Baltic Astronomy, vol. 7, 43–69, 1998.

PERIOD SEARCH IN LARGE DATASETS

Alex Schwarzenberg-Czerny[1,2]

[1] *Astronomical Observatory, Adam Mickiewicz University, ul. Słoneczna 36, 60-286 Poznań, Poland*

[2] *Copernicus Astronomical Center, ul. Bartycka 18, 00-716, Warsaw, Poland*

Received October 15, 1997.

Abstract. The number of papers on the analysis of unevenly sampled time series is scarce. The present article is an attempt to provide a fairly simple introduction to this topic. We start with the demonstration that some procedures for analysis of unevenly sampled time series, such as the power spectrum, suffer from a number of faults and traps which make them unreliable in practice. Then we consider the application of orthogonal models in statistics and testing of statistical hypotheses. Next, we demonstrate, how these classical principles of statistics can be adapted to the analysis of unevenly sampled time series. In this way we derive new, reliable methods for analysis of unevenly sampled time series. We discuss the relevant statistics, i.e. periodogram functions and their performance, and provide tools for planning efficiency of time series observations. These methods should be particularly useful for astronomers, since astronomical time series are often sampled unevenly in time.

Key words: methods: statistical – stars: variables

1. INTRODUCTION

Many observers would not read a paper on new statistical methods, arguing that by sticking to the old, classical methods (i) they know what they are doing and (ii) any conclusions drawn in a classical way are conservative and based on well tested methods. However, the interpretation of classical results relies on theoretical analysis. Until recently, the analysis was performed almost exclusively for evenly sampled time series. We shall demonstrate that extension of

this analysis onto the unevenly sampled time series, often encountered in astronomy, is questionable. The general aim of the present paper is to convince readers that the analysis of the unevenly sampled series differs significantly from the analysis of time series sampled unevenly. More specifically, we would like to demonstrate that (i) analyses, experiences and intuitions based on the application of the classical methods to evenly sampled signals often fail for unevenly sampled time series and (ii) that the old classical methods of statistics may be applied in a new way for the analysis of the uneven time series.

In most of our paper we refer to the classical theory of experiment elaborated by Fisher and collaborators early in this century. In Section 2 we discuss the power spectrum as a counter example, demonstrating how the well known classical method fails in new circumstances. Section 3 is devoted to fitting of the orthogonal models. The essential statistical concepts are presented in Section 4. The aspects of applications of these general concepts to a particular case of unevenly sampled time series are discussed in the following sections. In Section 5 we evaluate the performance of various methods. Short comments on analysis of multi-periodic signals are given in Section 6. Application of simulations to analysis of time series is discussed in Section 7. We conclude with discussion of time series analysis in large photometric surveys (Section 8). Many concepts discussed here are quite general. However, where the specific properties of a signal are considered, we draw most emphasis to periodic signals. In an astronomical context this means we pay more attention to stellar than to extra-galactic applications. The present review is rather biased by the author's preferences. For different aspects and contexts of time series analysis (hereafter TSA), the reader is referred to reviews by Feigelson (1997) and Scargle (1997). A readable introductory text on signal models and power spectrum was published by Deeming (1975).

2. A SINUSOIDAL SIGNAL SAMPLED UNEVENLY

2.1. Power spectrum

Let us start from observation that, for a given frequency ω, sine (cosine) discrete Fourier transform (DFT) constitutes a special case of scalar product of observations $x_k, k = 1, ..., n$ with the sine (cosine) function:

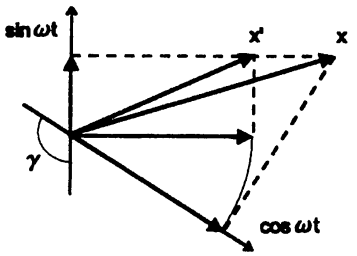

Fig. 1. The power, corresponding to $|x'|^2$, may be expressed by projections of a function x onto sine and cosine directions. For uneven sampling, sine and cosine are not orthogonal vectors ($\gamma \neq \pi/2$) and the power does not correspond to the squared amplitude: $\|x\|^2 \neq \|x'\|^2$.

$$(x, \sin\omega t) \equiv \sum_{k=1,n}^{n} x(t_k) \sin\omega t_k \qquad (1)$$

Let us assume, for this section alone, that our signal consists of a pure harmonic oscillation without noise. Adopting geometrical terminology it may be said that Fourier transforms are orthogonal projections onto the directions of sine and cosine in the space of functions of time. For uneven sampling, sine and cosine generally are not orthogonal functions in the sense of scalar product in Eq. (1):

$$(\cos\omega t, \sin\omega t) \neq 0 \qquad (2)$$

This is in marked contrast to the even sampling fast Fourier transform (hereafter FFT) case, where sine and cosine harmonics are orthogonal. Because of non-orthogonality, a sum of squares of sine and cosine projections, called the power $P(\omega)$, fails to satisfy the Pythagorean Theorem:

$$P(\omega) \equiv (x, \cos\omega t)^2 + (x, \sin\omega t)^2 \neq$$
$$\neq \|x\|^2 \equiv (x, x). \qquad (3)$$

The vector x and its sine and cosine projections do not form a rectangular triangle. A rectangular triangle formed of projections yields another vector x' (Fig. 1). Now recall that the squared amplitude corresponds to the vector norm $A^2 \equiv \|x\|^2 = (x, x)$ while power corresponds to another vector norm $P \equiv \|x'\|^2 \neq A^2$. Clearly, x coincides with x' only for $\gamma = \pi/2$, i.e. for sine and cosine orthogonal in Eq. (2). A simple but somewhat surprising conclusion following from

the inequality (Eq. (3)) is lack of any simple relation between the signal power P and the amplitude A for uneven sampling. By simple geometrical considerations one may demonstrate that the power remains restricted to a limited range $0 < P < 2A^2$, depending on the angle between sine and cosine components, $\pi > \gamma > 0$, respectively. Thus any popular methods exploiting DFT (Eq. (1)), such as power spectrum or CLEAN (Deeming 1975, Roberts et al. 1987), yield unreliable estimates of signal amplitudes for uneven sampling. For this reason one may question whether observers using power spectrum for unevenly sampled signals obtain a faithful estimate of amplitudes (Section 1 (i)). Clearly, better methods are required for the case of uneven sampling.

Please note that the failure of the power spectrum P in estimation of amplitudes is a different effect from aliasing (Section 3.3). This failure occurs at the signal true frequency ω_o while aliasing occurs because of interference between signal and sampling at some other frequency ω.

2.2. Modified spectrum

Lomb (1976) has observed that it is always possible to shift the phase of sine and cosine functions so that they become orthogonal in the sense of Eq. (2). The amount of phase shift required for this purpose generally depends on frequency, $\tau = \tau(\omega)$. Since the shifted sine and cosine functions are orthogonal, $(\sin(\omega t + \tau), \cos(\omega t + \tau)) = 0$, the projections onto their directions obey the Pythagoras theorem. Hence the corresponding modified power P_m carries full information on amplitude of the corresponding harmonic component $P_m = A^2$, if no components other than sine and cosine are present. If the signal consists of a sinusoid plus white noise then, for a given frequency ω, a direct correspondence exists between the modified power P_m and χ^2 for the least squares fit of the data with a sinusoid:

$$P_m(\omega) = \|x\|^2 - \chi^2(\omega) \tag{4}$$

where now $\|x\|^2 = (x, x)$ constitutes the total signal power in all frequencies ω (c.f. Eq. (7)). The total power is independent of ω and contains contributions from both signal and noise. Hence, according to Eq. (4), the plots of P_m or χ^2 against ω, called *periodograms*, are mirror reflections of each other.

So far we assumed that the signal contains no constant term. The value of the constant term is *a priori* unknown and it is usually

estimated from the data, e.g. as the average value of the data. Hence, the Lomb procedure involves fitting of three functions, sine, cosine and a constant term, instead of two functions. However, Ferraz-Mello (1981) has pointed out that for uneven sampling the constant term is not orthogonal with respect to the Lomb sine and cosine functions. If so, the Lomb spectrum is not equivalent to the least squares fit and yields a biased estimate of amplitude, correlated with the constant term. The distribution of the corresponding statistics is no longer exponential $\chi^2(2)$, as claimed. Ferraz-Mello (1981) has demonstrated, how to obtain a fully orthogonal model by application of the Gramm-Schmidt orthogonalization to the three functions listed above (c.f. Eqs. (19–21) in the Appendix). Examples and further discussion of the application of the Lomb functions with and without constant term are provided by Scargle (1982) and Foster (1994), respectively. Early references are listed by Ferraz-Mello (1981) and Press et al. (1992).

Unfortunately, the Gramm-Schmidt orthogonalization is so inefficient, that its application to any large set of functions would be prohibitively costly. An efficient procedure for generation of complex orthogonal trigonometric functions is discussed in Section 3.4. Summarizing, for uneven sampling the power spectrum yields a biased and non-optimal estimate of amplitude A. The Lomb-Scargle spectrum is better in that respect, but still involves a non-orthogonal constant term, hence it is neither unbiased nor optimal. We mean optimality in the sense of the least squares residuals and bias means that, despite improving signal-to-noise ratio (S/N), estimates do not converge to the true value of A. The model of Ferraz-Mello (1981) is fully orthogonal, hence it is free of problems haunting previous two models. We shall not discuss it here as it constitutes a special case of the model discussed in Section 3.4. Just for completeness, we observe that for even sampling all three models are identical, unbiased and optimum.

3. MODELS OF SIGNALS

3.1. Classification

Time series are functions, albeit discrete ones. Although there are theorems which allow classification of a given function, their assumptions such as continuity or infinite length of observation interval are too unrealistic to be useful in practice. In other words, by no

method one can be sure that a given discrete and finite series of observations represents a periodic function, a pure random process or the mixture. For example, a series appearing random may have the period longer than the interval of observations, or a purely random process can produce a perfect section of a sinusoid. We may only judge how likely or unlikely are these interpretations. In this judgment we often use our prejudice about the nature of the observed process. Thus, in the final instance, TSA stays on the assumed properties of a signal, i.e. on the *signal model.*

Three broad classes of signals exist: deterministic, chaotic and stochastic signals. These types of signals differ by the character of dependence of their value at time t on their previous value at time $t - \Delta t$. More precisely, the classification depends on the absolute value of the correlation $|\rho(t, t - \Delta t)| \approx |\rho(\Delta t)|$ as a function of the time interval Δt. For deterministic signals $|\rho| \approx 1$ for arbitrary long Δt. For stochastic signals, $|\rho| << 1$ for arbitrary short Δt. For chaotic process, $|\rho| \to 1$ for $\Delta t \to 0$ and decays to small values $|\rho| << 1$ for $\Delta t > \tau$, where τ is the correlation decay time, characteristic for a given chaotic process. It is related to the so-called Lyapunov exponent $1/\tau$. It is clear from the above definitions that deterministic and stochastic signals are extreme cases of chaotic signals corresponding to $\tau = \infty$ and 0, respectively. Often these definitions are relaxed in the sense that instead of *all* Δt one means *all* Δt *of interest.* General chaotic processes received little attention from astronomers (but c.f. Scargle 1989, Buchler 1993). Solar spots constituted an early example of stochastic time series. Nowadays astronomical stochastic time series are considered mostly in X-ray and extragalactic contexts. Deterministic signals, or, more specifically, (multi-)periodic signals, are often encountered in stellar astronomy. In TSA of both stochastic and deterministic signals, explicit function models are used. The technical difference is that for the stochastic processes one assumes a model shape of expected autocorrelation function or its Fourier transform power spectrum, while for the deterministic processes one assumes a shape of the signal itself (Deeming 1975). In the present review, when appropriate, we shall concentrate on periodic deterministic models.

3.2. Least squares

Rarely the observations x and the model values $x_{||}$ are exactly the same. Normally, the residuals $x - x_{||}$ are not vanishing. The

residuals are used to formulate the criteria for selection of optimum model parameters, such as the maximum likelihood and entropy criteria or least squares (hereafter LSQ) criterion. According to this criterion one selects a set of model parameters y minimizing norm of residuals: $\min_y = \|x - x_{\|}(y)\|^2$. For the Gaussian errors of observations and the linear model function $x_{\|}(y)$, the LSQ criterion is optimum in a certain sense. LSQ may be also used if errors are so small that within their range the linear approximation of a nonlinear $x_{\|}(y)$ is accurate enough. This property underlines the role of linear models in statistics. However, in general case, LSQ may perform worse than other criteria. For selection among models with different number of parameters one should apply a modified criterion, namely, the LSQ *per degree of freedom* criterion:

$$\min_y = \frac{\|x - x_{\|}(y)\|^2}{n - r} \tag{5}$$

where by the number of degrees of freedom $n-r$ we mean the number of observations in excess of r, the number of parameters y. A general version of this criterion was proposed by Akaike (1973).

3.3. General orthogonal models

All methods in time series analysis rely on a model of signal, built in either explicitly or implicitly. In Section 2 we have explained already that, depending on sampling, either the original power spectrum P or its modification P_m correspond to the least squares fit of the data with a sine function. By fitting a model one decomposes the observed signal x into its model and residual components, $x_{\|}$ and $x_{\perp} = x - x_{\|}$. A particularly efficient way of performing a linear LSQ fit is the projection onto orthogonal base of model functions $z^{(k)}$, $k = 1, ..., r$. In the language of linear algebra one speaks about the vector space of all possible signals x and its subspace of the model signals $x_{\|}$ with base z. Then the fitted model is computed as follows:

$$x_{\|} = \sum_{k=1}^{r} (x, z^{(k)}) z^{(k)}, \tag{6}$$

where $x = x_{\|} + x_{\perp}$ and $y_k \equiv (x, z^{(k)})$ are model parameters. By virtue of Fisher's lemma, model and residuals are orthogonal (Eq. 8),

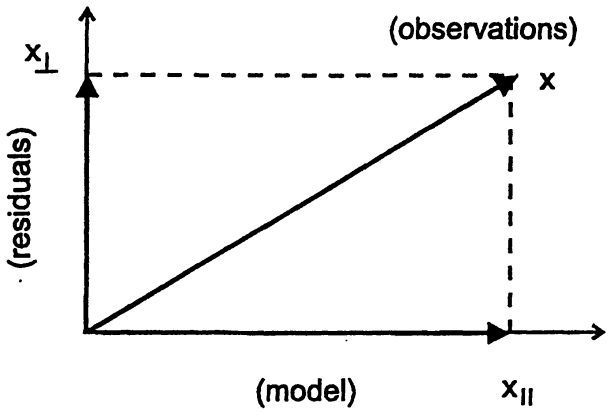

Fig. 2. Split of a signal into orthogonal components: model $x_{\|}$ and residuals $x_{\perp}$.

so that the components $x_{\|}$ and $x_{\perp}$ satisfy the Pythagoras relation (Fig. 2):

$$\|x_{\|}\|^2 + \|x_{\perp}\|^2 = \|x\|^2 \tag{7}$$

$$(x_{\|}, x_{\perp}) = 0 \tag{8}$$

Eq. (7) may be interpreted in terms of decomposition of the total signal power into the model and noise power. In the following sections we discuss important forms of the orthogonal functions used for modeling the signals.

On one hand, it is not difficult to see that as soon as observations are fixed, $\|x\|$ is independent of frequency ω and of other model details. On the other hand, the split of the power among the components $\|x_{\|}\|$ and $\|x_{\perp}\|$ depends on ω. Plots of $\|x_{\|}\|$, $\|x_{\perp}\|$ or their functions against ω are called *periodograms*. In most of the present paper we discuss the properties depending on orthogonality of model functions *for a given* ω. This kind of orthogonality does not guarantee orthogonality of the model functions corresponding to different frequencies. In fact, the only known set of periodic functions, orthogonal on a grid of frequencies, are FFT harmonics. Generally, no set of functions orthogonal between frequencies exists for uneven sampling. This follows from the fact that functions for each ω are already prescribed and only tunable parameters are frequencies. However, n observations may be converted by an orthogonal transformation to no more than n frequencies, not enough to satisfy $(n-1)r/2$ orthogonality conditions. The lack of orthogonality between the frequencies

for the uneven sampling results in a correlation between the values of periodogram at these frequencies. This effect persists with no regard for model and periodogram in use. The correlation of nearby values of periodogram is called the power leakage. The correlation of distant regions of periodogram is called aliasing. Further effects of correlation between frequencies are discussed in Section 4.4.2.

3.4. Fourier series model

Orthogonal sine and cosine functions are used as model functions in a modified power spectrum (Lomb 1976, Ferraz-Mello 1981, c.f. Section 2.2). However, there are lots of strongly non-sinusoidal signals for which the sinusoid constitutes a poor model. For these signals, a Fourier series consisting of multiple harmonics constitutes a better model. For even sampling, the Fourier harmonics are orthogonal. However, for uneven sampling they are no longer orthogonal, so that the orthogonal projections are useless. Fitting observations with the Fourier series directly by least squares or by the Gramm-Schmidt orthogonalization constitutes so slow and possibly ill-conditioned algorithms, that no period search method in the past used this model. The situation has changed with the notion that fast recurrence formulae exist for generation of trigonometric polynomials orthogonal on uneven grid of observations (Grenander & Szegö 1958). This enabled formulation of a computationally efficient yet sensitive multi-harmonic method, based on the fit of Fourier series by means of projection onto orthogonal trigonometric polynomials (Schwarzenberg-Czerny 1996). The advantage of this method may be best appreciated by inspection of Fig. 3. In this example, the multi-harmonic periodogram reveals a strong signal detection with no alias ambiguity. Yet the corresponding power spectrum is prone to multiple closely spaced aliases. There are two causes of good performance of the multi-harmonic method. The first reason is of a statistical nature: the multi-harmonic Fourier series fits non-sinusoidal signals considerably better than a pure sinusoid does. This means that less power is left in residuals $x_\perp$, corresponding to higher sensitivity. The second reason is of a graphical nature: since we plot $\|x_\parallel\|^2/\|x_\perp\|^2$, our periodogram is visibly sensitive to the small $\|x_\perp\|$, marking a good fit.

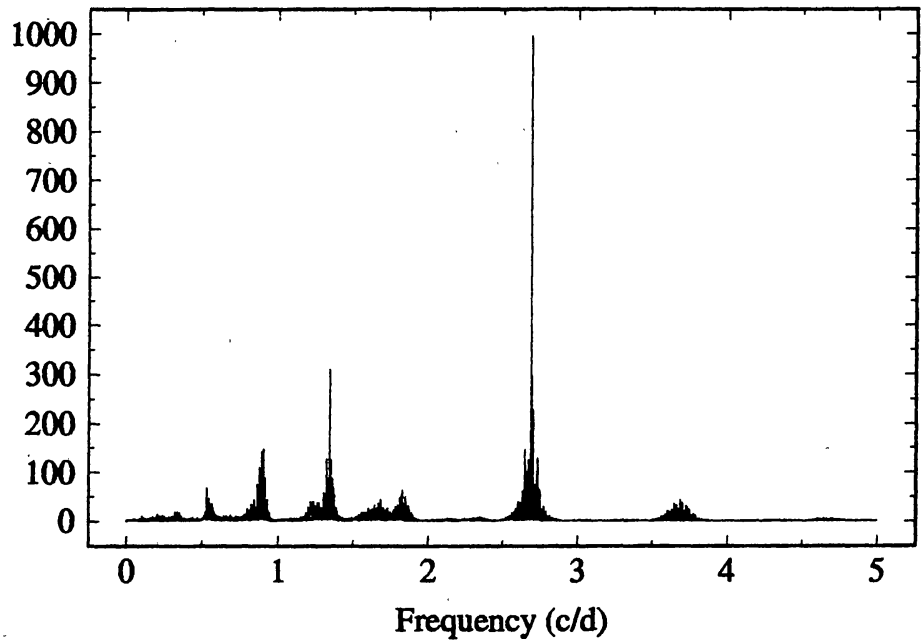

Fig. 3. Multi-harmonic periodogram using orthogonal trigonometric polynomials. Note the absence of alias ambiguity, in marked contrast to ordinary power spectrum for the same data (V15 in M68, c.f. Fig. 6 of Walker 1994). Sole ghost features are sub-harmonics of the base frequency.

3.5. Compact support models

Compact support base functions (hereafter CSF) vanish everywhere except for a narrow phase interval. An example of an orthogonal CSF base are top hat functions covering consecutive phase intervals. All phase folding and binning methods implicitly use these orthogonal top hat functions for modeling of signals. Another example of orthogonal CSF are spline bell functions employed by Akerlof et al. (1994). Although original bell functions are not orthogonal as they partially overlap, the MACHO team uses pre-computed matrices for transformation between bell functions and their orthogonal combinations. A distinct advantage of CSF in time series analysis is a weak dependence of their computational efficiency on complexity of a model. Still, for the complex models using many CSF functions, an adequate sampling requires a frequency grid which is dense in comparison to the corresponding Fourier grid, at extra computational

cost. Detailed investigations reveal that CSF models suffer from uneven sensitivity as a function of phase difference between signal and base functions. This may be caused by a poor correspondence between shapes of the real signals and CSF (Schwarzenberg-Czerny 1997b). Wavelets constitute another example of orthogonal CSF. Although attempts to model periodic functions with wavelets are scarce, efforts were made to use wavelets to model and subtract noise from unevenly sampled observations (Lehto 1997, Scargle 1997). Another interesting application of CSF is a time-frequency analysis. In this analysis one considers not just power of a signal but also its coherence length (Mallat & Zhang 1993, Roques et al. 1996).

4. FIT QUALITY

4.1. Quality statistic

It may appear a bit paradoxical that most information on statistical properties of a fitted model is contained in the residuals from the fit, $x_\perp$. This information concerns the type of distribution, correlation of observations, significance of model detection in observations, confidence intervals of the fitted parameters and their covariances. All period search methods involve a measure of the fit quality between observations and the model. This measure is a function of both model parameters and of observations. Since observations are affected by errors, they are random variables. A value of the function of random variable is a random variable too. Such a function is called a *statistic*. Among different types of statistics used as a measure of the fit quality, the most important ones are related to norm of residuals $\|x_\perp\|$. For the orthogonal models, the Pythagoras theorem (Eq. (7)) provides a simple relation of the model and residual norms:

$$\|x_\parallel\|^2 = \|x\|^2 - \|x_\perp\|^2 \tag{9}$$

where the total power $\|x\|$ is a constant independent of the model and frequency. Particularly useful are dimensionless ratios of these norms (c.f. Section 4.4.3). Such ratios are listed in Table 1.

A brief derivation of the corresponding probability distributions is given in Bickel & Doksum (1977). Note that because of Eq. (7) these ratios are unique functions of each other:

Table 1. Classification of the least squares statistics.

Statistic Θ	Distribution[a,b]	Applications
$\frac{\|x_\parallel\|^2}{\|x\|^2}$	$I_\Theta(d_\parallel, d_\perp)$	Lomb-Scargle Spectrum[c] Whittaker & Robinson statistic[d]
$\frac{\|x_\perp\|^2}{\|x\|^2}$	$I_\Theta(d_\perp, d_\parallel) = I_{1-\Theta}(d_\parallel, d_\perp)$	χ^2 method, PDM method[d]
$\frac{\|x_\parallel\|^2}{\|x_\perp\|^2}$	$F_\Theta(d_\parallel, d_\perp)$	AOV periodogram[e], Multi-harmonic periodogram[f]

a I and F are the Beta and Fisher-Snedecor distributions (Abramovitz & Stegun 1971, Bickel & Doksum 1977);
b $d_\parallel = r$ and $d_\perp = n - r$ are the numbers of degrees of freedom in model and residuals, respectively;
c Lomb 1976, Scargle 1982;
d Stellingwerf 1978;
e Schwarzenberg-Czerny 1989;
f Schwarzenberg-Czerny 1996.

$$\frac{\|x_\parallel\|^2}{\|x\|^2} = 1 - \frac{\|x_\perp\|^2}{\|x\|^2} = \frac{1}{1 + \left(\frac{\|x_\parallel\|}{\|x_\perp\|}\right)^{-2}}. \tag{10}$$

To illustrate the properties of these distributions, we consider here the phase dispersion minimization (hereafter PDM) statistic corresponding to the $\|x_\perp\|^2/\|x\|^2$ ratio for the step function model (Stellingwerf 1978). It follows from Table 1 that the PDM statistic is the Beta probability distribution (Fig. 4b). Its originally claimed distribution was the Fisher-Snedecor distribution. These two distributions are markedly different (Fig. 4a). Note that the Beta distribution is highly asymmetrical and thus is awkward to use by the observers who are used to the Gaussian distribution. For this reason we proposed to use the analysis of variance (hereafter AOV) statistic $\|x_\parallel\|^2/\|x_\perp\|^2$ or its log value, since they follow nearly symmetrical F-S or Snedecor distributions (Schwarzenberg-Czerny 1989, 1996). Since all statistics listed in Table 1 are uniquely related, statistical conclusions do not depend on which one is used, provided that the distribution and the statistic are matching.

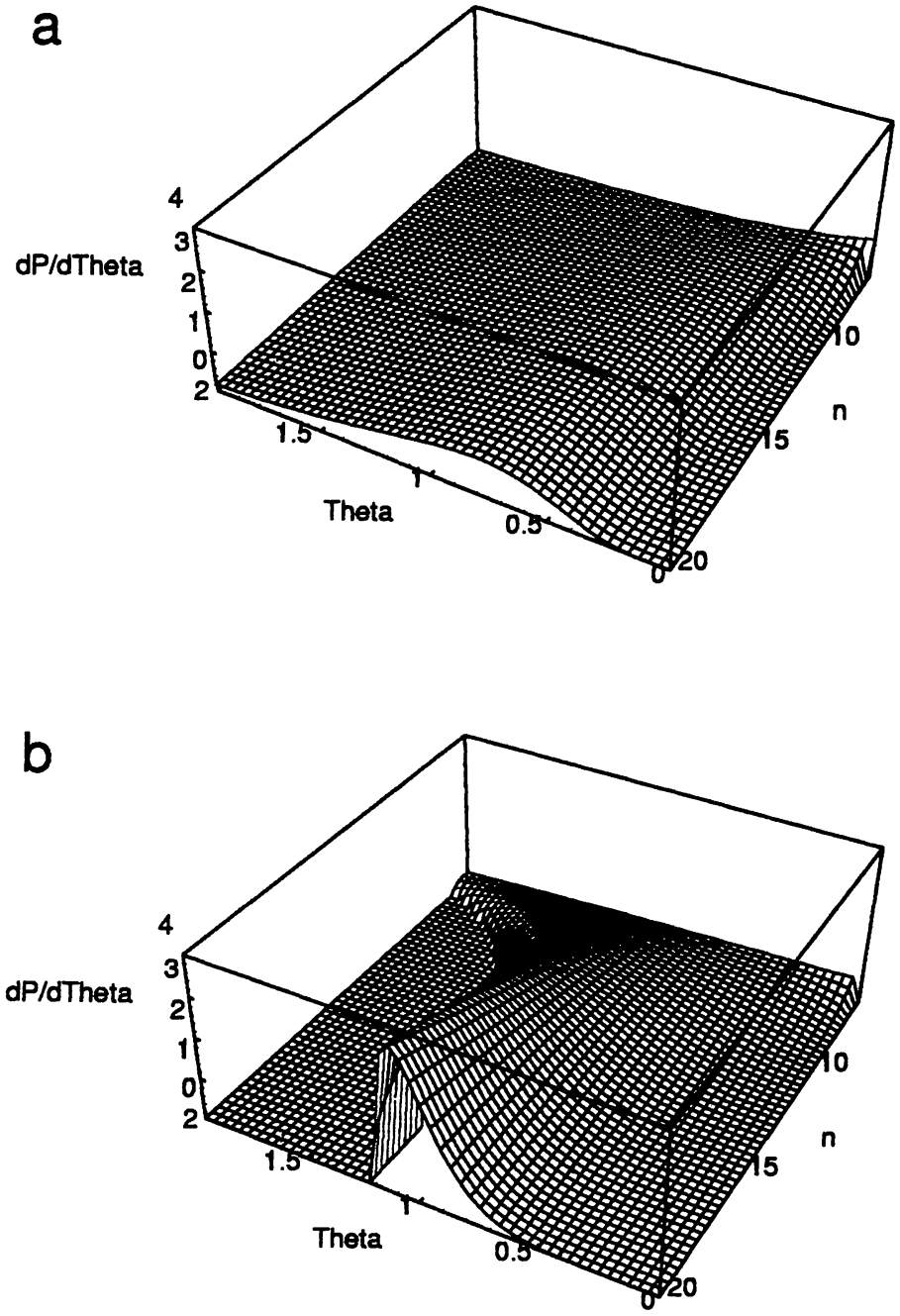

Fig. 4. Families of (a) incorrect, F-S, and of (b) correct, Beta distributions for the PDM periodogram. Each line in a family corresponds to the probability distribution for a given number of observations n.

It is remarkable that virtually every method used for a period search in astronomy may be reduced to one of these statistics with a suitable model. In the present paper we discuss the power spectrum, Lomb-Scargle spectrum, PDM and AOV methods. The asymptotic relation of string-length methods with the statistics of Table 1 was discussed by Schwarzenberg-Czerny (1989). These relationships enable us to find the correspondence between different methods used in practice. The relations combined with the test power theory presented in Section 5 enable the evaluation and comparison of sensitivity of different methods. Rarely used combinated methods do not fall within the present scheme, however.

4.2. Classical detection criterion

In order to asses the significance of the detected signal one compares two situations: either the observed signal contains a pure noise or it contains some deterministic (function) component plus random errors. In statistics the former, undesirable situation is called the null hypothesis H_0, the latter, desirable situation is called the alternative hypothesis H_1. Let the observed value of the statistic be Θ. One considers the probability of observing the larger value $P(\Theta' > \Theta)$ for the null hypothesis H_0 valid, i.e. for pure noise. The complement probability $Q = 1 - P$ is called the significance of detection of the model yielding Θ fit. This significance may be expressed directly as a probability, in percents or in corresponding deviations of the Gaussian distribution (e.g. 0.995, 99.5% or 3σ). Related questions in statistics are called the *hypotheses testing*. In the classical, non-Bayesian statistics one adopts *a priori* certain critical value P_c and considers the detection significant if $P < P_c$. A complement critical probability $\alpha = 1 - P_c$ is called the confidence level. Such a detection criterion depends on the distribution of Θ for the pure noise, i.e. for H_0 valid. Its independence of a possibly complex distribution for compound signal corresponding to H_1 constitutes a practical advantage. The relevant distributions are listed in Table 1.

4.3. Bayesian detection criterion

Unfortunately, statisticians are divided on the basic assumptions of the procedures they use. Bayesian statisticians for hypotheses testing adopt *a priori* distributions of model parameters, called the *prior distributions*. From these prior distributions and distribution

of observations they derive a detection criterion specific to the given problem. The dependence of their criterion on the problem is considered as a disadvantage by critics of Bayesian statistics. Fortunately, for the period search in large number of observations the Bayesian criterion reduces to the asymptotic form $P < P_B$ resembling the classical criterion. Although the Bayesian and non-Bayesian criteria are different in general, $P_B \neq P_c$, they coincide for a particular set of prior distributions. Critics of the classical method point out that these particular prior distributions seem unrealistic. In the absence of conclusive arguments in the Bayesian dispute, only taste and tradition motivate our preference of classical, non-Bayesian method in the present discussion.

4.4. Corrections to distributions

4.4.1 Correlated residuals

In Fig. 5 we display the Wolf solar spot number as a function of time. Formal errors of the least squares fit of this function with a sinusoid are surprisingly small, 1% in period. Yet inspection of Fig. 5 reveals that the observed and fitted curves are often up to half a period out of phase. In fact, the solar spot cycle constitutes a well known example of an non-periodic (stochastic) process, which forgets its phase of oscillation just after a few cycles. This example demonstrates one danger in TSA: you may get a formally excellent fit of no physical and statistical significance. Formal errors of periods derived from fast photometry, as a rule, suffer from this problem. Here we discuss methods which enable us to identify such problems and to rectify the results of analysis.

The distributions discussed in statistics for the H_0 hypothesis are derived for observations containing Gaussian white noise. This corresponds to assumptions that (i) all observations follow the same Gaussian distribution, (ii) their mean is 0 and (iii) *that they are uncorrelated.* Inspection of the residuals encountered in practice reveals that assumption (iii) is often violated. The residuals become correlated for the reasons connected either to the source and propagation of a signal or because the model is too coarse to follow the structure of a signal. These correlations affect the distribution of Θ by decreasing the effective number of observations $n_{\rm eff}$. It is quite easy to recognize the correlation by estimation of an average number of adjacent residuals of the same sign. If among $n_{\rm obs}$ observations the

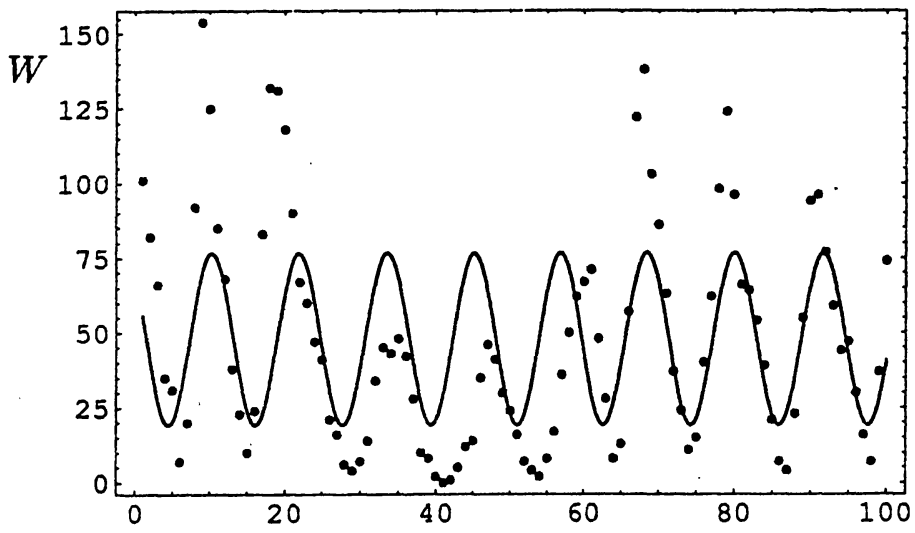

Fig. 5. Wolf number of solar spots as a function of year number and its least squares fit with a sinusoid. The formal error of a period fit is 1% and one would naively expect the consistency of true and fitted maxima within a few percent of their period, while, in fact, the discrepancies reach up to half of the period. See text for explanation.

groups of $n_{\rm corr}$ consecutive points are correlated, on average, then the effective number of observations is $n_{\rm eff} = n_{\rm obs}/n_{\rm corr}$ and the true errors are a factor of $\sqrt{n_{\rm corr}}$ larger than these obtained from the least squares fit routine (Schwarzenberg-Czerny 1991). In Fig. 5 up to 30 consecutive residuals have the same sign, indicating the true period error of the order of $1\sqrt{30} \approx 5\%$ and the phase error, accumulated over 10 cycles, of the order of 50%. Such a large error warns against inconsistency of the signal and the model.

4.4.2. Bandwidth penalty

The distributions discussed so far refer to the situation in which one considers a specific frequency. In practice, many frequencies in a periodogram are scanned in pursuit of a detectable signal. This corresponds in statistics to multiple trials. Clearly, the probabilities of success, i.e. of detection in single and N trials, P_1 and P_N, respectively, are different. If trials corresponding to different frequencies are independent, then

$$P_N = 1 - (1 - P_1)^N \approx NP_1 \qquad (11)$$

where the last equality holds only for $NP_1 \ll 1$. For uneven sampling, the Θ statistics at different frequencies are generally correlated because of power leakage and aliasing. Then the identification of the effective number of searched independent frequencies n becomes difficult. Horne & Baliunas (1986) discuss the Monte Carlo method of estimation of effective N. Although their method is correct, it is incorrect to extrapolate their results by means of their fitted empirical formulae. Far extrapolations using these formulae yield nonsensical results, namely, N exceeding either the number of computed frequencies or the number of observations n.

4.4.3. Theoretical versus empirical distributions

Following the null hypothesis H_0, one assumes that observations constitute a pure white noise of known variance σ^2. If so, then, except for a constant factor, norms $\|x_\parallel\|^2$ constitute sums of squares of normal random variables and thus they follow the χ^2 distribution. For this reason, statisticians advise one to use χ^2 in the data analysis. Distributions, obtained in this way, we call *theoretical distributions*. The constant normalization factor, variance σ^2, has to be used since the statistical tables are prepared for a unit variance. Unfortunately, since observers do not know *a priori* the variance of their data, they have to estimate it from the same data using another norm, say $\hat{\sigma}^2 = \|x\|^2/(n-1)$, where n is the total number of observations. However, the variance of the data σ^2 is a constant parameter of the distribution while $\hat{\sigma}^2$ is a statistical, i.e. a random variable depending on random observations. So, by dividing the norms by $\hat{\sigma}^2$, observers obtain the ratios of two random variables, $\|x_\parallel\|^2/\|x\|^2 \equiv \Theta_\parallel$ or $\|x_\parallel\|^2/\|x\|^2 \equiv \Theta_\perp$. These ratios follow the Beta distribution (Table 1). Distributions, obtained in this way, we call *empirical distributions*. Although it is often argued that for the large number of data empirical distributions converge to theoretical ones, this convergence is fast only near the center of the distribution and not in its tails. In fact, these tails are used most often by observers in evaluation of the significance of signal detection.

To demonstrate this subtle effect, we consider the L-S statistic. Its theoretical distribution derived by Scargle is exponential, $Q_{1\mathrm{exp}}(x) = 1 - P_1(x) = e^{-x}$, corresponding to χ^2 (2). Its empirical Beta distribution may be evaluated analytically, yielding $Q_{1\mathrm{Beta}}(x) = I_{1-x}(n/2,1) = (1-(2z/n))^{n/2}$. For realistic applications, one requires P_N rather than P_1 (Section 4.4.2). In this

example we assume so good frequency sampling that $N \approx n$ and employ Eq. (11) to convert Q_1 to Q_N. The empirical and theoretical distributions Q_N computed in this way are plotted in Fig. 6. In Gaussian units, the differences reach $\pm 1\sigma$ at probabilities corresponding to 3σ. Clearly, the Beta and χ^2 distributions are not the same.

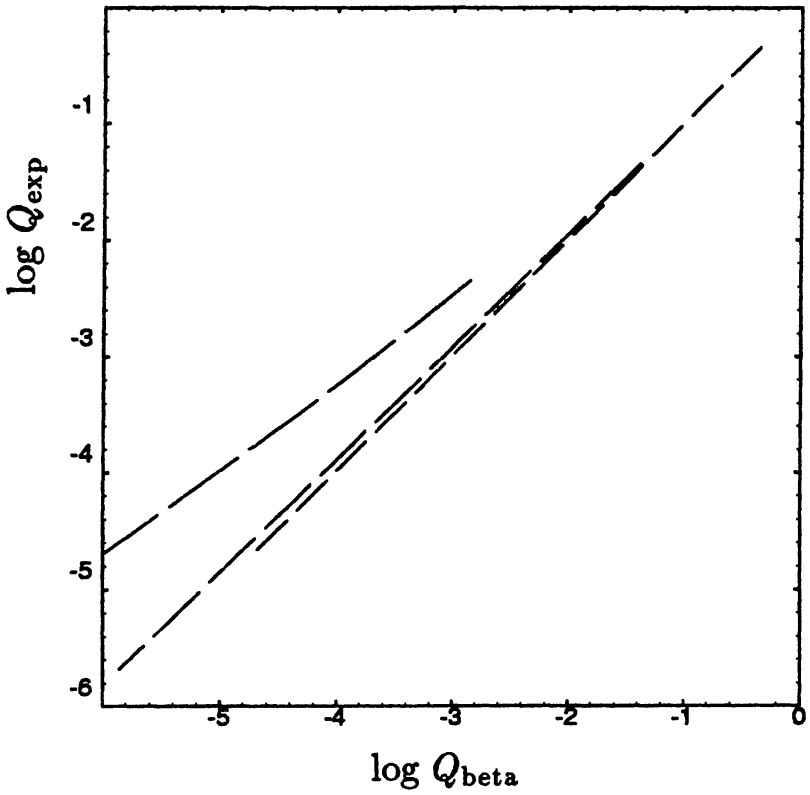

Fig. 6. Comparison of theoretical and empirical probability distributions, corrected for bandwidth, for the Lomb-Scargle statistic. Long, medium and short dash curves correspond to n =100, 1000 and 10000 observations. Note that in Gaussian units the differences reach $\pm 1\sigma$ at probabilities corresponding to 3σ.

5. SENSITIVITY AND TEST POWER

One advantage of classical, non-Bayesian statistics is the independence of detection criteria on the shape of the input signal. On one hand, by setting the fixed significance level α we accept the fixed rate $1-\alpha$ of false detection with no regard for the method and quality of observations. On the other hand, the rate of detected true signals

β depends on the method and signal-to-noise ratio S/N. These effects cannot be avoided in non-Bayesian statistics. The fraction β is called by statisticians the *test power*. The larger the test power β, the more sensitive is a given method for detection of particular signals. The sensitivity depends on the following properties:

(1) on the signal amplitude and its shape;
(2) it increases with the number of observations and signal-to-noise ratio;
(3) for real smooth signals, the smooth (e.g. Fourier) models perform better than the step models (binning);
(4) the best sensitivity is obtained for the models (harmonics) matching the data in resolution.

The sensitivity does not depend on which of the statistics listed in Table 1 is used, provided that the detection criterion is based on matching probability distribution (Schwarzenberg-Czerny 1997a). This means that all methods from a broad class using χ^2-like norms of residuals are equivalent, as long as they use the same model of a signal. It follows from property (1) that no single method exists which is optimum for all kinds of signals. Because of this property, a comparison of performance of different methods is not possible in general.

Fairly general formula may be obtained for an asymptotic case of small amplitudes, $A \to 0$ (Schwarzenberg-Czerny 1997b). The formula enables a quantitative evaluation of gains or losses in sensitivity β due to effects (1)–(4):

$$1 - \beta \approx R\left(R^{-1}(1-\alpha) - A^2 d \frac{\|s_\parallel\|^2}{2\sqrt{d_\parallel}} \right), \qquad (12)$$

where in the asymptotic limit of large number of observations the function R is defined as follows: $R(\Gamma) \to [1 - erf(\Gamma)]/2$. Thus in the limit, the power is a unique function of the fractional fitted power $\|s_\parallel\|^2 \equiv \|x_\parallel\|^2/\|x\|^2$. In Fig. 7 we present the results of the sample calculations of $\|s_\parallel\|^2$. The calculations were performed for a family of von Mieses input signals, a periodic analogue of the Gaussian Bell. Their results demonstrate that the use of trigonometric polynomials ensures higher sensitivity than the use of step functions (i.e. phase binning), at least for the assumed signal shape.

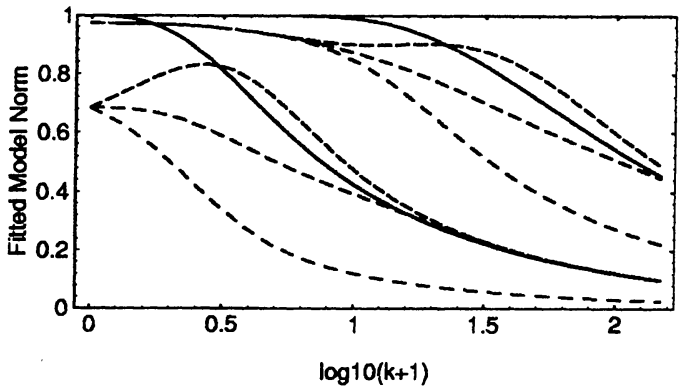

Fig. 7. Fractional fitted power $\|x_{\|}\|/\|x\|$ plotted against von Mieses shape parameter κ. The detection sensitivity is a unique function of $\|x_{\|}\|/\|x\|$. Small and large κ correspond to near sinusoidal and narrow pulse input signals, respectively. Continuous lines correspond to trigonometric orthogonal polynomials, families of dashed lines correspond to phase bins (step functions) with different phase offsets. The top curves correspond to an 11 parameter model (5 harmonics or 11 bins) and the bottom curves are from the three-parameter model (pure sinusoid or 3 bins).

6. MULTI-PERIODIC SIGNALS

Multiple periodic stars play an important role in research of pulsating stars. The detection of multiple periods is a difficult task from a statistical point of view. The use of truly multi-periodic models would require multi-dimensional frequency grids. The number of statistical trials, corresponding to such large grids, should be large enough to make rather strong detections insignificant after bandwidth correction (c.f. Section 4.4.2). A more practical method relies on a consecutive identification of the strongest oscillation in the remaining signal and its subsequent removal by prewhitening, i.e. subtraction of the least squares sinusoid. The methods of such type in statistics are called sequential analysis (e.g. Eadie et al. 1971). As the course of analysis in sequential methods is driven by the data, their statistical properties depend on the data too. Hence, no general discussion of properties of these methods is feasible. In particular, the sensitivity of these methods strongly depends on the data.

7. SIMULATIONS

Statistical properties of the period search methods can be investigated using simulations. Two types of numbers are used in simulations: Monte Carlo simulations rely on random number generators, and bootstrap and jack-knife methods rely on shuffling of the original observations. The simplicity, adaptability to complex situations and reliable estimates of low moments of relevant distributions constitute the advantages of simulations.

Unfortunately, simulations suffer from disadvantages, too. The tails of simulated distributions rely on rare events from random number generators. Random number generators and Monte Carlo algorithms are untested and unreliable for these rare events. In extreme cases, the effects of discrete representation of machine numbers may influence the results for rare events. At the same time, the generation of sufficient number of rare events for reasonable accuracy in the distribution tails becomes computationally very costly. For similar reasons, the estimation of high moments of distributions by Monte Carlo methods becomes both unreliable and inaccurate. We conclude this section with the suggestion not to use simulations in analysis of large data samples. For such data samples, many classical statistical methods have known analytical asymptotic expansions of distributions and their moments, thus they are easy and reliable to use. Simultaneously with increasing number of observations, errors of these classical methods decrease.

8. LARGE SURVEYS

8.1. Specific statistical aspects

Let us discuss the implications of the general principles discussed previously for large surveys. Because of property (2) (Section 5) large surveys potentially enable the detection of low S/N ratio signals. However, this may be prevented by prefiltering data using "general variability criteria". These general criteria, usually based on the total variance, are not particularly sensitive for detection of periodic signals. This is so because by virtue of the Parceval Equation, the total variance is proportional to the sum of power over all $n_{\rm eff}$ independent frequencies. Consequently, an increase of the power at a given frequency by a large factor, $A^2/\sigma^2 \gg 1$, can produce insignificant increase of the total variance by a factor of $A^2/(\sigma^2 n_{\rm eff}) \ll 1$.

Thus, the general variability criterion leaves many significant oscillations undetected.

8.2. Efficiency

Scanning of the large number of low amplitude signals requires the use of efficient algorithms. For near sinusoidal signals sampled nearly uniformly, with no large gaps, the modified power spectrum has good sensitivity and its FFT implementation performs with $\mathcal{O}(n_{\rm FFT}\log_2 n_{\rm FFT})$ efficiency, where $n_{\rm FFT}$ is the number of interpolated points covering the whole observation time interval (Press & Rybicki 1989). However, such a combination of signals and sampling occurs in astronomy rarely. Methods, relying on phase folding and binning, have better sensitivity for non-sinusoidal signals and perform as $\mathcal{O}(\epsilon n_{\rm FFT} n_{\rm obs})$, where $\epsilon \approx 1$ for a good frequency sampling. For large gaps, $n_{\rm FFT}$ becomes very large and then the phase folding methods outperform the FFT power spectrum, particularly if only the fraction $\epsilon \ll 1$ of $n_{\rm FFT}$ frequencies is searched: $n_{\rm FFT}\log_2 n_{\rm FFT}) > \epsilon n_{\rm FFT} n_{\rm obs})$. The phase folding and binning methods are still less sensitive than the methods employing Fourier series (property (3) of Section 5). With introduction of the orthogonal projection algorithm it becomes feasible to apply the multi-harmonic Fourier series method for large data samples (Schwarzenberg-Czerny 1996). The orthogonal projection method performs as $\mathcal{O}(\epsilon n_{\rm FFT} n_{\rm obs} n_{\rm harm})$, where $n_{\rm harm}$ denotes the number of harmonics in use. Thus in circumstances discussed above, the multi-harmonic method may outperform the FFT based methods in terms of numerical efficiency. In any case, the multi-harmonic periodogram is more sensitive than the power spectrum for non-sinusoidal oscillations.

8.3. Practical experience

The OGLE collaboration performs a large imaging survey of selected fields, collecting in the process over 50 000 light curves of variable stars. A manual scanning of these light curves and corresponding periodograms is hardly feasible. The statistical detection criteria discussed above are well suited for automatic search of periodic stars in the large data base. The OGLE data were used to perform detailed tests of the AOV and PDM methods (Schwarzenberg-Czerny 1989, Stellingwerf 1978). The AOV method relies on phase folding

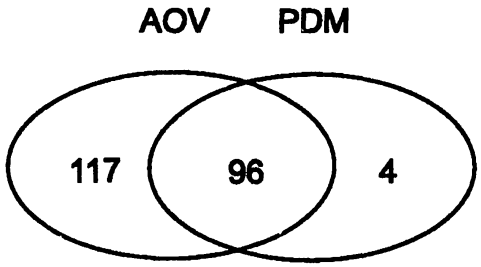

Fig. 8. Results of the OGLE variable star search in Baade's Window, using (i) AOV periodogram and (ii) uncorrected PDM periodogram. Numerals indicate the number of periodic variable stars found by each/both methods.

and binning of observations and on the AOV statistic Θ_0 (Table 1). PDM is an older method, popular among observers. It uses the same folding and binning scheme as AOV but differs in using of the $\Theta_\perp$ statistic and F-S distribution. According to Table 1, the statistic and the distribution do not match, so, in its original form, PDM is statistically incorrect.

The first test concerned the sensitivity of the AOV method. It was performed for stars laying in overlapping areas of frames, so that two independent light curves were available for each object. Analysis of these data revealed a nearly 100% detection efficiency for amplitudes exceeding 0.2 mag and stars brighter than 17 mag. The second test involved a comparison of performances of the AOV and PDM methods. Both the AOV and PDM methods were applied to variable stars found by OGLE in Baade's Window. In this field, 96 periodic variable stars were identified by both AOV and PDM. Only 4 other stars were discovered with PDM and missed by AOV. Additionally, 117 stars, i.e. most of periodic variables discovered in this field, were found by AOV and missed by PDM (Fig. 8). The periodic nature of all stars detected in this test was confirmed by eye inspection of the folded light curves. The conclusion drawn from both tests was that AOV gives the best results for different types of periodic variables (Udalski et al. 1994). Note, however, that PDM with the matching distribution derived according to Section 4 should perform equally well as AOV. Application of a better model, namely of Fourier series as in multi-harmonic periodogram, should improve the performance further.

ACKNOWLEDGMENTS. This work was supported by the Committee for Scientific Research (KBN) under grant No. 2P03D. 003.13.

REFERENCES

Abramovitz M., Stegun I. 1971, Handbook of Mathematical Functions, Dover, N.Y.

Akaike H. 1973, in Proceedings of the 2nd Intern. Symp. in Information Theory, eds. B. N. Tetroc & F. Caski, Akad. Kiado, Budapest, p. 267.

Akerlof C. et al. 1994, ApJ, 436, 787

Bickel P. J., Doksum K. A. 1977, Mathematical Statistics, Holden-Day, San Francisco

Buchler J. R. 1993, in Nonlinear Phenomena in Stellar Variability, eds. M. Takeuti & J.-R. Buchler, Ap&SS, 210, 9.

Davies S. R. 1990, MNRAS, 244, 93

Deeming T. J., 1975, Ap&SS, 36, 137

Eadie W. T., Drijard D., James F. E., Roos M., Sadoulet B. 1982, Statistical Methods in Experimental Physics, North-Holland Publ. House, Amsterdam.

Ferraz-Mello S. 1981, AJ, 86, 619

Feigelson E. D. 1997, in Proceedings Astronomical Time Series (Tel-Aviv), eds. D. Maoz et al., Kluwer Academic Publishers, Dordrecht, p. 13

Foster G. 1995, AJ, 109, 1889

Grenander U., Szegö G. 1958, Toeplitz Forms and Their Applications, Univ. of California Press, Berkeley, p. 40

Grison P. 1994, A&A, 289, 404

Horne J. H., Baliunas S. L. 1986, ApJ, 302, 757

Lehto H. J. 1997, in Proceedings Astronomical Time Series (Tel-Aviv), eds. Maoz D. et al., Kluwer Academic Publishers, Dordrecht, p. 269

Lomb N. R. 1976, Ap&SS, 39, 447

Mallat S., Zhang Z. 1993, IEEE Trans. Signal Process., 41(12), 3397

Press W. H., Rybicki G. B. 1989, ApJ, 338, 277

Press W. H, Teukolsky S. A., Vetterling W. T., Flannery B. P. 1992, Numerical Recipes, Cambridge Univ. Press

Roberts D. H., Lehar J., Dreher J. W. 1987, AJ, 93, 968

Roques S., Bourzeix F., Bouyoucef K. 1996, A&SS, 239, 297

Scargle J. H. 1982, ApJ, 263, 835

Scargle J. D. 1989, ApJ 359, 469

Scargle J. D. 1997, in Proceedings, Astronomical Time Series (Tel-Aviv), eds. Maoz D. et al., Kluwer Academic Publishers, Dordrecht, p. 1

Schwarzenberg-Czerny A. 1989, MNRAS, 241, 153
Schwarzenberg-Czerny A. 1991, MNRAS, 253, 198
Schwarzenberg-Czerny A. 1996, ApJ, 460, L107
Schwarzenberg-Czerny A. 1997a, ApJ, 489, 941
Schwarzenberg-Czerny A. 1997b, in preparation
Stellingwerf R. F. 1978, ApJ, 224, 953
Udalski A. et al. 1994, Acta Astron., 44, 317
Walker A. R. 1994, AJ, 108, 555

APPENDIX A: ORTHOGONAL BASES

Equation (1) suggests that it would be convenient to do time series analysis in a mathematical environment where functions behave as vectors and form scalar products. Note that since it is often convenient to use a complex exponential function instead of sine and cosine, complex values should be allowed. The relevant mathematical space is called *Hilbert space*. Hilbert spaces encountered in quantum mechanics, $\mathcal{H}_\infty$, have infinite number of dimensions and, correspondingly, complex theory. As long as we consider only discrete time series of, say, n complex observations $x_k, k = 1, ..., n$, we should concern ourselves only with a finite dimension Hilbert space $\mathcal{H}_n$. For studying properties of $\mathcal{H}_n$, often suffices an analogy with ordinary geometry and no arcane knowledge of operator and spectral theory is required. $\mathcal{H}_n$ differs from a familiar n-dimensional real vector space $\mathcal{R}_n$ only slightly because of a complex scalar product:

$$(x, y) = \overline{(y, x)} \tag{13}$$

$$(\alpha x + y, z) = \overline{\alpha}(x, z) + (y, z) \tag{14}$$

$$\|x\|^2 \equiv (x, x) \geq 0 \quad \text{and} \tag{15}$$

$$\|x\| = 0 \Leftrightarrow x = 0 \tag{16}$$

where $\overline{x}$ denotes the complex conjugate of x. A scalar product definition suitable for the present purposes is

$$(x, y) = \sum_{k=1}^{n} g_k \overline{x_k} y_k, \tag{17}$$

where g_k are real and positive, to satisfy Eq. (15). The presence of complex conjugate terms in Eqs. (13)–(17) is necessary to satisfy Eq. (15) for the pure imaginary x. The vector norm $\|x\|$ and the angle

between vectors $\gamma(x,y)$ are defined in a usual way: $\|x\|^2 = (x,x)$ and $\cos\gamma = (x,y)/\|x\|\|y\|$.

Similarly to ordinary vector space, $\mathcal{H}_n$ must contain a set of n independent base vectors $z^{(j)}, j = 1,...,n$, such that $|\cos\gamma(z^{(j)}, z^{(k)})| \neq 1$ for $j \neq k$. From these vectors one may always construct the orthonormal vector base $x^{(j)}, j = 1,...,n$, such that

$$(x^{(j)}, x^{(k)}) = \delta_{jk}, \tag{18}$$

where $\delta_{jk} = 1$ for $j = k$ and 0 otherwise. The construction, called Gramm-Schmidt orthonormalization, is a recurrence process:

$$x^{(1)} \Leftarrow z^{(1)}/\|z^{(1)}\|, \tag{19}$$

$$x^{(j)} \Leftarrow z^{(j)} - \sum_{k=1}^{j-1} x^{(k)}(x^{(k)}, z^{(j)}), \tag{20}$$

$$x^{(j)} \Leftarrow x^{(j)}/\|x^{(j)}\| \quad \text{for} \quad j = 2,...,n. \tag{21}$$

Eq. (20) times $x^{(l)}, \ell = 1, j$ demonstrates that $x^{(j)}$ satisfies Eq. (18) if $x^{(k)}, k = 1,...,j-1$ do. Some reasons which make orthogonal (orthonormal) bases convenient were discussed already in Section 2. Arbitrary vectors, say, y, z and their scalar product and norm have a simple expansion in an orthonormal base $x^{(k)}$:

$$y = \sum_{k=1}^{n} x^{(k)}(x^{(k)}, y), \tag{22}$$

$$(y,z) = \sum_{k=1}^{n} (y, x^{(k)})(x^{(k)}, z), \tag{23}$$

$$\|y\|^2 = \sum_{k=1}^{n} |(y, x^{(k)})|^2. \tag{24}$$

The validity of Eq. (22) is best demonstrated by observation that its scalar product with $x^{(k)}$ reduces using Eq. (18) to an identity $(x^{(k)}, y) = (x^{(k)}, y)$ for all n values of k. Since there are n components of y to satisfy these n identities, the expansion in Eq. (22) must be exact and unique. Eq. (23) is obtained by substitution of Eq. (22)

for y and z in (y, z) and subsequent use of Eq. (18). Eq. (24) is obtained from Eq. (23), as a special case for (x, x), using Eqs. (15) and (13). Equations (22)–(24) are familiar from real vector spaces. They retain validity for the complex $\mathcal{H}_n$. These complex equations have an interesting interpretation for even sampled observations and the FFT model. For the FFT frequency grid, harmonics $\exp(\omega t)$ are orthonormal with suitable normalization. Then for real y, the FFT transform is $\mathcal{F}y(k) = (y, x^{(k)})$ and Eq. (22) reduces to a familiar identity $y = \mathcal{F}^{-1}(\mathcal{F}y)$. The equation $\|\mathcal{F}y\|^2 = \|y\|^2$ corresponds to Eq. (24). The convolution theorem $y * z = \mathcal{F}y\mathcal{F}z$ does not generalize for arbitrary orthonormal bases since generally $x^{(k+1)} \neq \alpha x^{(k)}$ for α independent of k. However, Eq. (23) demonstrates that for a special case of zero lag, $(y * z)(0)$, the convolution theorem holds for general orthonormal functions.

APPENDIX B: ORTHOGONAL LEAST SQUARES FIT

Let us consider for the moment a model $x_\parallel$ built of a linear combination of orthogonal base functions, $z^{(k)}$:

$$x_\parallel = \sum_{k=1}^{r} y_r z^{(k)}, \tag{25}$$

where y_k are model parameters. We restrict ourselves to observations and models represented by vectors of the real numbers, x and $x_\parallel$, respectively. Let us fit the parameters y by least squares. Then $0 = (\partial/\partial y_k)\|x_\perp\|^2 = (\partial/\partial y_k)[(\underline{x - x_\parallel}, x - x_\parallel) + (x - x_\parallel, \underline{x - x_\parallel})]$, where differentiation acts only on underlined terms. Noting that x does not depend on y and that swapping of real arguments does not change the second product, one obtains: $0 = -2(\partial x_\parallel/\partial y_k, x - x_\parallel) = -2(z^{(k)}, x - \sum y_k z^{(k)}) = -2[(z^{(k)}, x) - y_k(z^{(k)}, z^{(k)})]$. We exploit here orthonormality of the base vectors (Eq. 18). In this way one derives the final result:

$$y_k = (z^{(k)}, x). \tag{26}$$

In this way we demonstrate that for the orthogonal model a (unique) least squares solution for model parameters is obtained by an orthogonal projection.