Detecting non-sinusoidal periodicities in observational data using multiharmonic periodograms

Roman V. Baluev

 Sobolev Astronomical Institute, St Petersburg State University, Universitetskij prospekt 28, Petrodvorets, St Petersburg 198504, Russia

Accepted 2009 February 12. Received 2009 February 9; in original form 2008 November 5

ABSTRACT

We address the problem of assessing the statistical significance of candidate periodicities found using the so-called ‘multiharmonic’ periodogram, which is being used for detection of non-sinusoidal signals and is based on the least-squares fitting of truncated Fourier series. The recent author’s investigation made for the Lomb–Scargle periodogram is extended to the more general multiharmonic periodogram. As a result, closed and efficient analytic approximations to the false alarm probability, associated with multiharmonic periodogram peaks, are obtained. The resulting analytic approximations are tested under various conditions using Monte Carlo simulations. The simulations showed a nice precision and robustness of these approximations.

Key words: methods: data analysis – methods: statistical – surveys.

1 INTRODUCTION

The Lomb (1976)–Scargle (1982) (LS) periodogram is a well-known powerful tool, which is widely used to search for periodicities in observational data. The main idea used in the LS periodogram is to perform a least-squares fit of the data with a sinusoidal model of the signal and then to check how much the resulting weighted rms have decreased for a given signal frequency. The maximum value of the LS periodogram (i.e. the maximum decrement in the least-squares goodness-of-fit measure) corresponds to the most likely frequency of the periodic signal. This natural idea is quite easy to implement in numerical calculations.

However, random errors in the input data inspire noise peaks on the periodogram, so that we can never be completely sure that the peak that we actually observed was produced by a real periodicity. The common way to assess the statistical significance of the observed peak is based on the associated ‘false alarm probability’ (FAP). The FAP is the probability that the observed or larger periodogram peak could be produced by random measurement errors. The smaller is FAP, the larger is the statistical significance.

Given some tolerance value FAP, (say 1 per cent), we could claim that the detected candidate periodicity is statistically significant (if FAP < FAP) or is not (if FAP > FAP).

From the statistical viewpoint, the FAP is tightly connected with the probability distribution of periodogram maxima, which are calculated within some a priori fixed frequency segment. However, even approximate calculation of this distribution is a non-trivial task. It represented a trouble for astronomers for about three decades. It is worthwhile to mention here, for instance, the papers by Horne & Baliunas (1986), Koen (1990), Schwarzenberg-Czerny (1998a,b), Cumming, Marcy & Butler (1999), Cumming (2004) and Frescura, Engelbrecht & Frank (2008). Recently, a significant progress in this field was attained in the paper (Baluev 2008), where closed and simultaneously rather efficient approximations of the FAP for the LS periodogram are given, basing on results in the theory of extreme values of stochastic processes.

However, periodic signals being dealt within astronomy often are significantly non-sinusoidal. Then, the use of the LS periodogram is not optimal, since the corresponding periodic variation would be fitted inadequately. For instance, it is the case for light curves of variable stars of several types and for radial velocity curves of stars orbited by a planet on an eccentric orbit. Several ways to deal with this issue were proposed (for further references, see e.g. Schwarzenberg-Czerny 1998a,b). In this paper, we focus attention on the so-called multiharmonic periodogram (Schwarzenberg-Czerny 1996), which is based on the least-squares fitting of truncated Fourier series. Note that in the paper (Baluev 2008), a general class of periodograms based on the least-squares data fitting was considered as well, but from theoretical positions only. Here, our aim is to apply these general results to the multiharmonic periodograms.

The plan of this paper is as follows. In Section 2, we formulate the problem rigorously and introduce the necessary mathematical definitions. In Section 3, basing on the work of Baluev (2008), we derive closed approximations of the FAP, associated with multiharmonic periodogram peaks. In Section 4, we use numerical Monte Carlo simulations to quantify the accuracy of these analytic approximations.

E-mail: roman@astro.spbu.ru

1 Speaking more precisely, the periodogram maxima are always calculated over some discrete set of values. However, in practice the periodograms are usually plotted on a dense frequency grid, which is practically equivalent to a continuous segment. It is the case that we consider in this paper.
2 GENERAL DEFINITIONS

Let us write down the temporal model of the putative periodic signal using a trigonometric polynomial of some a priori stated degree \( n \):

\[
\mu(t, \theta, f) = \sum_{k=1}^{n} (a_k \cos 2\pi kf t + b_k \sin 2\pi kf t),
\]

where \( f \) is the signal frequency and the vector \( \theta \) incorporates \( d = 2n \) Fourier coefficients \( a_k, b_k \). Further, we adopt exactly the same notations as those used in Balnev (2008). Clearly, the model \( \mu \) is linear: \( \mu(t, \theta, f) = \theta \cdot \varphi(t, f) \), where the vector \( \varphi(t, f) \) incorporates the first \( n \) harmonics of the Fourier basis. In addition to the signal model \( \mu \), we define the base temporal model \( \mu_{n, 1}(t, \theta_{n, 1}) = \theta_{n, 1} \cdot \varphi_{n, 1}(t) \), which is assumed to be linear with respect to \( \theta_{n, 1} \). This base model may represent, for instance, a constant or a long-term polynomial (e.g. linear or quadratic) temporal trend. Therefore, the alternative (full) model is given by \( \mu_{n, 1}(t, \theta_{n, 1}) = \mu(t, \theta, f) \), where \( \theta_{n, 1} \) incorporates all parameters in \( \theta_{n, 1} \) and \( \theta \). From the viewpoint of the statistical tests theory, we need to test the base hypothesis \( H_{\theta} : \theta = 0 \) against the alternative one \( \theta : \theta \neq 0 \).

The input data set consists of \( N \) measurements \( x_i \), taken at timings \( t_i \) and having uncertainties \( \sigma_i \). We assume that the random errors of the measurements are statistically independent and normally distributed. Below we will deal with the least-squares periodograms defined in Balnev (2008). These periodograms are based on the linear least-squares fitting procedure. The basic one, \( z(f) \), represents the half difference

\[
z(f) = \left[ \chi^2_{n, 1}(f)^2 - \chi^2_{n}(f)^2 \right] / 2,
\]

where \( \chi^2_{n, 1} \) and \( \chi^2_{n} \) represent the minimum values of the \( \chi^2 \) goodness-of-fit statistic calculated under the two corresponding hypotheses, \( H_{\theta} \) and \( K \). Note that under the base hypothesis \( H_{\theta} \) the random quantities \( \chi^2_{n, 1} \) and \( \chi^2_{n} \) follow the \( \chi^2 \) distributions with \( N_{\theta} = N - d_{\theta} \) and \( N_{\theta} = N - d_{\theta} \) degrees of freedom and thus indeed represent \( \chi^2 \) variates. The periodogram \( z(f) \) can only be calculated if the variances \( \sigma_i \) of the observational errors are known exactly. Usually we do not know these variances exactly, and can fix only the statistical weights \( w_i \propto 1/\sigma_i^2 \), so that \( \sigma_i^2 = \kappa / w_i \) with the common factor \( \kappa \) being unconstrained a priori. Therefore, we will also consider three modified least-squares periodograms:

\[
\begin{align*}
z_{\kappa}(f) &= N_{\theta} \chi^2_{n, 1}(f) - \chi^2_{n}(f) \over 2 \chi^2_{n}(f), \\
z(f) &= N_{\theta} \chi^2_{n, 1}(f) - \chi^2_{n}(f) \over 2 \chi^2_{n}(f), \\
z_{\kappa}(f) &= N_{\theta} \chi^2_{n, 1}(f) - \chi^2_{n}(f) \over 2 \chi^2_{n}(f).
\end{align*}
\]

These periodograms do not depend on \( \kappa \) and can be calculated even if \( \kappa \) is unknown. The periodograms, \( z(t, f) \) and \( z_{\kappa}(f) \), represent normalizations of the basic periodogram \( z(f) \) by the sample variances of the residuals, calculated under one of the two hypotheses, \( H_{\theta} \) or \( K \). The periodogram \( z(t, f) \) is proportional to the logarithm of the likelihood ratio statistic. More discussion of these definitions can be found in Balnev (2008). A discussion of several issues associated with the least-squares interpretation of the periodograms introduced above can also be found in Schwarzenberg-Czerny (1998a,b) and Zechmeister & Kürster (2009). The modified periodograms \( z_{1,2,3} \) are unique-value monotonic functions of each other and thus are entirely equivalent for the practical use.

The definitions (3) imply that the statistical weights \( w_i \) should be known with sufficient precision, and only the proportionality factor is unknown. This framework is usually adopted for the period analysis of astronomical data (e.g. Gilliland & Baliunas 1987; Irwin et al. 1989; Zechmeister & Kürster 2009), and we adopt it here. Nevertheless, sometimes this model may not work well. For instance, the paper (Balnev 2009) discusses the case in which the weights of observations are not known a priori with sufficient precision. In this case, the traditional multiharmonic periodograms being discussed here may not work well.

We do not discuss in detail the numerical algorithms for calculation of the periodograms introduced. The form of the above definitions is more suitable for quantifying the statistical distributions of the corresponding periodograms. Fast numerical algorithms of practical evaluation of the multiharmonic periodograms are given in Schwarzenberg-Czerny (1996) and Palmer (2009).

3 FALSE ALARM PROBABILITY

Let us pick any of the periodograms introduced above, and denote it as \( Z(f) \). If the frequency of the putative signal was known, the false alarm probability \( \text{FAP}_{\text{single}}(Z) \), associated with the given value \( Z(f) \), could be calculated as \( \text{FAP}_{\text{single}}(Z) = 1 - P_{\text{single}}(Z) \), where \( P_{\text{single}}(Z) \) is the cumulative distribution of the corresponding periodogram value, calculated under the base hypothesis \( H_{\theta} \). It is well-known that within simple constant scalefactors these distributions are \( \chi^2(d, F_d, N_d) \) and \( B(d, N_d) \) for the periodograms \( Z, z_2 \) and \( z_1 \), respectively (see e.g. Schwarzenberg-Czerny 1998a,b; Balnev 2008). Here, the quantities in brackets mark the necessary numbers of degrees of freedom.

When the signal frequency is unknown a priori, we need to search for a maximum of \( Z(f) \) within some wide frequency band \([f_{\text{min}}, f_{\text{max}}]\). From now on we will assume, for the sake of definiteness, that \( f_{\text{min}} = 0 \). In practice it is a frequent case, and also this assumption allows us to simplify the formal expressions. All results presented below can be easily extended to the case of arbitrary \( f_{\text{min}} > 0 \). For example, we will need to replace certain lower integration limits appropriately and to change the expressions for the frequency bandwidth from \( f_{\text{min}} \) to \( f_{\text{max}} - f_{\text{min}} \). According to Balnev (2008), to estimate the FAP associated with the observed maximum, we use the Davies (1977, 1987, 2002) bound

\[
\text{FAP}_{\text{max}}(Z, f_{\text{max}}) \leq \text{FAP}_{\text{single}}(Z) + \tau(Z, f_{\text{max}}).
\]

Exact expressions for function \( \tau \) are given in Balnev (2008) for the general least-squares periodogram \( z \) and for its modifications \( z_{1,2,3} \) (see equations 7 and 8 in that paper). In fact, the right-hand side in the inequality (4) represents something more than just an upper bound. It was demonstrated by Balnev (2008) that in the LS periodogram case the inequality (4) appears rather sharp, especially for practically important low-FAP levels. In addition to the bound (4), we will deal with the following approximation:

\[
\text{FAP}_{\text{max}}(Z, f_{\text{max}}) = 1 - \text{P}_{\text{max}}(Z, f_{\text{max}}).
\]

As it was discussed in Balnev (2008), the formulae (5) should provide a good approximation to \( \text{FAP}_{\text{max}} \) uniformly (i.e. for all FAP levels) in the case of small aliasing. Note that the approximation (5) and the bound (4) yield almost coinciding results if \( \text{FAP} < 0.1 \), so that the mentioned property of the approximation (5) probably will not have direct practical application. In this paper, we use (5) just to plot a reference ‘alias-free’ FAP curve.

Now we need to obtain the function \( \tau(z, f_{\text{max}}) \) for our special case of the multiharmonic periodograms. In particular, we need to calculate the factor \( A(f_{\text{max}}) \), present in the expressions for \( \tau \). In general, this factor depends in a rather unpleasant way on the models of the data, on the time series sampling and on the sequence of...
the statistical weights of observations. To attain some technical simplicity, let us first assume that, like in the classical LS periodogram, the base model is empty: $d_k = 0$ and $\mu_k(t) \equiv 0$. In this case, we need to find first the eigenvalues $\lambda_k$ of the $d \times d$ matrix $M$, which is defined as

$$Q = \varphi \otimes \varphi, \quad S = \varphi \otimes \varphi', \quad R = \varphi' \otimes \varphi, \quad M = Q^{-1}(R - S^T Q^{-1} S).$$

(6)

Here, the overline denotes the weighted averaging over the time series and the binary operation $\otimes$ is the dyadic product of vectors ($x \otimes y = x \cdot y^T$) (see appendix A in Baluev 2008). The notation $\varphi'$ stands for the partial derivative of the vectorial function $\varphi(t, f)$ over $f$. We use the expressions from the paper (Davies 1987) to calculate the factor $A(f_{\text{max}})$. We need to combine equations (3.2 and 3.3) and the unnumbered equation following after the equation (3.4) from Davies (1987) to obtain the formula (7) in the paper (Baluev 2008) with

$$A = \frac{\pi^{n-1}}{\Gamma(n + \frac{1}{2})} \int_0^{f_{\text{max}}} df \int_0^{\infty} \left[ 1 - \frac{1}{\prod_{k=1}^{n} \sqrt{1 + x^2 \lambda_k(f)}} \right] dx \frac{1}{\gamma^{1/2}}.
$$

(7)

We need to obtain some more simple, although possibly approximate, expression for the factor $A$. To do this, we first obtain a suitable approximation to the matrix $M$ and hence to its eigenvalues $\lambda_k$. After that, we can substitute the approximations for $\lambda_k$ to (7), in order to derive the final approximation to $A(f_{\text{max}})$. We give the associated details, as well as an assessment of the practical precision of the resulting approximation, in Appendix A. Here, we give only the final result, which seems to be sufficiently accurate in practice. The matrix $M$ can be approximated by the following diagonal block form:

$$M \approx \pi T_{\text{eff}}^2 \begin{pmatrix} I_2 & 0 & \ldots & 0 \\ 0 & 2I_2 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & n^2 I_2 \end{pmatrix},$$

(8)

where $I_2$ is the $2 \times 2$ identity matrix, $T_{\text{eff}}$ is the effective time-span ($T_{\text{eff}} = \sqrt{4\pi D_t}$, where $D_t$ is the weighted variance of timings $t$; see Baluev 2008). The approximate equality (8) implies that the $2n$ eigenvalues required are grouped into $n$ pairs $\lambda_{2k-1} \approx \lambda_{2k} \approx \pi T_{\text{eff}}^2 k^2$, $k = 1, 2, \ldots, n$. Finally,

$$A(f_{\text{max}}) \approx 2\pi^{n+\frac{1}{2}} \alpha_n W,$$

(9)

where $W = f_{\text{max}} T_{\text{eff}}$ and

$$\alpha_n = \frac{2^n}{(2n - 1)!} \sum_{k=0}^{n} (-1)^{n-k} k^{2n+1}$$

(10)

Here, the quantity $(2n - 1)!$ represents the product of all odd integers from $(2n - 1)$ down to $1$. The numerical values of the constants $\alpha_n$ for a few values of $n$ are given in Table 1.

Therefore, using equations (7) and (8) from Baluev (2008), we obtain for the basic multiharmonic periodogram $z(f)$

$$\tau \approx W \alpha_n e^{-\pi^2 z^{n+\frac{1}{2}}},$$

(11)

Table 1. The constants $\alpha_n$ for a few values of $n.$

| $n$ | 1  | 2  | 3  | 5  | 8  | 15 |
|-----|----|----|----|----|----|----|
| $\alpha_n$ | 1 | 1.556 | 1.062 | 0.136 | 9.921 $\times 10^{-4}$ | 1.037 $\times 10^{-10}$ |

Using multiharmonic periodograms

4 NUMERICAL SIMULATIONS

Speaking in terms of the statistical tests theory, two kinds of mistakes can be made in the signal detection problem: the false alarm and the false non-detection. Our primary goal was to keep the FAP at some a priori small levels $\text{FAP} < \text{FAP}$. This is guaranteed by the theoretical inequality (4). Now our goal is to characterize (given the condition of bounded FAP) the detection power, which is provided by the actual precision of the FAP estimation. We noted above that the right-hand side in (4) is expected to provide some approximation to the FAP, not just an upper bound. However, the error of this approximation depends on conditions: in the case when distant periodogram values are weakly correlated, this approximation should be precise, and in the case when there exist pairs (or more complicated combinations) of strongly correlated distant periodogram values, this precision decreases (see Baluev 2008, appendix B). In practice, the absence of strongly correlated peaks means that the periodograms are free from aliases.

Since we have been already prevented (at the given probability FAP) from false alarms by the upper character of the Davies bound (4), now we are more interested in precise approximation of detection thresholds [i.e. such critical values $z_\alpha$ that provide FAP ($z_\alpha$) = FAP] rather than of the FAPs themselves, because it is the detection threshold $z_\alpha$ that determines the detection probability. This means that we should pay major attention to horizontal deviations between the simulated and theoretical FAP curves, rather than to vertical ones.

We now proceed to testing the precision of the theoretical approximations obtained above using Monte Carlo simulations of $\text{FAP}_{\text{max}}$, in the same way as in Baluev (2008). When the order of the approximating trigonometric polynomial grows, the volume of necessary calculations increases significantly due to the following reasons.

(i) The calculations of single values of the multiharmonic periodogram require to solve higher dimensional linear least-squares problem (or to orthogonalize higher dimensional functional bases).

(ii) As the simulations have shown, the average density of peaks on the multiharmonic periodograms increases roughly as $O(n)$.

This requires for the calculations to be performed on a more dense frequency grid, in order to obtain enough accurate values of periodogram maxima.

Therefore, our abilities in making numerical simulations are severely limited to small $n$ only.

First, let us deal with the case when the time series does not produce any aliasing in the classical sense, i.e. on the LS periodogram. This is the case of a large number of evenly distributed observations. The corresponding simulated $\text{FAP}_{\text{max}}$ curves are shown in Fig. 1 for the periodograms $z(f)$. We can see that the theoretical approximations work quite well. Nevertheless, the small deviations for the cases $n \geq 2$ contrast with the LS case $n = 1$, for which we cannot see any deviation at all. Probably these small deviations emerged because of an extra correlation of distant periodogram...
One may argue that such understanding of the notion ‘aliasing’ is not traditional, because the associated effect is not connected with uneven time series sampling, and is only a result of an interplay between the main period and its subharmonics. Nevertheless, for the sake of a uniform terminology, we name here all ‘wrong’ periodogram peaks as aliases, and the associated phenomenon of correlativity of distant periodogram values as aliasing.

The paper (Baluev 2008) in fact paid undeservedly small attention to the modified LS periodograms $z_{1,2,3}(f)$. It was assumed that the behaviour of their FAP curves is similar to the behaviour of the FAP curves of the basing LS periodogram. However, they are the modified periodograms which are usually used in practice. Here, we try to correct this mistake. It appears that the FAP curves of modified periodograms are considerably less sensitive to an uneven time series sampling. Consequently, the precision of the Davies bound (4) and of the alias-free approximation (5) appears significantly better (see Fig. 3). For the modified multiharmonic periodograms, the random distribution of timings does not introduce any significant perturbation of the FAP curve even for $N$ as small as 30. In this case, the FAP curves for the modified periodograms perfectly agree with the alias-free approximation (5). For periodically gapped timings, the precision of the analytic FAP estimations improves too. Moreover, this precision does not decrease and even seem to increase when the order $n$ or the frequency bandwidth $f_{\text{max}}$ grows. The reason for such refinement of the precision of the analytic estimations of the FAP for the modified periodograms is unclear.

It is harder to complete a similar series of Monte Carlo simulations for more complicated cases, e.g. with the base model $\mu_{\text{H}}$ incorporating at least a constant or a linear trend. We present only a few examples of such simulations, which nevertheless further certify the practical efficiency of the closed expressions for the FAP described above (Fig. 4). Actually, it looks that a low-order polynomial trend in the base model $\mu_{\text{H}}$ does not introduce any visible deviation in the simulated FAP curve, at least in this particular case.

Finally, let us take some realistic time series sampling and consider the associated FAP curves and their approximations under some realistic conditions. For this purpose, as in the paper (Baluev 2008), we use the observational dates and standard errors of the high-precision radial velocity data for the stars 51 Peg and 70 Vir (Naef et al. 2004). The number of observations in the first time series is $N = 153$ and in the second one $N = 35$. The time-span of these time series is about a decade. In both cases, significant aliasing is present (e.g. corresponding to the annual and diurnal periods). We can see, however, that in both cases the analytic formulae for the periodogram $z_1$ work very well (Fig. 5).

It is worth noting that in all the cases discussed above, the Davies bound (4) indeed bounds the simulated FAP curves from the upper-right side, at least for not very small levels FAP $> 10^{-3}$, which can be reliably modelled using $10^5$ Monte Carlo trials.

5 CONCLUSIONS

In this paper, previous results by Baluev (2008) are applied to the case when the model of the signal to be detected represents a truncated Fourier polynomial. Closed analytic expressions for the false alarm probabilities, associated with multiharmonic periodogram peaks, are given. They are tested under various conditions using Monte Carlo simulations. The simulations have shown that the accuracy of the mentioned theoretical estimations usually is quite suitable in practice. Also, these simulations have revealed an unexpected (but pleasant) phenomenon: the accuracy of the above theoretical approximations of the FAP is considerably better for the normalized multiharmonic periodograms than for the basic, purely least squares, ones. Since in practice the observational noise variance is rarely known precisely, they are the normalized periodograms that are usually dealt with. Therefore, the better behaviour of the FAP curves for the normalized periodograms has high practical value.

The necessary amount of Monte Carlo simulations for the multiharmonic periodogram is bigger than for the LS one. It may appear...
Using multiharmonic periodograms

Figure 2. Simulated versus analytic FAP for the basic multiharmonic periodogram $z(f)$. For the panels in the left-hand and middle columns, the time series consisted of $N = 100$ and 30 randomly spaced data points, respectively. For the panels in the right-hand column, $N = 100$ data points were clumped in 10 evenly spaced groups. Each group consisted of 10 points and spanned only $1/50$ fraction of the total time-span (instead of the natural $1/10$ fraction). In each panel, four converging bunches of curves from left to right correspond to $n = 1, 2, 3, 5$. For the top row $f_{\text{max}}^T = 50$, for the middle one $f_{\text{max}}^T = 500$ and for the bottom one $f_{\text{max}}^T = 5000$.

Figure 3. Same as in Fig. 2, but for the modified periodogram $z_1(f)$.

very difficult to obtain a sufficiently precise Monte Carlo estimation of the FAP even in the case of a single time series. Most likely, for surveys dealing with large numbers of separate time series, it would be impossible to perform the necessary amount of Monte Carlo simulations. For example, a single CPU at 2 GHz would complete all Monte Carlo simulations presented above in a few months only. On the contrary, the closed theoretical estimations presented in this paper do not require any simulations at all, and simultaneously often have a nice accuracy. This indicates that the mentioned estimations represent a promising practical tool and may be used in a wide variety of astronomical applications, involving search for non-sinusoidal periodicities in observational data. The corresponding research fields...
Simulated versus analytic FAP for the multiharmonic (n = 2) periodogram \( z_1(f) \), constructed from \( N = 100 \) evenly spaced observations, in the main frequency band \( f_{\text{max}}T = 50 \). The graph shows four simulated FAP curves for different degrees of the polynomial trend in the base model \( \mu_{3t} \): empty base model (\( d_{3t} = 0 \)), a constant term (\( d_{3t} = 1 \)), a linear trend (\( d_{3t} = 2 \)), and a quadratic trend (\( d_{3t} = 3 \)). All these curves appear almost coinciding. For an intercomparison, we show here the theoretical distribution curves for all the modified periodograms, \( z_1, z_3 \) and \( z_2 \) (from left to right). Note that we plot them only for the case \( d_{3t} = 0 \), because the similar curves for \( d_{3t} = 1, 2, 3 \) did not show any visible deviation.

Figure 4.

Simulated versus analytic FAP for the modified multiharmonic (\( n = 2 \)) periodogram \( z_1(f) \), constructed from \( N = 100 \) evenly spaced observations, in the main frequency band \( f_{\text{max}}T = 50 \). The graph shows four simulated FAP curves for different degrees of the polynomial trend in the base model \( \mu_{3t} \): empty base model (\( d_{3t} = 0 \)), a constant term (\( d_{3t} = 1 \)), a linear trend (\( d_{3t} = 2 \)), and a quadratic trend (\( d_{3t} = 3 \)). All these curves appear almost coinciding. For an intercomparison, we show here the theoretical distribution curves for all the modified periodograms, \( z_1, z_3 \) and \( z_2 \) (from left to right). Note that we plot them only for the case \( d_{3t} = 0 \), because the similar curves for \( d_{3t} = 1, 2, 3 \) did not show any visible deviation.

Figure 5.

Simulated versus analytic FAP for the multiharmonic (\( n = 3 \)) periodogram \( z_1(f) \) constructed from the radial velocity time series of 51 Peg (top panel) and 70 Vir (bottom panel). The base model \( \mu_{3t} \) incorporated a free constant term (\( d_{3t} = 1 \)). In each panel, three bunches of curves correspond to \( P_{\text{max}} = 1/f_{\text{max}} = 100 \) and 1 d (from left to right).

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (Grant 09-02-00230) and by the Russian President Programme for the State Support of Leading Scientific Schools (Grant NSh-1323.2008.2). I am grateful to the anonymous referee for providing important suggestions, which helped me to improve the manuscript.

References

Baluev R. V., 2008, MNRAS, 385, 1279
Baluev R. V., 2009, MNRAS, 393, 969
Cumming A., 2004, MNRAS, 354, 1165
Cumming A., Marcy G. W., Butler R. P., 1999, ApJ, 526, 890
Davies R. B., 1977, Biometrika, 64, 247
Davies R. B., 1987, Biometrika, 74, 33
Davies R. B., 2002, Biometrika, 89, 484
Frescura F. A. M., Engelbrecht C. A., Frank B. S., 2008, MNRAS, 388, 1693
Gilliland R. L., Baliunas S. L., 1987, ApJ, 314, 766
Horne J. H., Baliunas S. L., 1986, ApJ, 302, 577
Irwin A. W., Campbell B., Morbec C. L., Walker G. A. H., Yang S., 1989, PASP, 101, 147
Koen C., 1990, ApJ, 348, 700
Lomb N. R., 1976, Ap&SS, 39, 447
Nael D., Mayor M., Beuzit J. L., Perrier C., Queloz D., Sivan J. P., Udry S., 2004, A&A, 414, 351
Palmer D. M., 2009, ApJ, in press (arXiv:0901.1913)
Scargle J. D., 1982, ApJ, 263, 835
Schwarzenberg-Czerny A., 1996, ApJ, 460, L107
Schwarzenberg-Czerny A., 1998a, MNRAS, 301, 831
Schwarzenberg-Czerny A., 1998b, Balt. Astron., 7, 43
Zechmeister M., Kürster M., 2009, A&A, in press (arXiv:0901.2573)

Appendix A: The Factor \( F_{\text{MAX}} \)

The elements of the matrices \( Q, S, R \) can be transformed in the way similar to equations (10) in Baluev (2008). The matrix \( Q \) will contain the averages of the kind \( \sin koT \) and \( \cos koT \). The matrix \( S \) will contain components \( i \sin koT, i \cos koT \) and also \( i \). The matrix \( R \) will contain components of the kind \( i \sin koT, i \cos koT \) and also \( i^2 \). Here, \( k = 1, 2, \ldots, 2n \). Therefore, we deal with quantities having the form

\[
\Omega_i(f) = \sin koT, \quad \Omega_i(f) = \cos koT,
\]

\[
\Lambda_i(f) = i \sin koT, \quad \Lambda_i(f) = i \cos koT,
\]

\[
\Xi_i(f) = i^2 \sin koT, \quad \Xi_i(f) = i^2 \cos koT.
\]

(A1)

and with the similar overtonic quantities, calculated at the frequencies \( 2f, 3f, \ldots, 2nf \). Now our goal is to show that under certain conditions the quantities \( \Omega_i, \Lambda_i, \Xi_i \) have small magnitude in comparison with the quantities \( i, i, i^2 \), respectively, and thus can be neglected. Let us assume that at the given frequency \( f \) the phases \( \omega, t \) are distributed approximately uniformly in the segment \([0, 2\pi]\). This means that the multipliers \( \sin koT, \cos koT \) in (A1) may be considered as random quantities. Their values are jumping randomly in the segment \([-1, 1]\), whereas the functions 1 and \( i^2 \) are varying slowly. Therefore, the mentioned sines and cosines may be treated as random quantities not correlated with the timings \( t \). This quasi-random property allows us to write down approximations like

\[
\sin koT \sim 1/\sqrt{N} \quad \text{and} \quad \cos koT \approx (i)(\sin koT) \sim i/\sqrt{N}.
\]

Therefore, all the quantities (A1) may be expected to be negligible (at the given frequency \( f \)) when the values of \( \Omega_i, \Lambda_i, \Xi_i \) are small. It is not hard to see that \( \Omega(f) = \Omega_i(f) + i\Xi_i(f) = \Re \text{e}^{\phi}(i) \) (with \( i \) being the imaginary unit) represents the complex spectral window of the time series and the square of its module is the usual spectral window. Typically, the spectral window contains a strong narrow peak at \( f = 0 \) and a series of smaller peaks, corresponding to aliasing frequencies. Therefore, in the case when the spectral window does not contain any strong peaks at the frequencies \( f, 2f, \ldots, 2nf \), we can keep in the matrices \( Q, S, R \) only the terms which do not contain sines or cosines inside the averaging operation. In this approximation, the matrix \( M \) can be calculated easily. The result
is given in (8), and the eigenvalues required are approximated as \( \lambda_N \approx \lambda_{2N-1} \approx \pi T_N^2 k^2 \).

It is not hard to check that when our base model \( \mu_N \) is not empty but contains a free constant term or a low-order polynomial drift with free coefficients, the same approximation for the matrix \( \mathbf{M} \) holds true under similar conditions. In this case, the base model \( \mu_N \) appears approximately orthogonal to the signal model \( \mu \) in the sense that the cross averages \( \varphi^T \mu \) can be neglected in comparison with the respective elements of the matrices \( \varphi^T \varphi \) and \( \varphi \otimes \varphi \).

Coupled with the obtained approximate expressions for the eigenvalues \( \lambda_k \), the equation (7) yields the equation (9) with

\[
\alpha_k = \frac{1}{2\pi i} \int_0^\infty \left[ 1 - \frac{1}{\prod_{i=1}^n (1 + xk^2)} \right] dx/\pi \approx R_n + \infty \mathrm{f}_i(\varphi_0 x \text{in the upper } C^+ \otimes \lambda_1)!!
\]

(\( k \approx R_n + \infty \))

\[
\alpha_k = \frac{2^n}{(2n-1)!!} \int_{-\infty}^{+\infty} \left[ 1 - \frac{1}{\prod_{i=1}^n (1 + x^2k^2)} \right] dx/k^2.
\]

(A2)

We can calculate the integral in (A2) using the theory of functions of a complex variable. Denoting the integrand in the last integral in (A2) as \( f(x) \), we can easily check that \( \lim_{x \to \infty} |xf(x)| = 0 \) (where \( x \) is considered as a complex variable). This means that we can replace the integration line \( (-\infty, +\infty) \) by a closed contour \( C_R \), representing a semicircle of the radius \( R \to \infty \) in the upper complex semiplane. Indeed, the integral over the semicircle arc decays at least as rapidly as \( \sim \pi R f(R) \sim \pi/\sqrt{R} \to 0 \) when \( R \to \infty \), and the integral over the diameter of the semicircle \( (-R, R) \) tends to the integral within \( (-\infty, +\infty) \) that we need to compute.

The integrand \( f(x) \) can be represented as a ratio of two algebraic polynomials: \( f(x) = P(x)/Q(x) \), where \( Q(x) = \prod_{i=1}^n (1 + x^2k^2) \) and \( P(x) = |Q(x) - 1|x\) [it is not hard to see that \( P(x) \) is indeed a polynomial of degree \( 2n - 2 \), because the free term in \( Q(x) \) is unit and hence the denominator \( x^2 \) is reduced]. Therefore, the integral over \( C_R \) can be expressed via the sum of residues of \( f(x) \) in the points \( x_i = i/k, k = 1, 2, \ldots, n \) (with \( i \) being the imaginary unit), which represent the roots of \( Q(x) \) in the upper complex semiplane. That is,

\[
\frac{1}{2\pi i} \int_{-\infty}^{+\infty} f(x)dx = \lim_{R \to \infty} \int_{C_R} f(x)dx = \sum_{k=1}^n \text{Res } f(x_k).
\]

(A3)

Since the singularities \( x_k \) are simple poles, the corresponding residues can be evaluated as

\[
\text{Res } f(x_k) = \frac{P(x_k)}{Q(x_k)} = \frac{k}{2i \prod_{j=1, j \neq k}^n (1 - j^2/k^2)} = \frac{k2^{n-1}}{2i \prod_{j=1, j \neq k}^n (k^2 - j^2)} = \frac{(-1)^{n-k}k^{2n+1}}{i(n+k)(n-k)!}.
\]

(A4)

The formulae (A2–A4) yield the final expression (10).

Note that alternatively we could use the treatment involving ellipsoidal surfaces in multidimensional spaces (see equation B7 by

---

**Figure A1.** The figure shows the precision of the alias-free approximation of the factor \( A(f_{\text{max}}) \). Top-left panel: the graph of the ratio of the derivative \( \Lambda' \left( f_{\text{max}} \right) \) (the inner integral in 7) to its alias-free approximation \( 2\pi^{11/2} \alpha_{T_{\text{eff}}} \). On an almost horizontal graph, we can see a sequence of strong but narrow splashes corresponding to aliasing periods. Bottom-left panel: the similar graph for the function \( A(f_{\text{max}}) \) itself. The splashes at the aliasing frequencies exist but are very small and do not produce significant perturbations. The data were obtained for \( n = 3 \) and \( d_{T_{\text{eff}}} = 1 \) (with a free constant term in the model \( \mu_N \)). The \( N = 100 \) timings of the mock input time series were periodically gapped with a frequency corresponding to \( T f \approx 28.5 \). At this gapping frequency, the folded phases spanned only \( \approx 10 \) per cent of the full period. Right-hand panels show similar graphs for the case of \( N = 30 \) randomly spaced observations, \( n = 5 \) and \( d_{T_{\text{eff}}} = 1 \).
Baluev 2008). This way seems to be less convenient to obtain exact formulae for \( \alpha_n \), but nonetheless it yields a simple upper bound

\[
\alpha_n \leq \frac{1}{(n-1)!} \sqrt{\frac{1}{n} \sum_{k=1}^{n} k^2} = \frac{1}{(n-1)!} \sqrt{\frac{(n+1)(2n+1)}{6}}.
\]  
(A5)

The comparison of this bound with numerical values from Table 1 shows that this bound is remarkably sharp.

Formally, the approximation (9) was based on certain assumptions of negligible aliasing, which we have discussed above. Nevertheless, it was demonstrated in Baluev (2008) for the LS periodogram that this approximation of the factor \( A(f_{\text{max}}) \) is quite precise in practice, even when the aliasing effects are strong. We may expect the same behaviour of \( A(f_{\text{max}}) \) for multiharmonic periodograms. This is due to the integral character of the representation (7). Indeed, the aliasing may result in a strong perturbation of the eigenvalues \( \lambda_k \) and hence of the inner integral in (7). However, these perturbing effects are locked in very narrow frequency intervals of the typical width \( \Delta W \sim 1 \). After integration over a wide frequency range with \( W \gg 1 \), the resulting perturbation in the whole integral appear insignificant. This is illustrated in Fig. A1.

Therefore, the only practically important source of a possible inaccuracy of the analytic FAP estimation lies in the possible unsharpness of the Davies bound (4) itself and in the possible inaccuracy of the associated alias-free approximation (5) itself.

This paper has been typeset from a \TeX/\LaTeX\ file prepared by the author.