New Horizons in Time-Domain Astronomy
Proceedings IAU Symposium No. 285, 2011
R.E.M. Griffin, R.J. Hanisch & R. Seaman, eds.

doi:10.1017/S1743921312000294

On the Sensitivity of Period Searches

A. Schwarzenberg-Czerny[1,2]

[1]Nicolaus Copernicus Astronomical Centre, 00-716 Warsaw
[2]Adam Mickiewicz University Observatory, PL 60-286, Poznań, Poland
email: alex@camk.edu.pl

Invited Talk

Abstract. Astronomical time series are special in that time sampling in them is uneven yet often with periodic gaps due to daytime, moon and seasons. There is therefore a need for special-purpose time-series analysis (TSA) methods. The emergence of massive CCD photometric surveys from the ground and space raises the question of an automatic period search in $\gg 10^5$ light curves. We caution that already at the planning stage it is important to account for the effects of time sampling and analysis methods on the sensitivity of detections. We present a transparent scheme for the classification of period-search methods. We employ tools for evaluating the performance of those methods, according to the type of light curves investigated. In particular we consider sinusoidal and non-sinusoidal oscillations as well as eclipse or transit light curves. From these considerations we draw recommendations for the *optimum analysis of astronomical time series*. We present briefly the capability of an automatic period-search package TATRY. Finally we discuss the role of Monte Carlo simulations in the analysis of detection sensitivity. As an example, we demonstrate a practical method to account for the bandwidth (multi-trial) penalty in the statistical evaluation of detected periods.

Keywords. methods: data analysis, methods: statistical, (stars:) binaries: eclipsing, stars: oscillations (including pulsations), (stars:) planetary systems, (Galaxy:) globular clusters: general, (galaxies:) Magellanic Clouds

1. Introduction

We present a biased overview of methods for enhancing period searches in astronomical data, relying heavily on our own work. In this context, astronomical data are ones with uneven time sampling. Data having regular time sampling are best analysed using FFT-based methods and are not discussed here. The massive present-day photometric surveys yield $\gg 10^5$ light curves, so we concentrate on methods that permit a fully automatic period analysis. First we turn attention to the proper planning of observation sampling. We then discuss the orthogonal models of periodic signals as they permit analytical evaluation of statistical properties in period search. From statistics we employ the concept of test power in order to evaluate the sensitivity of period-search methods. On that basis we are able to derive recommendations for optimum period analysis. We continue with the discussion of two correction effects to be accounted for in calculating realistic probability distributions. Finally we turn our attention to pitfalls and profits of Monte Carlo simulations for statistical analysis of time series.

2. Sample Planning

A frequent fatal error in planning astronomical observations is observing objects at the same position with respect to the meridian, say on D consecutive nights. The corresponding sampling pattern may then be represented as a product $\Pi_D(t) \cdot \text{Ш}_1(t)$, where Π and

III denote the Heaviside top-hat and Dirac picket-fence functions in the time domain. Applying the convolution $*$ and its theorem to the Fourier transforms $\mathcal{F}$, one obtains the corresponding window function in the frequency domain, $W = |\mathcal{F}\Pi_D * \mathcal{F}\mathrm{III}_1|^2 = |\mathrm{sinc}(D\nu) * \mathrm{III}_1(\nu)|^2$. It corresponds to an infinite series of peaks/aliasses of width $1/D$. Next, let us consider the same number of observations scattered uniformly over a fraction d of nights. Such a pattern may be represented by $\Pi_D(t) \cdot [\mathrm{III}_1(t) * \Pi_d(t)]$ and $W = |\mathrm{sinc}(D\nu) * [\mathrm{III}_1(\nu) \cdot \mathrm{sinc}(d\nu)]|^2$. Now all aliases beyond width $1/d$ of the $\mathrm{sinc}(d\nu)$ function are greatly reduced in size. Such a reduction is obtained by reshuffling observations on different nights without changing their time span *per night*. The effect resembles the aperture synthesis obtained by shifting radio antennæ between observations.

3. Period Search by Quadratic Norms

3.1. *Statistical Principles of Detection: Periodogram Statistics and Distributions*

Our considerations are based on R.A. Fisher's statistical theory of the least squares (LSQ) fit of n observations $\mathbf{x} = (x_1, \cdots, x_n)$ with the orthogonal model $\mathbf{x}_\parallel = \sum_i c_i \mathbf{p}_i(\varphi)$, where vectors/functions $\mathbf{p}_i(\varphi) \equiv \mathbf{p}_i(2\pi\mathbf{t})$ are orthogonal with respect to the scalar product defined by observed phases: $0 = (\mathbf{p}_i, \mathbf{p}_j) \equiv \sum_\varphi w_\phi p_i(\varphi) p_j(\varphi)$ for $i \neq j$. Hereafter we assume that the average values $(\mathbf{1}, \mathbf{x}) = (\mathbf{1}, \mathbf{x}_\parallel) = 0$. By virtue of the Fisher lemma the model $\mathbf{x}_\parallel$ and residuals from fit $\mathbf{x}_\perp \equiv \mathbf{x} - \mathbf{x}_\parallel$ are orthogonal $(\mathbf{x}_\parallel, \mathbf{x}_\perp) = 0$. In consequence, an n-dimensional analogue of Pythagoras theorem holds:

$$\begin{array}{llll} \|\mathbf{x}\|^2 = & \|\mathbf{x}_\parallel\|^2 + & \|\mathbf{x}_\perp\|^2 & \text{where} \\ n = & n_\parallel + & n_\perp & \\ \text{observation} = & \text{model} + & \text{residuals} & \end{array} \tag{3.1}$$

$\|\mathbf{x}\|^2 \equiv (\mathbf{x}, \mathbf{x})$ and n, $n_\parallel$ and $n_\perp$ denote the number of observations, the number of model parameters and the number of degrees of freedom of the residuals. Because of the relation $\|\mathbf{x} - \mathbf{x}_\parallel\|^2 = \|\mathbf{x}\|^2 - 2(\mathbf{x}, \mathbf{x}_\parallel) + \|\mathbf{x}_\parallel\|^2$, where only the middle term depends on the frequency ν, our considerations for LSQ also apply to the case of the cross-correlation function (CCF) periodogram.

Suitable families of orthogonal functions $\mathbf{p}_i$ are either Szegö trigonometric polynomials or top-hat functions corresponding to the phase bins (Schwarzenberg-Czerny 1996, 1989). Nominally, Szegö polynomials follow from Gramm-Schmidt orthonormalization of Fourier harmonics, yet convenient recurrence formulæ also exist. Phase folding and binning of data is equivalent to LSQ fitting a step function composed of a linear combination of top-hats. The box function employed for planetary transit searches corresponds to two phase bins of unequal width (Schwarzenberg-Czerny & Beaulieu 2006), so the present considerations apply in this area too. Quite unique orthogonal functions were employed by MACHO (Akerlof *et al.* 1994).

A statistics $\Theta(\nu, \mathbf{x})$ is the merit figure indicating the quality of the fit. A periodogram is the plot of $\Theta(\nu, \mathbf{x})$ against ν. Patterns in the periodogram may relate to the presence in the data of oscillations with the corresponding frequency. The significance of those frequencies depends on the probability distribution of Θ for hypothetical data consisting of pure noise. In Statistics, that case is called a null hypothesis, H_0. Θ must be dimensionless, as no statistical conclusions may depend on units. There are three ways to construct dimensionless Θ statistics from the dimensioned $\|\mathbf{x}\|$, $\|\mathbf{x}_\parallel\|$ and $\|\mathbf{x}_\perp\|$ (Table 1). Because of Eq. (3.1), all these Θ's are uniquely related:

$$\Theta_\perp = 1 - \Theta_\parallel = \frac{1}{1 + \Theta_{AOV}}, \tag{3.2}$$

https://doi.org/10.1017/S1743921312000294 Published online by Cambridge University Press

Table 1. Basic Classes of Period Statistics

Statistics	Definition	Distribution	Name	Analogue
$\Theta_{AOV} =$	$\frac{\|\mathbf{x}_\parallel\|}{\|\mathbf{x}_\perp\|}$	$F(n_\parallel, n\perp; \Theta_{AOV})$	Fisher-Snedecor	AOV$^{(1)}$,mhAOV$^{(2)}$
$\Theta_\parallel =$	$\frac{\|\mathbf{x}_\parallel\|}{\|\mathbf{x}\|}$	$\beta(n_\parallel, n\perp; \Theta_\parallel)$	β distribution	Power$^{(3)}$, L-S$^{(4,5)}$
$\Theta_\perp =$	$\frac{\|\mathbf{x}_\perp\|}{\|\mathbf{x}\|}$	$\beta(n_\perp, n\|; \Theta_\perp)$	β distribution	χ^2,PDM$^{*\ (6,7)}$

References: (1)- Schwarzenberg-Czerny (1989), (2)- Schwarzenberg-Czerny (1996), (3)- Deeming (1975), (4)- Lomb (1976), (5)- Scargle (1982), (6)- Stellingwerf (1978), (7)- Schwarzenberg-Czerny (1997)

so the corresponding F and β distributions may be obtained from each other by suitable changes of variable. From this we find that conclusions drawn from the Θ_{AOV}, $\Theta_\parallel$ and $\Theta_\perp$ periodograms must all be identical *if and only if* the model $\mathbf{x}_\parallel$ remains the same. In other words, what counts is not the shape of the periodogram peak but its probability (Schwarzenberg-Czerny 1998). Turning that argument *ad absurdum*, one may state that obtaining a clean, single-peak periodogram is sufficient to raise any periodogram to the power of 1000 or so. As no additional information is supplied, such a nice view has spurious meaning. In practical terms it is sufficient to discuss periodograms in which oscillations correspond to peaks. The results would also apply to the periodograms showing through at the corresponding frequencies.

However, to the human eye equivalent periodograms may look deceptively different. For a high S/N and χ^2 periodogram, an alias minimum of $\Theta_\perp$ that is twice as high as the true minimum would not look significant. At the same time the corresponding alias peak power $\Theta_\parallel = 1 - \Theta_\perp$ would almost match the true peak, pretending to be significant. For the human eye it is therefore better to plot a $\log \Theta_{AOV}$ periodogram, as the probability distribution of its values is close to the normal one. Then a twice-higher peak has twice the σ significance.

3.2. *Sensitivity of Detection: Test Power*

To evaluate the sensitivity of detection we must consider two different hypothetical data sets: for a pure noise with standard deviation 1, and for noise plus a periodic signal of amplitude A (same units). In Statistics these two cases are called the "null" and the "alternative" hypotheses, H_0 and H_1, respectively. Accordingly, for H_0 and H_1, Θ obeys different probability distributions, $P_0(\Theta)$ and $P_1(\Theta)$. Ideally the two distributions are separated by a critical value, Θ_c. We could say that $\Theta < \Theta_c$ corresponds to a pure noise and $\Theta > \Theta_c$ to the detected of a signal. However, in reality the two distributions overlap for a range of Θ. Two kinds of errors thus arise: one claims detection while in reality H_0 is true (*false positives*), and conversely one claims no detection while in reality H_1 remains true (*misses*). In classical statistics we fix Θ_c so that false positives seldom occur, i.e. the significance level $\alpha = P_0(\Theta < \Theta_c)$ is close to 1. Then the *test power* of the criterion Θ_c is defined as $\beta = P_1(\Theta < \Theta_c)$, where the probability of misses is $1 - \beta$. Thus, for a fixed Θ_c, large β corresponds to good detection sensitivity. The analytical formulæ for P_0 are listed in Table 1. No corresponding formulæ are known for P_1, as they depend in a complex way on signal shape. However, for small signal-to-noise, $A/1 \ll 1$, it is possible to derive approximate asymptotic formulæ for P_1 (Schwarzenberg-Czerny 1999). In that approximation, P_0 and P_1 retain the same shape yet are shifted, in units of their standard deviation, by

$$\Delta\Theta / \sqrt{Var\{\Theta\}} = A^2 n \frac{\|s_\parallel\|^2}{\sqrt{2n_\parallel}} \quad \text{where} \tag{3.3}$$

https://doi.org/10.1017/S1743921312000294 Published online by Cambridge University Press

$$\|s_{\|}\|^2 = \frac{(x_{\|signal}, x_{\|model})^2}{\|x_{\|signal}\|^2 \|x_{\|model}\|^2}. \tag{3.4}$$

$x_{\|signal}$ and $x_{\|model}$ denote the shapes of the real signal and of the fitted model, respectively. The bigger $\Delta\Gamma$ is, the more sensitive is our method/model for a given signal.

After feeding into Eqs. (3.3, 3.4) the von Mieses function $e^{-\kappa \cos^2 \varphi}$ as a signal and Fourier harmonics or top-hat functions as model functions, our calculations yield the test power for the most popular Fourier and binning periodograms. Small κ values corespond to near sinusoidal input signals, and large ones to narrow gaussian spikes of width $\sqrt{2\kappa}$. These calculations reveal that excessively crude or excessively fine models both yield decreased sensitivity because (respectively) of the factors $\|s_{\|}\|^2$ and $1/\sqrt{n_{\|}}$ in Eq. (3.3). Thus, for optimum sensitivity, the resolution of the model should just resolve the features in the signal.

4. Related Issues

4.1. *Strength and Pitfalls of Monte Carlo TSA*

Monte Carlo (MC) simulations have venerable origins, and date back to von Neumann's work in 1940 at Los Alamos. MC constitutes a powerful method for studying likely events and their expectation integrals, $E\{F(x)\} \equiv \int F(x) f(x) dx$. The application of MC methods follows rules of Statistics as a branch of Mathematics. Current prevailing editorial policies seem to expect any observer or referee to be able to analyse the reliability of conclusions by MC simulations. In my opinion it is as justifiable as expecting an observer to perform state-of-the-art hydrodynamic simulations and/or to support observations with quantum mechanical calculations of involved atoms and transition probabilities. It is doable, but by no means by all. The policy results in the emergence of many poor simulations at best, and in the publication (in otherwise respectable journals) of chains of logically linked wrong papers based on shabby simulations, at worst. In particular, the application of MC methods for rare events, such as in significance analysis, is always

https://doi.org/10.1017/S1743921312000294 Published online by Cambridge University Press

uneconomic and often risky, because of untested statistical properties of random number generators and discrete computer arithmetic for these rare events.

4.2. *Corrected Significance: Bandwidth Penalty*

From Table 1 one may derive the analytic tail probability of large Θ for a single frequency: $Q_1(\Theta > \Theta_0) \equiv 1 - P_1(\Theta < \Theta_0)$. As more and more frequencies are examined in the periodogram, the probability of a spurious occurrence of a peak due to pure noise increases, in the same way as the probability of winning a lottery increases with the number of trials. This increased probability, called *bandwidth penalty*, has to be accounted for any realistic statistical evaluation. Because the aliasing values of a periodogram at different frequencies may be strongly correlated, out of N investigated frequencies only $N_{\mathit{eff}} \leqslant N$ may be independent. In that case the postulated tail probability, according to Horne & Baliunas 1986, could be:

$$Q_N(\Theta_0) = Q_1(\Theta_0)^{N_{\mathit{eff}}}. \tag{4.1}$$

The hitch is in the unknown value of N_{eff}. Paltani (2004) proposed a useful method to estimate N_{eff} by MC simulations, though relying on their mean, rather than extreme, values. The Paltani procedure, improved by us in steps (e) and (f), may be summarized in the following way:

(*a*) Replace observations x with a simulated white noise;
(*b*) Calculate the periodogram Θ for a given frequency grid;
(*c*) Find the extreme value Θ_s of the simulated periodogram;
(*d*) Repeat steps (a)–(c) as many times as desired for accuracy;
(*e*) Find median value Θ_m of Θ_s, where $Q_N(\Theta_m) = 0.5$;
(*f*) Solve Eq. (4.1) for $N_{\mathit{eff}} = \frac{\ln 0.5}{\ln[Q_1(\Theta_m)]}$;
(*g*) Calculate $Q_N(\Theta_0)$ from Eq. (4.1) for Θ_0.

In this way one makes use of likely events $Q_N(\Theta_m) \gg Q_N(\Theta_0)$, which would be rejected in the brute-force simulations.

4.3. *MC Study of Bandwidth Correction*

The modified Paltani method enables a convenient study of N_{eff} by MC simulations for several realistic, uneven sampling patterns. For illustration, we included a case with a large time gap in the middle of the observations. The simulations depend on several parameters: the number of calculated frequencies N, the number of observations N_{obs}, the maximum number of resolved frequencies $N_{max} = \Delta t \Delta \nu$ where Δt and $\Delta \nu$ denote the ranges of time and frequency spanned by the observations and their periodogram, respectively, and the number of model parameters $N_{\|}$.

In all simulations the condition $N_{\mathit{eff}} < N$ held strictly. However, another condition, $N_{\mathit{eff}} \leqslant N_{max}\sqrt{N_{\|} - 1}$, held only approximately. It is expected that a finer model would yield a more precise phase determination, so some factor involving $N_{\|}$ seems in place. Surprisingly, the condition $N_{\mathit{eff}} \leqslant N_{obs}$ does not hold in our simulations. To our knowledge, that effect was not mentioned in the literature. At this stage no definite explanation seems available. However, one could suspect that periodogram values depend on observations in such a complicated way that effectively they become chaotic. In the same sense, the consecutive values of a random number generator show no correlation despite them all depending on just one seed value. An alternative explanation would be that Eq. (4.1) never held, i.e. that in a periodogram there were no truly independent frequencies.

https://doi.org/10.1017/S1743921312000294 Published online by Cambridge University Press

4.4. *Corrected Significance: Correlation or Red Noise Effect*

The presence of a *correlation (red noise)* in observations may ruin simplistic statistical estimates. For example, the LSQ fit of a sine to the solar spot Wolfer numbers spanning 100 years yields the nominal period of $P \approx 11$ y with an error of the order of $0.002P$. However, the propagation of such an ephemeris for the next 50 years demonstrates that a realistic period error was $\approx 0.1P$. This has happened because consecutive residuals from the fit are correlated (they keep the same sign for decades), while the standard LSQ error estimates implicitly *assume* that residuals are *(uncorrelated) white noise*. Conversely, for simulated data consisting of white noise plus the oscillation of the same variances or amplitudes as above, the $0.002P$ error estimate proves realistic.

This remarkable correlation effect is seldom discussed in texts on LSQ. In fact, the correlation of every N_{corr} consecutive observations decreases the effective number of observations by roughly a factor of N_{corr}, and hence increases the real LSQ errors by a factor of $\sqrt{N_{corr}}$(Schwarzenberg-Czerny 1991). A simple way to estimate N_{corr} is by counting the number of sign changes in the residuals from the fit (the *post mortem* analysis). For white noise, one expects $N_{obs}/2$ changes of sign (every second residual should change sign, on average). If the observed number of sign changes in the residuals is $N_{sign} < N_{obs}/2$, then the number of consecutive correlated observations is $N_{corr} \approx N_{obs}/(2N_{sign})$.

Conclusion: Statistics does work for planning and analysis of astronomical time series observations, though care is needed.

References

Akerlof, C., Alcock, C., Allsman, R., *et al.* 1994, *AJ*, 436, 787
Deeming, T. J. 1975, *Astrophys. Sp. Sc.*, 36, 137
Gieronimus, Ya.L., 1958, *Orthogonal polynomials of circle and real intervals*, Moscow, GIFML (in Russian)
Horne, J. H. & Baliunas, S. L. 1986, *ApJ*, 302, 757
Lomb, N. R. 1976, *Atrophys. Sp. Sc.*, 39, 447
Paltani, S. 2004, *A&A*, 420, 789
Scargle, J. D. 1982, *ApJ*, 263, 835
Schwarzenberg-Czerny, A. 1989, *MNRAS*, 241, 153
Schwarzenberg-Czerny, A. 1991, *MNRAS*, 253, 198
Schwarzenberg-Czerny, A. 1996, *ApJ*, 460, L107
Schwarzenberg-Czerny, A. 1997, *ApJ*, 489, 941
Schwarzenberg-Czerny, A. 1998, Baltic Astronomy, 7, 43
Schwarzenberg-Czerny, A. 1999, *ApJ*, 516, 315
Schwarzenberg-Czerny, A. & Beaulieu, J.-P. 2006, *MNRAS*, 365, 165
Stellingwerf, R. F. 1978, *ApJ*, 224, 953

Appendix

The following recurrence yields Szegö polynomial expansion coefficients $c_n, n = 0, 1, \cdots, 2N$ for observations $t_m, x_m, m = 1, \cdots, M$ (Gieronimus 1958, Schwarzenberg-Czerny 1996):

$$\mathbf{p}_0(\mathbf{z}) = 1 \qquad c_n = \frac{(\mathbf{f}, \mathbf{p}_n)}{\|\mathbf{p}_n\|^2} \qquad \alpha_n = \frac{(\mathbf{z}\mathbf{p}_n, 1)}{\|\mathbf{p}_n\|^2} \tag{1}$$

$$\mathbf{p}_{n+1}(\mathbf{z}) = \mathbf{z}\mathbf{p}_n(\mathbf{z}) - \alpha_n \mathbf{z}^n \overline{\mathbf{p}_n(\mathbf{z})} \tag{2}$$

where $z_m = e^{2\pi i \omega t_m}$, $\mathbf{f} = \mathbf{z}^N \mathbf{x}$, i.e. component-by-component product of vectors $\mathbf{z}^N$ and $\mathbf{x}$. The modified Eq. (1) for α_n remains valid for the FFT limit case $\mathbf{p}_n(\mathbf{z}) \to \mathbf{z}^n$.

https://doi.org/10.1017/S1743921312000294 Published online by Cambridge University Press