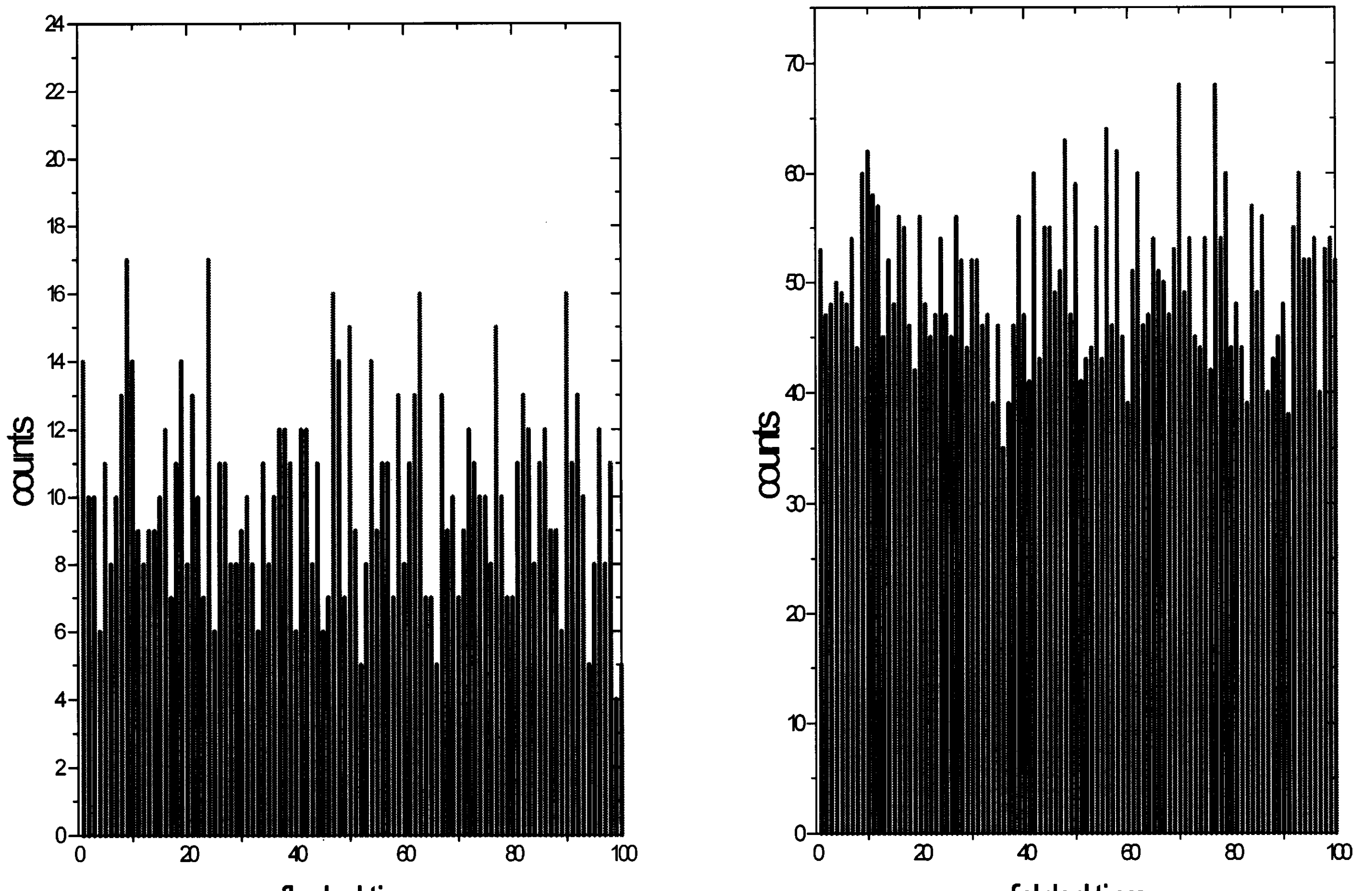

FIG. 2*a* FIG. 2*b*

FIG. 2.—Histograms of folded time series. Signal period 0.3, folding period 0.33, arrival times: (*a*) 1000 photons and (*b*) 5000 photons.

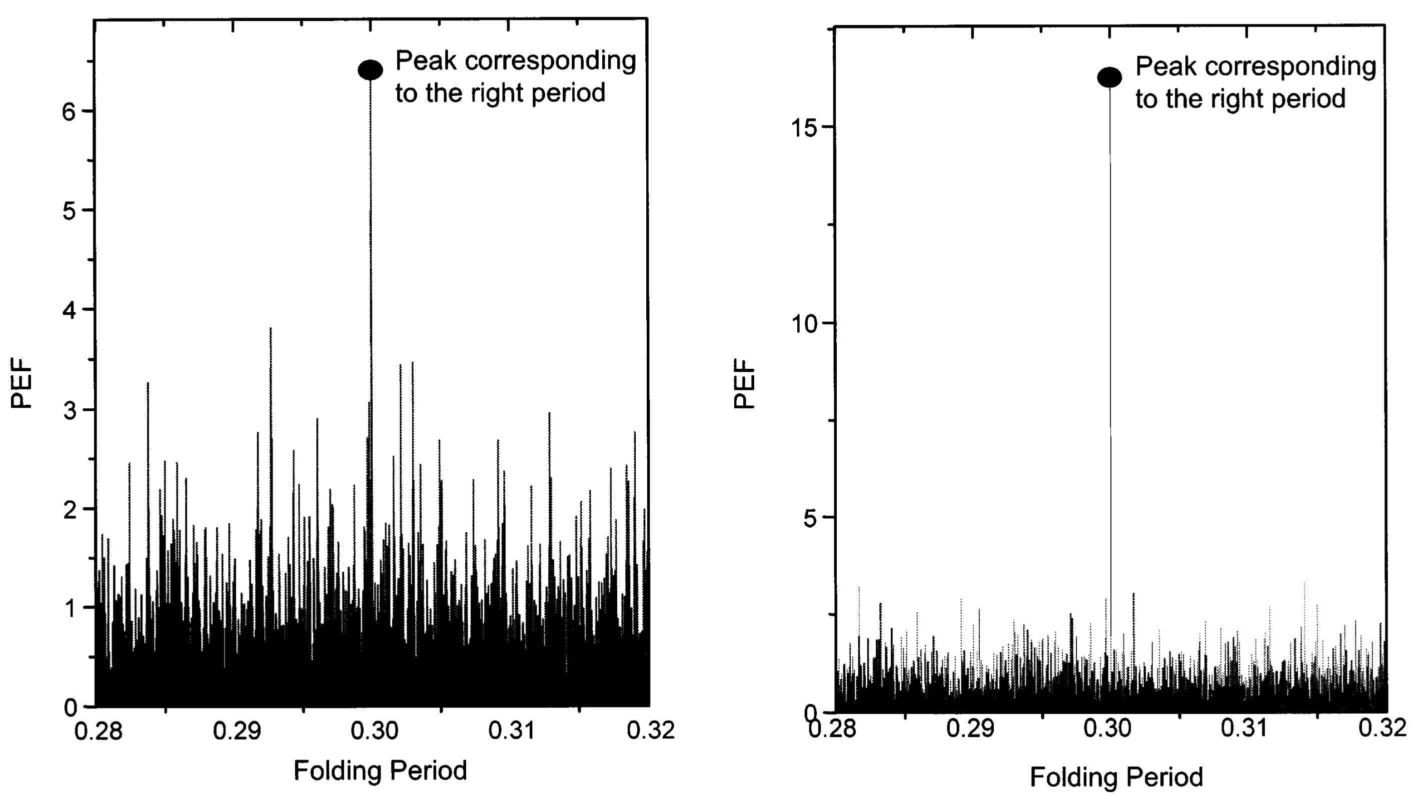

FIG. 3*a* FIG. 3*b*

FIG. 3.—Period estimation function of (*a*) 1000 photons time series and (*b*) 5000 photons time series.

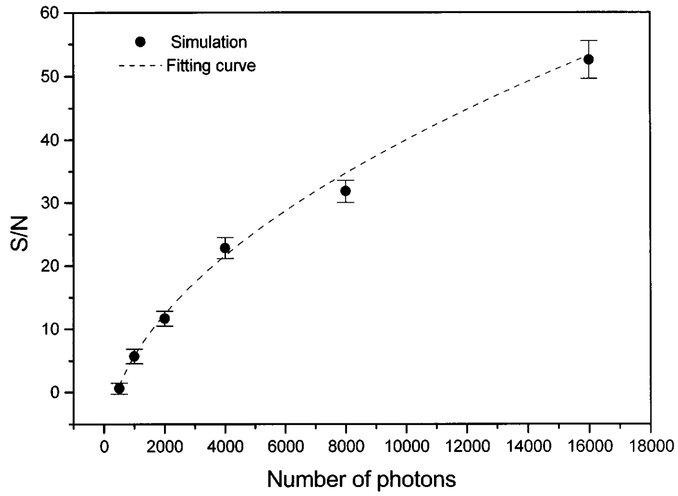

FIG. 4.—signal-to-noise ratio of PEF for time series with different number of photons. The error bars correspond to 2 σ. The analytic expression of the fitting curve is $(1/2)\sqrt{N} - 10$, where N is the number of photons.

$\lambda_2 = \lambda_0 \delta(1 + \sin \omega t)$, respectively. This allows us to consider some of the photons coming from a time-constant source and others coming from a pure periodic source. We can consider the photons emitted by the time-constant source (described by λ_1) as the *noise*, N_0, and the others as the *signal*, S_0. therefore, we can state the rate of their mean number as the signal-to-noise ratio of our series of arrival times.

The parameter λ_1 and λ_2 are the mean number of photons emitted per unit time; then, computing their temporal mean value, we obtain

$$\frac{S_0}{N_0} = \frac{\lambda_0 \delta}{\lambda_0(1 - \delta)} = \frac{\delta}{(1 - \delta)} . \tag{18}$$

We have generated two time series with 1000 and 5000 photon arrival times, respectively, both having $\lambda_0 = 1$, $\delta = 0.2$, and a period of 0.3 arbitrary units (see Fig. 1). In both cases the period after folding has been divided in $k = 100$ intervals.

Figures 2*a* and 2*b* display the same histograms as Figures 1*a* and 1*b* but obtained with a folding period of 0.33. The extent to which noise dominates the histograms is evident in all four cases.

We have applied the present method on both previous time series scanning the period from 0.28 to 0.32. The PEF is represented in Figure 3*a* for 1000 photons and in Figure 3*b* for 5000 photons, respectively.

Table 1 shows some statistical results that enables evaluation of the sensitivity of the method.

We have also generated time series of different lengths with the same mean interval time between consecutive photons. These series were generated with $\tau_0 = 10$ (arbitrary units), $\delta = 0.2$, and $\lambda_0 = 1$. The periodicity of the series has been explored for periods between 5 and 15 using a variable period step

$$\Delta\tau = \frac{\tau^2}{4L} , \tag{19}$$

which is half of the upper limit of the period step given in equation (8).

Figure 4 shows the achieved signal-to-noise ratio, where the error bars correspond to two standard deviations of the PEF. The fitting curve shows a dependency on the square root of the photon number for each time series, which is an expected feature for a typical Gaussian process. To construct the histogram representing the folded data it is possible to use different numbers of bins denoted by the corresponding parameter k. We have set the parameter $k = 20$, but similar results were obtained with values as high as $k = 200$, showing that k is not a critical parameter of the method.

4. CONCLUSIONS

The present method has proved to be suitable to process a small amount of data meaningfully. Most of the noncorrelated noise is suppressed by means of the autocorrelation function. An important feature of the method is its phase independence that makes the time reference irrelevant. The final signal-to-noise ratio is sufficiently acceptable even for as few as 1000 sparse photons, where, on average, 800 photons correspond to a time-constant source. We have applied the method to periodic time series with periods in the range of 10^{-2} to 10^{+2} of the mean interval time of consecutive photons, obtaining similar results to those presented here.

REFERENCES

Bai, T. 1992, ApJ, 397, 584
Bignami, G. F., Caraveo, P. A., & Mereghetti, S. 1993, Nature, 357, 704
Cincotta, P., Mendez, M., & Núñes, J. A. 1995, ApJ, 449, 231
Colavita, A. 1996, in Proc. 4th School Non-Accelerator Particle Astrophysics, ed. E. Bellotti, R. A. Carrigan, G. Giacomelli, & N. Paver (Singapore: World Scientific), 490
Fichtel, C. A., et al. 1994, ApJS, 94, 551
Gregory, P. C., & Loredo, T. J. 1992, ApJ, 398, 146
Kanbach, G., et al. 1988, Space Sci. Rev., 46, 69
Leahy, D. A., et al. 1983, ApJ, 272, 256
Sacco, B., et al. 1993, Nuovo Cimento C, 16, 715
Scargle, J. D. 1989, ApJ, 343, 874
Swanepoel, J. W. H., & De Beer, C. F. 1990, ApJ, 350, 754