Detecting non-sinusoidal periodicities in observational data: the von Mises periodogram for variable stars and exoplanetary transits

Roman V. Baluev\textsuperscript{1,2}\textsuperscript{*}

\textsuperscript{1} Central Astronomical Observatory at Pulkovo of the Russian Academy of Sciences, Pulkovskoje shosse 65, St Petersburg 196140, Russia
\textsuperscript{2} Sobolev Astronomical Institute, St Petersburg State University, Universitetskij prospekt 28, Petrodvorets, St Petersburg 198504, Russia

Accepted 2013 February 6. Received 2013 February 5; in original form 2012 December 26

\textbf{ABSTRACT}

This paper introduces an extension of the linear least-squares (or Lomb–Scargle) periodogram for the case when the model of the signal to be detected is non-sinusoidal and depends on unknown parameters in a non-linear manner. The problem of estimating the statistical significance of candidate periodicities found using such non-linear periodograms is examined. This problem is related to the task of quantifying the distributions of the maximum values of these periodograms. Based on recent results in the mathematical theory of extreme values of a random field (the generalized Rice method), a general approach is provided to find a useful analytic approximation for these distributions. This approximation has the general form $e^{-P(\sqrt{z})}$, where $P$ is an algebraic polynomial and $z$ is the periodogram maximum.

The general tools developed in this paper can be used in a wide variety of astronomical applications, for instance in the study of variable stars and extra-solar planets. With this in mind, we develop and consider in detail the so-called von Mises periodogram – a specialized non-linear periodogram in which the signal is modelled by the von Mises periodic function $\exp(v \cos \omega t)$. This simple function with an additional non-linear parameter $v$ can model the light curves of many astronomical objects that show various types of periodic photometric variability. We prove that our approach can be perfectly applied to this non-linear periodogram.

We provide a package of auxiliary C++ programs, attached as online-only material. These programs should facilitate the use of the von Mises periodogram in practice.

\textbf{Key words:} methods: analytical – methods: data analysis – methods: statistical – surveys.

\section{1 Introduction}

The Lomb (1976)–Scargle (1982) (hereafter LS) periodogram is a well-known powerful tool that 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 (harmonic) model of the signal and then to check how much the resulting value of the $\chi^2$ function improves for a given signal frequency. The maximum value of the LS periodogram (i.e. the maximum decrement in the $\chi^2$ 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. The linearity of the harmonic model with respect to unknown parameters (two coefficients near the sine and cosine) introduces additional simplifications.

Any signal detection tool is not of much use without an accompanying method of estimating the statistical significance of candidate periodicities. Indeed, the random errors contaminating the input data inspire noise fluctuations on the periodogram, so that we can never be completely sure that the peak that we are actually observing is a result of real periodicity in the data. To assess the statistical significance of the observed periodogram peak, we need to calculate the ‘false alarm probability’ (hereafter FAP) associated with this peak. The FAP is the probability that the observed or a larger periodogram peak could be produced by random measurement errors. The smaller the FAP is, the larger the statistical significance. Given some small tolerance value FAP\textsubscript{\text{R}} (say, 1 per cent), we can claim that the detected candidate periodicity is statistically significant (when FAP < FAP\textsubscript{\text{R}}) or is not (when FAP > FAP\textsubscript{\text{R}}).

From the statistical point of view, the FAP is closely connected with the probability distributions of the periodogram considered under the null hypothesis (i.e. the hypothesis of no signal in the data). If the frequency of the putative signal were known a priori, a single value of the LS periodogram would be enough to check whether the presence of this periodicity is likely or not. In this case, the FAP is given by the well-known exponential distribution of any single value $z$ of the LS periodogram, so that FAP = $e^{-z}$. However, the case in which the frequency of a possible signal is basically unknown is much more common. In this case, the FAP is...
provided by the distribution function of the maximum value of the periodogram (corresponding to the frequency range being scanned).

The calculation of the latter distribution is a non-trivial task. The absence of an accurate and/or rigorous analytic expression for this distribution (even for the plain LS periodogram) represented a significant problem for astronomers for about three decades. In addition to the Lomb and Scargle works, it is worthwhile to mention here the papers by Horne & Baliunas (1986), Koen (1990), Schwarzenberg-Czerny (1998a), Cumming, Marcy & Butler (1999), Cumming (2004) and Frescura, Engelbrecht & Frank (2008). We believe that this obstacle is the main reason why only very simple extensions of the LS periodogram have attained widespread popularity. Theoretically, it is not difficult to construct a periodogram in which intricate models of the data are used. Armed with modern computers, it may be possible to evaluate such periodograms in practice, even if they rely on CPU-greedy numerical algorithms. But what should be done next? How do we decide which of the signals detected are real and which belong to the noise? The only general solution available is the Monte Carlo simulation technique, which might be useful for the basic LS periodogram but is not, unfortunately, for more complicated cases.

Recently, significant progress in this field was made by Baluev (2008), who give entirely analytic and simultaneously accurate approximations of the FAP, based on results from the theory of extreme values of stochastic processes (the ‘Rice method’). In brief, the main result presented in Baluev (2008) for the LS periodogram is:

\[ \text{FAP}(z) \leq M(z) \approx e^{-z^2/2}, \]  

(1)

where \( z \) is the maximum periodogram value corresponding to a given frequency range, and \( W \) is the width of this range multiplied by a certain effective length of the time series (which is usually close to the plain time span). The symbol ‘\( \leq \)’ in (1) means that FAP(\( z \)) will never exceed \( M(z) \), and simultaneously \( M(z) \) represents an asymptotic approximation for FAP(\( z \)), with the error decreasing for small FAP (or large \( z \)). The high practical importance of the approximation in (1) is based on three things: (i) it is entirely analytic, eliminating any need for Monte Carlo simulations; (ii) its practical accuracy usually appears good or at least quite satisfactory; and (iii) its possible errors never lead to more false alarms than we expect, as we are dealing with an upper limit on FAP.

The LS periodogram can be easily generalized in multiple ways to encompass more complicated models (Schwarzenberg-Czerny 1998a,b; Baluev 2008; Zechmeister & Kürster 2009; Ferraz-Mello 1981). First, we can introduce a base model of an expected underlying variation (typically a long-term polynomial trend) and check whether the addition of a probe sinusoidal signal offers enough improvement in \( \chi^2 \). These cases were briefly considered in Baluev (2008), where the general conclusion was that such a modification does not typically invalidate the result (1). Second, we can deal with more complicated (but still linear) models than just a sinusoid. In particular, in Baluev (2009b) so-called multiharmonic periodograms were considered, where the periodic signal is modelled by a trigonometric polynomial involving the leading few terms of the Fourier series (Schwarzenberg-Czerny 1996). In this case, formula (1) is generalized to

\[ \text{FAP}(z) \leq M(z) \approx \alpha_n e^{-z^2/2}, \]  

(2)

where \( \alpha_n \) are certain numbers depending on the degree \( n \) of the approximating trigonometric polynomial. Note that \( n = 1 \) corresponds to the LS case.

However, non-sinusoidal periodic signals, which are dealt with in astronomy, often obey non-linear models. Then the use of the LS periodogram or periodograms from Baluev (2008, 2009b) is not optimal, as the corresponding periodic variation might be fitted by an inadequate model. For instance, this is the case for light curves of variable stars and for radial velocity curves of spectral binaries involving elongated orbits. Theoretically, we could use a high-order Fourier expansion to approximate a non-sinusoidal periodicity, but this solution is obviously inefficient owing to the unnecessarily large number of extra free parameters. The aim of the present paper is to extend the results from Baluev (2008) and Baluev (2009b) to the case of an arbitrary model of the periodic signal, incorporating a few parameters in a non-linear manner. As will be demonstrated, we can apply roughly the same technique (the Rice method) to this case, with the major difference that we now deal with random fields instead of random processes. That is, a closed approach to constructing the limiting approximation \( M(z) \) in the form \( \text{We}^{-z^2/2} \cdot \text{P}(\sqrt{z}) \) will be provided, where \( \text{P} \) is an algebraic polynomial.

The structure of the paper is as follows. In Section 2, a general definition extending the LS periodogram to the non-linear case is introduced. In Section 3, the problem of assessing the statistical significance of candidate periodicities detected with the non-linear periodogram is considered. This description is followed by an auxiliary Section 4 devoted to the methods of practical evaluation of the theoretical approximations of Section 3. In Section 5, the consequences implied by various noise models of the data are discussed. In Section 6, two concrete practical applications of these results are supplied. In the first (rather tutorial) example, the aim is to detect a periodic signal of arbitrary (but a priori fixed) shape, when the unknown parameters are the amplitude and the phase of the signal. In the second example, a more complicated periodogram is considered based on the so-called von Mises model of the signal, essentially \( \exp(\nu \cos x) \), which involves an additional non-linear parameter \( \nu \).

2 DEFINITION OF THE NON-LINEAR PERIODOGRAM

Let \( x_1 \) denote the outcomes of \( N \) observations made at times \( t_i \). The errors of these measurements are assumed to follow Gaussian distributions and to be statistically independent (hence, uncorrelated). The standard deviations of these errors, \( \sigma_i \), are assumed to be known a priori. We want to test whether these observations are consistent with some base model of variation, or whether a certain deterministic periodicity is also present.

The model to be tested for consistency with the data is \( \mu_N(t, \theta_N) \), where the vector \( \theta_N \) incorporates \( d_N \) unknown parameters, which should be estimated from the data. We assume that this model is linear with respect to unknown parameters:

\[ \mu_N(t, \theta_N) = \theta_N \cdot \varphi_N(t), \]  

(3)

where the vector of base functions \( \varphi_N(t) \) is set a priori. Typically, the model \( \mu_N \) incorporates a free constant term, and, possibly, a long-term polynomial trend with free coefficients. The LS periodogram, by the way, assumes that \( \mu_N \equiv 0 \), implicitly requiring some preliminary centring of the time series.

The model of the periodic signal is given by \( \mu(t, \theta, \nu) \), where the vector \( \theta \) contains \( d \) unknown parameters to be estimated from the data together with the frequency \( \nu \). The united vector \( \theta_N = [\theta_N, \theta] \)
parametrizes the compound alternative model of the data,\(^1\) which is given by
\[
\mu_{\kappa}(t, \theta_K, f) = \mu_{\kappa_0}(t, \theta_K) + \mu(t, \theta, f).
\]

We denote the \(d\)-dimensional domain, where \(\theta\) is supposed to reside, by \(\Theta\). The signal is supposed to vanish when \(\theta\) belongs to some ‘null domain’ \(\Theta_0 \subset \Theta\) and not to vanish when \(\theta\) does not belong to \(\Theta_0\). Therefore, we wish to test whether the data are consistent with the base hypothesis \(H : \theta \in \Theta_0\) (implying that the model \(\mu_{\kappa_0}(t, \theta_K)\) fits the data satisfactorily) or whether this base hypothesis should be rejected in favour of the alternative \(K : \theta \in \Theta \setminus \Theta_0\) (implying the model \(\mu_{\kappa}(t, \theta_K, f)\)). The model \(\mu\) may be non-linear with respect to \(\theta\).

The unknowns \(\theta_K\), \(\theta\) and \(f\) can be estimated using the least-squares approach. Under the hypothesis \(H\), the best-fitting estimation of \(\theta_K\) can be obtained by minimizing the functions \(x_{\kappa_0}^2(\theta_K) = \left(\bar{x} - \mu_{\kappa_0}(t, \theta_K)\right)^2\). Under the hypothesis \(K\), the best-fitting estimations of \(\theta_K\), \(\theta\) and \(f\) should correspond to the minimum value of the function \(x_{\kappa}(\theta_K, f) = \left(\bar{x} - \mu_{\kappa}(t, \theta_K, f)\right)^2\). Because \(\mu_{\kappa_0}\) is linear, the minimizations by \(\theta_K\) can be performed rapidly and precisely using the usual linear least-squares algorithms. The minimization of \(x_{\kappa_0}^2\) over the remaining variables is equivalent to the maximization of the non-linear function
\[
\zeta(\theta, f) = \frac{1}{2} \left[ \min_{\theta_K} x_{\kappa_0}^2(\theta_K) - \min_{\theta_K} x_{\kappa}^2(\theta_K, \theta, f) \right],
\]
which simultaneously characterizes the improvement in the \(\chi^2\) fit quality, which is achieved by means of adding to the base model the model of the periodic signal with given values of \(\theta\) and \(f\). Note that the maxima of \(\zeta(\theta, f)\) do not depend on the choice of the parametrization. That is, they are invariant with respect to a non-degenerated transformation of the vector \(\theta\) (and any non-degenerated linear transformation of \(\theta_K\)).

We can perform the minimization over the frequency \(f\) in a traditional manner by means of looking for the highest peak on the graph of the function
\[
z(f) = \max_{\theta_K} \zeta(\theta, f),
\]
which can be called the ‘non-linear least-squares periodogram’. This definition means that for any fixed frequency \(f\) we perform the fit of our model via the remaining \(d \theta\) parameters \(\theta\). The value of the periodogram characterizes the relevant advance in the \(\chi^2\) fit quality. When the model \(\mu(t, \theta, f)\) is linear with respect to \(\theta\), this definition of \(z(f)\) coincides with the definition of the linear least-squares periodogram from Balau (2008).

In the majority of practical applications, one of the parameters in \(\theta\) is the amplitude \(K\) of the periodic variation. This means that
\[
\mu(t, \theta, f) = K h(t, \xi, f),
\]
where the vector \(\xi\) contains the \((d - 1)\) remaining unknown parameters of the signal. We assume that \(\xi\) belongs to some domain \(\Xi\) in \((d - 1)\) dimensions, so that the domain \(\theta\) represents the Cartesian product \(\{0, +\infty\} \times \Xi\) or \(( -\infty, +\infty) \times \Xi\), and the null domain is the domain of zero amplitude: \(\Theta_0 = \{K = 0\} \times \Xi\). For simplicity, let us first consider the case when \(d_L = 0\) and the hypothesis \(H\) states that the data do not contain anything but the white Gaussian noise. In this case, \(x_{\kappa_0}^2 = \langle x^2 \rangle\) and \(\zeta(\theta, f) = \langle x^2 \rangle - \langle x^2 \rangle/2\). The function \(K = \langle h^2 \rangle/\langle h^2 \rangle\) is a quadratic polynomial of \(K\), which can be easily maximized given fixed \(f\) and \(\xi\). This results in a least-squares estimation \(K^* = \langle h^2 \rangle/\langle h^2 \rangle\), and in the maximum \(K^* = \eta^2/2\), where \(\eta(\xi, f)\) represents the new function to be maximized by the remaining parameters. It can be expressed as
\[
\eta(\xi, f) = \langle x \psi \rangle = \langle x \psi \rangle / \sqrt{\langle h^2 \rangle},
\]
where \(\psi(t, \xi, f) = h(t, \xi, f) / \sqrt{\langle h^2 \rangle}\). A similar result can be obtained for the case when the relation (7) is still valid, but the model \(\mu_{\kappa_0}\) is no longer empty. It is not hard to check that, if the models \(\mu\) and \(\mu_{\kappa_0}\) were orthogonal in the sense that \(\langle h\phi_{\eta_0} \rangle = 0\), the maximum of \(\zeta(\eta, K)\) could be expressed in exactly the same way as was described in the previous paragraph.

In the general case the models are not orthogonal, and we introduce the new model function
\[
h(t, \xi, f) = h(t, \xi, f) - \langle Q_{\eta_0, \kappa_0}^{-1} Q_{\eta_0, \kappa_0} \psi \cdot \psi \rangle(t),
\]
where \(Q_{\eta_0, \kappa_0} = \langle \phi_{\eta_0} \otimes \phi_{\eta_0} \rangle\) is the \(d_L \times d_L\) Fisher information matrix associated with \(\eta_0\), and \(Q_{\eta_0, \kappa_0} = \langle \phi_{\eta_0} h \rangle\) is the \(d_L \times 1\) Fisher information matrix for the parameters \(\theta_{\eta_0}\) and \(K\). Because the identity \(\langle \phi_{\eta_0} h \rangle = 0\) holds true, the new model of the signal is orthogonal to the base model. Now \(h\) should replace \(h\) in the expression for \(\psi\), so that
\[
\psi(t, \xi, f) = \frac{h(t, \xi, f) - \langle Q_{\eta_0, \kappa_0}^{-1} Q_{\eta_0, \kappa_0} \psi \cdot \psi \rangle(t)}{\sqrt{\langle h^2 \rangle - Q_{\eta_0, \kappa_0}^{-1} Q_{\eta_0, \kappa_0} \psi \cdot \psi}},
\]
After that, we can directly calculate the quantity \(\eta\) from equation (8).

Finally,
\[
\max_K \eta = \eta^2/2,
\]
with
\[
K^* = \eta / \sqrt{\langle h^2 \rangle - Q_{\eta_0, \kappa_0}^{-1} Q_{\eta_0, \kappa_0} \psi \cdot \psi}.
\]
The best-fitting values of \(\xi\) and \(f\) correspond to the maximum of \(\eta\).

Often it might be useful to assume that negative values for \(K\) are not allowed. In that case we should make a small amendment to the last formula (11) and (12). Namely, they can be used only for \(\eta > 0\), while for \(\eta < 0\) we set \(x + \kappa_0 = 0\) and \(K^* = 0\) with the best-fitting values of \(\xi\) and \(f\) undefined.

The formulae become simpler for an important practical case when \(d_L = 1\) and \(\phi_{\eta_0} = 1\) reflects a free constant offset of the data.

In this case, let us first define
\[
x_c = (x - c)/(1), \quad h_c(\xi, f) = (h_c - c)/(1), \quad D = (h_c - h_c)^2,
\]
and then evaluate
\[
\eta = ((x - x_c)(h - h_c)) / \sqrt{D}, \quad \max_K \eta = \eta^2/2, \quad K = \eta / \sqrt{D}.
\]

Note that the quantity (1) represents the sum of weights of all observations.

3 APPROXIMATING THE FALSE ALARM PROBABILITY USING THE RICE METHOD

3.1 General introduction to the problem

In this paper we are interested in the FAP associated with the observed maximum peak \(z_{\max} = \max_{|t| , f} z(f)\), where \(f_{\max}\) is some
a priori given maximum frequency. This FAP can be formally defined as follows:
\[
\text{FAP}(\zeta_{\text{max}}) = \Pr[\exists \theta \in \Theta, f \in [0, f_{\text{max}}] : \zeta(\theta, f) > \zeta_{\text{max}}],
\]
again with the probability operator calculated under the null hypothesis (no actual signal in the data). It can be seen that to assess the FAP, the distribution function of the maximum values of \(\zeta(\theta, f)\) should be known. The function \(\zeta\) represents a real-valued random field defined on a domain of dimension \((d + 1)\).

It could be claimed that because the model of the extra signal contains \(d\) free parameters and because the quantity \(z(f)\) is the logarithm of the likelihood ratio statistic, the distribution of \(2z(f)\) (for a fixed \(f\)) should tend to the \(\chi^2\) distribution with \(d\) degrees of freedom, when the number of observations grows. This was assumed, for instance, by Cumming (2004), who considered the case of Keplerian velocity variation with four unknown parameters (plus the period). The reader is cautioned that in general this assumption is incorrect because the conditions of the corresponding limiting theorem are not satisfied. The most important reason comes from the fact that the parameters \(\xi\) have no physical sense (are undefined) when \(K = 0\). Speaking mathematically, they are not identifiable for \(K = 0\). The lack of identifiability under the null hypothesis destroys the usual asymptotic properties of the likelihood ratio (and \(\chi^2\)) tests (Dacunha-Castelle & Gassiat 1999). This is because typically a valid Taylor series of the signal model at \(K = 0\) cannot be constructed, except for rare special cases. Without that, the model cannot be linearized under the null hypothesis, which is critical for the validity of the asymptotic \(\chi^2\) distribution. An exception is provided by the LS periodogram with the harmonic model of the signal. In this special case, the following re-parametrization can be performed: \(K \cos(2\pi f t + \lambda) = a \cos(2\pi f t) + b \sin(2\pi f t)\). While the phase \(\lambda\) was not identifiable at \(K = 0\), the new parameters \(a\) and \(b\) are identifiable, and even linear. In this case, the distribution of each single value of \(2z(f)\) is indeed the \(\chi^2\) one with two degrees of freedom. Unfortunately, similar tricks are not possible for the majority of other models, even apparently simple ones.

Matters become even worse for the more practical case when \(f\) is also unknown. In this case, the frequency, treated as a new free parameter, is not identifiable (at \(K = 0\)) even for the LS periodogram. Actually, it can be noted that the non-identifiability of the frequency is the primary obstacle that has made the treatment of the significance levels of the LS periodogram so difficult and non-rigorous over decades. The LS periodogram of the noise contains an infinite sequence of similar noisy narrow peaks, but none of them can serve as a reference position for a quadratic Taylor approximation that would be valid in the whole frequency range. If it were not for that obstacle, we could just use the \(\chi^2\) distribution with three degrees of freedom (two for \(a\) and \(b\) plus one for \(f\)) to approximate the necessary distribution of the LS periodogram. However, Baluev (2008) managed to deal with this obstacle using the so-called ‘Rice method’, treating the noisy LS periodogram as a random process depending on a real argument \(f\), which is a single non-linear parameter of the model. The case of non-linear periodograms just adds more non-linear arguments of \(\zeta\), but the issue of their non-identifiability at \(K = 0\) remains qualitatively the same. Therefore, we can try to treat this more general situation using the same or a similar method.

Of course, it is hardly possible to derive an exact expression for FAP, but it would be good if we found at least an approximation analogous to what was obtained in previous works. Namely, the aim is to obtain something like
\[
\text{FAP}(\zeta_{\text{max}}) \lesssim M(\zeta_{\text{max}}),
\]
where \(M\) represents simultaneously an upper bound for FAP and its more or less good asymptotic approximation for large \(\zeta_{\text{max}}\) (small FAP).

### 3.2 Basic ideas of the Rice method

The modern comprehensive theory of the Rice method and relevant topics can be found in the reviews by Kratz (2006) and Azaïs & Wschebor (2009). Here we present only a very brief extract from the results that is most useful in our present paper. Suppose we deal with some arbitrary random process or field \(Z(x)\) and we need to find the probability that its maximum (within some domain \(x \in \Xi\)) will lie beyond a specified level \(Z(x) = z\). In our signal detection task this probability is equal to \(\text{FAP}(z)\), and this is obviously a complementary probability to the distribution function of the maximum of \(Z\). The general Rice method to estimate such probabilities is based on two main steps. First, we should construct some derived integer random variable \(N(z)\), such that the event \(N(z) = 0\) is equivalent (or almost equivalent) to the event \(\{Z(x) < z \forall x \in \Xi\}\), and the event \(N(z) \geq 1\) is (almost) equivalent to \(\{\exists x \in \Xi : Z(x) > z\}\). The boundary event when \(Z(x) \leq z\) everywhere in \(\Xi\) and there is at least one \(x\) such that \(Z(x) = z\) should usually correspond to \(N(z) = 1\). The term ‘almost’ refers here only to the effects at the boundary of \(\Xi\) (the boundary maxima); if somehow knew for sure that boundary maxima were impossible than this word could be omitted.

For random processes, a good choice for \(N\) is the number of up-crossings of the specified level; that is, the number of points \(x\) such that \(Z(x) = z\) and \(Z(x) > 0\). For random fields the term ‘up-crossing’ is meaningless, and in this case we choose \(N(z)\) to be the number of local maxima beyond \(z\) (and inside \(\Xi\)), that is, the number of points \(x\) where \(Z > z\), \(Z = 0\) and \(Z\) is negative-definite.

Given such a counter variable \(N\), we can estimate the required FAP, namely the probability for \(Z(x)\) to exceed a given level \(z\) somewhere in \(\Xi\), as
\[
\text{FAP}(z) \lesssim M(z) = M_{\text{boundary}}(z) + \tau(z).
\]
Here the term \(M_{\text{boundary}}\) refers to the maxima attained at the boundary of \(\Xi\); it may or may not be neglected, depending on other conditions of the task. We will discuss it in detail later. The primary term is \(\tau(z)\), which is equal to the mathematical expectation of the selected counter. This formula is basically the same as (16) with concretized function \(M(z)\).

The second proposition of the Rice method is the generalized Rice formula for \(\tau\). For random processes we have
\[
\tau(z) = \int_{\Xi} \mathbb{E}[(Z'(x))_+] | Z(x) = z | p_Z(z) \, dx,
\]
where the function \(p\) denotes the probability density function of the quantity \(Z(x)\) shown as a subscript (this \(p\) also depends on \(x\)), and \(|[\cdot]|_+ = \max(0, \cdot)|\). If necessary, the expression (18) can be rewritten in terms of the joint distribution of \(Z(x)\) and \(Z'(x)\) (see Baluev 2008). The terms ‘Rice method’ and ‘Rice formula’ are after Rice (1944), who constructed his original Rice formula for a stationary Gaussian random process.

In this paper we will use the generalized Rice formula for random fields. Actually, we now have three formulae of that type. The first one is introduced in Azaïs & Delmas (2002); it can be written as
\[
\tau(z) = \int_z^\infty dZ \int_{\Xi} \mathbb{E}([Z']_+ | Z' = 0; Z) | p_{Z,Z}(Z, 0) | dx.
\]
where \( \det(Z') \) is equal to \( \det(Z'') \) when \( Z' \) is negative-definite, and zero otherwise. This formula is generally similar to (18), although considerably more complicated. Azaïs & Wschebor (2009), in their chapter 8, introduced a variant of (19) with \( \det(Z') \) replaced by \( \det(Z'') \). Such replacement obviously increases the value of the right-hand side of (19), keeping its upper-limit property, but making the computations slightly easier. It counts \( \eta \) the critical points of \( Z(x) \) above \( z \), not just the local maxima. However, both these formulæ are usually too difficult for computations, and the formula that is typically used in practice contains just the ‘naked’ \( \det(Z'') \) instead of \( \det(Z') \) or \( \det(Z'') \). Such a formula gives the mathematical expectation of the Euler–Poincaré characteristic (EPC) of the level-section set \( \{ x \in X : Z(x) > z \} \) (which is also called the ‘excursion set’).

Unfortunately, the quantity \( \mathbb{E}(\text{EPC}) \) does not strictly retain the upper-limit property of \( \mathbb{E}(N) \). However, it is known [at least for Gaussian fields, see for example chapter 8 in Azaïs & Wschebor (2009)] that for large levels \( z \) the quantities \( \mathbb{E}(N) \) and \( \mathbb{E}(\text{EPC}) \) are asymptotically equivalent, and their difference decreases quickly (we will provide further details on this below). This is because beyond a large \( z \)-level all critical points of the random field are local maxima with almost unit probability, so the relevant excursion set represents a number of solid ellipsoids encompassing the positions of these maxima. Each such solid ellipsoid has \( \text{EPC} = 1 \), and thus \( \text{EPC} \approx N \) for large \( z \). All this means that we can typically use \( \mathbb{E}(\text{EPC}) \) as a good practical approximation for \( \tau \). Even if \( \mathbb{E}(\text{EPC}) \) does not provide an entirely strict upper bound, the relevant errors usually appear negligible for practical levels of \( z \). Of course we must admit that ‘usually’ and ‘typically’ are not the same as ‘always’, but nonetheless this approximation appears quite satisfactory in the examples considered below, as well as in a few other cases that are under preparation for a future publication.

The Rice method usually provides good practical accuracy, so that the mentioned upper bound (16) (applies close to the actual value of \( \mathbb{P}(\text{FAP}) \)) at least for practically important cases of small FAP levels. The Rice method is not widely known, because it is not mentioned in typical handbooks on mathematical statistics. Therefore, its usage in applications (e.g. in astronomy) is rare. However, rare does not mean absent: a variant of this method was applied by Bardeen et al. (1986) to study cosmological density fluctuations that were modelled by a Gaussian random field.

### 3.3 Applying the Rice method to non-linear periodograms

Mathematically, the condition \( \max \xi \leq z \) is equivalent to \( \max |\eta| \leq \sqrt{2\pi} \) (for arbitrary \( K \)) or \( \max \eta \leq \sqrt{2\pi} \) (for \( K \geq 0 \)). Therefore, we need to calculate the distribution of the maximum values of the random function \( \eta(\xi, f) \) to estimate the FAP. In this subsection, we limit ourselves to the single-sided case \( K \geq 0 \), bearing in mind that to obtain the formulæ for the case of arbitrary \( K \) we need to double the right-hand side of (17), because then we need to honour the maxima of \( \eta \) above \( \sqrt{2\pi} \) as well as its minima below \( -\sqrt{2\pi} \), which are entirely analogous.

The random field \( \eta(\xi, f) \) possesses quite simple statistical properties. From the definition (8) it clearly follows that if the noise in our observations is Gaussian, \( \eta \) represents a Gaussian random field. It is easy to check from (8) and (10) that \( \mathbb{E}\eta = 0 \) and the variance \( \mathbb{D}\eta = 1 \). This places us in the framework of theorem 1 in Azaïs & Delmas (2002). In this case, we can use (17) with

\[
\tau(z) \approx \mathbb{E}(\text{EPC}) = \sum_{j=0}^{[n/2]} a_j P_{n+1-2j}(z).
\]  

In this relationship, the integer \( n \) represents the number of free arguments of \( \eta \). It is equal to \( \dim \xi = d - 1 \) or \( \dim \xi + 1 = d \), depending on whether we consider the case of fixed or free frequency \( f \). Below we will consider only the more practical second case with \( n = d \), but in order to avoid misunderstandings it is preferable to keep different notations for \( n \) and \( d \). The notation \( [\cdot] \) denotes the integer part of the argument. The functions \( P_n(z) \) represent the tail probability associated with the \( \chi^2 \) distribution with \( k \) degrees of freedom:

\[
P_n(z) = \frac{1}{T(k/2)} \int_z^\infty x^{k/2-1} e^{-x} \, dx.
\]  

When \( z \to \infty \), the error of the approximation in (20) decreases more quickly than \( e^{\delta^2 + k/12} \) with some positive \( \delta \), while \( \tau \) itself typically decreases as \( e^{\zeta(\eta, n + 1/2)/2} \). This means that the relevant relative error decreases more quickly than \( e^{\delta \zeta(n + 1/2)/2} \).

We can see that the expression (17) involves a linear combination of \( \chi^2 \) tails with different numbers of degrees of freedom. However, the sum of the coefficients \( a_j \) is not necessarily one, so that the expansion in (17) is not a mixture of distributions in the rigorous meaning of this notion. For large \( z \), the first term with \( P_{n+1} \approx \zeta(n + 1/2)/2 \zeta^{-2} \) dominates, whereas the relative magnitudes of the remaining terms decrease as \( \sim 1/z^2 \).

The calculation of the coefficients \( a_j \) represents an appreciable technical difficulty. These coefficients are proportional to the quantities \( k_{2j} \) introduced in theorem 1 by Azaïs & Delmas (2002). We altered the original coefficients \( k_{2j} \) in order to have the functions \( P_{kr} \) explicitly in the sum (20).

Let us denote the variance–covariance matrix of the gradient of \( \eta \) as \( G \) (it is denoted as \( \Lambda \) by Azaïs & Delmas 2002). If it can easily be calculated from equation (8). The part of the matrix \( G \) corresponding to only the parameters \( \xi \) is equal to

\[
\frac{\partial \psi}{\partial \xi} \otimes \frac{\partial \psi}{\partial \xi}.
\]  

and the remaining elements due to the frequency parameter can be expressed in an entirely analogous manner.

First, let us consider a simplified situation when \( G \) does not depend on \( \xi \) and \( f \). Then we can use proposition ‘a’ of theorem 1 in Azaïs & Delmas (2002). We find that in this case the coefficients \( a_j \) (or \( k_{2j} \)) are proportional to the coefficients of the Hermite polynomials \( H_n \) (Korn & Korn 1968, section 21.7), so that after some elementary transformations we have

\[
\tau \approx AF_n(z)
\]  

with

\[
A = \frac{\sqrt{\det G}}{2\pi^{(n+1)/2}} \mathrm{Vol}(\mathbb{Z}) f_{\max}.
\]  

The last equality in (24) can just be checked by direct differentiation or it can be derived ‘honestly’ using the Rodrigues representation for \( H_n \). Notice that the \( H_n \) are normalized here so that their leading coefficients are equal to one and the weighting function is \( e^{-z^2/2} \).

Let us write down a few of the functions \( F_n \):

\[
F_1(z) = e^{-z}, \quad F_2(z) = e^{-z} \sqrt{z}, \quad F_3(z) = e^{-z} \left( z - \frac{1}{2} \right),
\]  

\[
F_4(z) = e^{-z} \left( z - \frac{3}{2} \right) \sqrt{z}, \quad F_5(z) = e^{-z} \left( z^2 - 3z + \frac{3}{4} \right).
\]
We plot the graphs of these functions in Fig. 1. Furthermore, a notable recursive relation
\[ F_{n+1}(z) = -F_n(z) \sqrt{\pi} \] can be used.

When \( G \) is non-constant, which is the more frequent case in practice, we can easily calculate only the primary term in (20). We have in this case
\[ \tau \approx Ae^{-z^2/(n-1)/2} \] (26)
with the relative error of \( \sim 1/z \), which is worse than the error of (20). Here
\[ A = \frac{1}{2\pi^{(n+1)/2}} \int_0^{\infty} df \int_{\xi} \sqrt{\det G} \frac{d\xi}{\xi} \] (27)
which is an evident generalization of \( A \) from (24).

The remaining \( z \)-power terms are different from the case of constant \( G \), and usually they are very hard to evaluate, because they involve conditional covariances of the second-order derivatives of \( \eta \) in quite unpleasant combinations (see proposition 'b' of theorem 1 by Azais & Delmas 2002). It can be noted that for the simple case \( n = 1 \) we have only one term in (20).

### 3.4 The role of the boundary effects

The quantity \( \tau \) by itself does not yet provide a closed solution of the problem. According to (17), we need to assess a similar term, which is related to the number of the so-called 'boundary maxima' (which are the local maxima of the random field restricted to the domain \( \mathcal{X} \) boundary).

This term takes into account the situation when all local maxima in the domain interior appear smaller than some boundary maximum. It can be calculated very similarly to the term \( \tau \); we just need to restrict our task to this boundary and apply the Rice method in a recursive manner. We should approximate the quantity \( M_{\text{boundary}} \) by a formula similar to (17) and (20), but with \( n \) replaced by \( n - 1 \). Hence, its relative contribution decreases for large \( z \), but at the rather slow rate of \( \sim 1/\sqrt{z} \).

We must, however, take care of one small but important issue. When we restrict our field \( \eta \) to the boundary of our domain, a maximum at the boundary does not necessarily represent a good candidate for the global maximum. Whether a particular boundary maximum is 'good' or 'bad' depends on the sign of the derivative of \( \eta \) in the direction of the outward normal to the boundary (i.e. the projection of the gradient \( \eta \) to this normal). If it is negative, we can definitely find larger values of \( \eta \) when stepping from the boundary inwards. Thus, such a boundary maximum can never provide the global maximum of \( \eta \), so we call it 'bad'. When counting the boundary maxima, we must filter out all 'bad' ones, keeping only the 'good', which offer a positive outward derivative. Mathematically, all this requires from us is to replace the gradient \( Z \) and the Hessian \( Z' \) in (19) by their projections to the tangent plane to the boundary (denote them, say, by \( Z_t \) and \( Z_t' \)), and to consider the operator \( \mathcal{E} \) and the probability density \( p_{Z_t Z_t'} \) to be subject to an additional condition \( Z_t > 0 \) (where the index \( \perp \) denotes the projection on the outward normal to the boundary surface).

In practice this usually just means that we need to halve (precisely or approximately) the estimated number of the boundary maxima, \( \tau_{\text{boundary}} \). It must be admitted that this issue has not been investigated in the literature with enough detail. Azais & Delmas (2002) prove this '1/2-rule' under certain restrictive assumptions, among which the most important is the requirement of constant \( G \). From their theorem 3 it follows, basically, that
\[ M_{\text{boundary}} = \frac{1}{2} \tau_{\text{boundary}} + \ldots. \] (28)

where \( \tau_{\text{boundary}} \) can be evaluated in essentially the same manner as \( \tau \), considering the restriction of the task to the boundary surface (which implies, in particular, a decrease in \( n \), and under '…' we mean here some terms having a faster decrease rate than \( \tau_{\text{boundary}} \). Although Azais & Delmas (2002) leave the case when \( G \neq \text{const} \) aside, investigation of their detailed proofs suggests that there is no obstacle to generalizing their single-term asymptotic formula for \( M_{\text{boundary}} \) to a more general case with \( G \neq \text{const} \). Moreover, we find that the main neglected term contained in '…' of (28) has the relative magnitude of \( \sim 1/z \).

Because in the general case we keep only the largest terms in \( \tau \) and similar quantities, we can use a two-term approximation such as
\[ \text{FAP}(z) \lesssim M(z) \simeq A z^{(n-1)/2} + \frac{1}{2} A_{\text{boundary}} z^{n/2-1} e^{-z}. \] (29)

The principal error of the right-most expression has a relative magnitude of \( \sim 1/z \) and results from the omitted terms in \( \tau \), while the omitted terms in \( \tau_{\text{boundary}} \) are of an even smaller order \( \sim 1/z^{1/2} \).

In more complicated cases, the boundary itself may be non-smooth owing to 'sub-boundaries' of smaller dimension (edges, vertexes), which will generate extra terms in \( M_{\text{boundary}} \). It is rather difficult to formulate a simple and general recipe of how to deal with them, because the geometry of the boundary might be quite complicated in general. Later, however, we explain this procedure with a concrete example of the von Mises periodogram, when the parametric domain is a rectangle.

### 4 EVALUATING THE COEFFICIENT \( A \)

#### 4.1 Assuming uniform phase coverage

It would be useful to construct analytic expressions of the FAP for the case when our observations are distributed approximately uniformly in time. We assume that the timings \( t_i \), when they are phased to some frequency \( f \), cover the relevant phase more or less uniformly. In this case, we can approximate the summation \( \langle \omega \rangle \) over the time series by means of integration over its time span \([t_1, t_2]\). Putting it more accurately, we approximate the time series average \( \langle \omega \rangle / \langle 1 \rangle \) by the integral average over the time span. Moreover, the periodic character of the model \( \mu \) usually allows for this integration to be performed over a single period only. The periodicity of \( h \) implies that \( h(t, \xi, f) = g(2\pi ft + \lambda, \psi) \), where the function \( g(x, \psi) \) is \( 2\pi \)-periodic in \( x \), and \( \lambda \) is the phase parameter. The remaining \( (d-2) \) parameters form the vector \( \psi \). This means that \( \Xi = [0, 2\pi] \times \mathcal{Y} \) and \( \Theta = [0, +\infty) \times [0, 2\pi] \times \mathcal{Y} \), where \( \mathcal{Y} \)}
is some \((d - 2)\)-dimensional domain of parameters \(\nu\). Now we can approximate, for instance, the mean value of \(h\) as
\[
\langle h \rangle \approx \frac{w}{P} \int_0^P h(t, \xi, f) \, dt \approx \frac{w}{2\pi} \int_{-\pi}^{\pi} g(x, \nu) \, dx = w \overline{g},
\]  
for example let us write down the derivatives of \(\psi\) as
\[
\frac{\partial \psi}{\partial f} \approx \frac{2\pi g'_{\nu} \sqrt{w_{g^2}}}{g}, \quad \frac{\partial \psi}{\partial \lambda} \approx \frac{2\pi g'_{\nu} \sqrt{w_{g^2}}}{g}, \quad \frac{\partial \psi}{\partial \nu} \approx \frac{g'_{\nu} \sqrt{w_{g^2}}}{g},
\]
where the function \(g\) and its derivatives \(g'_{\nu} = \frac{\partial g}{\partial \nu} \) and \(g'_{\lambda} = \frac{\partial g}{\partial \lambda} \) outside the averaging are calculated for \(x = 2\pi ft + \lambda\). This allows us to calculate the elements of the matrix \(G\):
\[
\left< \left( \frac{\partial \psi}{\partial f} \right)^2 \right> \approx q = \frac{\pi^2}{g^2},
\]
\[
\left< \frac{\partial \psi}{\partial \lambda} \cdot \frac{\partial \psi}{\partial \nu} \right> \approx v_{ij} = \frac{\pi^2}{g^2} - \frac{\pi^2}{g^2},
\]
\[
\left< \frac{\partial \psi}{\partial \nu} \cdot \frac{\partial \psi}{\partial \nu} \right> \approx v_{ij} = \frac{\pi^2}{g^2} - \frac{\pi^2}{g^2},
\]
which do not depend on the frequency and phase. When calculating the elements of the matrix \(G\), we also deal with summations such as, for instance,
\[
\left< \left( \frac{\partial \psi}{\partial f} \right)^2 \right> \approx 4\pi^2 \left( \sqrt{w_{g^2}} \right)^2 (2\pi ft + \lambda),
\]
The line over \(\int\) denotes the weighted averaging of the squared timings: \(\overline{t^2} = \langle t^2 \rangle / \langle t \rangle\), which can be easily evaluated directly (without approximating it by a continuous integral). Similar arguments lead to
\[
\left< \frac{\partial \psi}{\partial f} \cdot \frac{\partial \psi}{\partial \nu} \right> \approx 2\pi q, \quad \left< \frac{\partial \psi}{\partial \lambda} \cdot \frac{\partial \psi}{\partial \nu} \right> \approx 2\pi iq v_i,
\]
where \(\overline{t}\) is the weighted average of \(t\). Thus, the full matrix \(G\) can be written in the following block form:
\[
G \approx \begin{pmatrix}
4\pi^2 \overline{t^2} q & 2\pi q & 2\pi \nu^T \\
2\pi q & q & v^T \\
2\pi \nu & v & V
\end{pmatrix},
\]
where the elements of the vector \(v\) and those of the matrix \(V\) are defined in (32). In this approximation, the matrix \(G\) depends only on the parameters \(\nu\). Using simple elementary transformations of \(G\) we can finally obtain that
\[
\text{det} G \approx \pi q^2 T_{\text{eff}}^2 \text{det} R,
\]
where \(R = V - v \otimes v/q\) (that is, \(R_{ij} = V_{ij} - v_i v_j/q\)) and \(T_{\text{eff}} = \sqrt{4\pi (\overline{t^2} - 1)}\) is the effective length of the time series, as defined in Balavev (2008). Integrating (27) over \(\lambda\) and substituting (37), we can write:
\[
A \approx \frac{W}{q_{\nu^2}/2 - 1} \int_{\tau} q \sqrt{\text{det} R} \, dv, \quad W = f_{\text{max}} T_{\text{eff}}.
\]

The factor \(A\) can now be substituted in (23) for use in (17) and (16). In the degenerate case \(n = 2\), we have \(\dim \nu = 0\) and put \(\text{det} R = 1\) by definition.

The main advantage of this method of calculation of \(A\) is that the result depends on a particular time series only via the single quantity \(T_{\text{eff}}\). Given the model \(g\) and the domain \(\lambda\), we can evaluate an approximation for \(A\) once and then use it for all time series, substituting the proper values of \(f_{\text{max}}\) and \(T_{\text{eff}}\). For the LS periodogram, for instance, we have \(A \approx W\), which is in full agreement with Balavev (2008). The main disadvantage is that this approximation may have insufficient accuracy in practice.

When deriving this approximation for \(A\), we assumed uniform phase coverage for all frequencies \(f\) in the scan range. When the original time series is not uniform, this assumption may become invalid at some frequencies \(f\), corresponding to periodic leakage patterns of \(t_i\) (and including the zero frequency). However, these perturbations usually appear only inside very short frequency segments (\(\Delta f \sim 1/T_{\text{eff}}\)) around the leakage frequencies; for most values of \(f\) in the range \([0, f_{\text{max}}]\) the phase coverage is still uniform. However, our formula for \(A\) in (27) involves an integration over the wide frequency band \(\Delta f \sim W/T_{\text{eff}}\), with \(W \gg 1\), and hence the perturbations of its integrand have almost no effect on the result, because they are limited by much smaller segments with \(\Delta f \sim 1/T_{\text{eff}}\). This means that the accuracy of our approximation for \(A\) does not significantly degrade even for non-uniform time series. For sinusoidal signals this was demonstrated in Balavev (2008), where it was shown that even ultimately strong aliasing has only a negligible effect on the resulting A, when \(W \gtrsim 10\).

In the case of non-sinusoidal signals, an important source of errors of the approximation (38) comes from elsewhere. If the signal model contains some quickly varying structures, for example narrow peaks, the observations may cover these structures with insufficient sampling, so the resulting approximation for \(A\) may appear poor even when the \(t_i\) are perfectly uniform. This effect is important for the von Mises periodogram below, for example.

### 4.2 Evaluating \(A\) directly

The direct evaluation of the factor \(A\) by means of substitution of (22) to (27) involves some rather unpleasant manipulations with huge formulae, especially when one is working in the general terms of Section 2. To simplify them, let us write down the gradient of the model \(h\) over the compound vector of all non-linear parameters \(\omega = \{f, \lambda, \nu\}:
\[
\gamma = \frac{\partial h}{\partial \omega} = \left\{ 2\pi t g'_{\nu}, g'_{\nu}, g'_{\nu} \right\}.
\]

Then define
\[
Q_{\gamma_{\omega}, \lambda} = \langle \phi_{\gamma} \otimes \phi_{\nu} \rangle.
\]
\( Q_{\theta, K} = \langle \varphi_\gamma g \rangle \),
\( Q_{\theta, \omega, \omega} = \langle \varphi_\gamma \otimes \varphi_\gamma \rangle \),
\( T = \langle y \otimes y \rangle - Q_{\theta, \omega, \omega}^{-1} Q_{\theta, \gamma, \gamma} Q_{\theta, \omega, \omega}^{-1} \),
\( y = \langle g y \rangle - Q_{\theta, \omega, \omega}^{-1} Q_{\theta, \gamma, \gamma} Q_{\theta, \omega, \omega}^{-1} \),
\( D = \langle y^2 \rangle - Q_{\theta, \gamma, \gamma}^2 Q_{\theta, \omega, \omega}^{-1} Q_{\theta, \omega, \omega}^{-1} \).  
\( (40) \)

where \( \varphi_\gamma \) is the functional base of the linear null model (3). Finally,
\( G = \frac{T - y \otimes y}{D^2}. \)  
\( (41) \)

We can also offer another evaluation sequence. Let us construct the full Fisher information matrix of all the parameters involved \((\theta, K, \omega)\):
\( Q = \begin{pmatrix} \varphi_\gamma \otimes \varphi_\gamma & \varphi_\gamma g & \varphi_\gamma \otimes y \\ g \varphi_\gamma g & g^2 & g y^T \\ y \otimes \varphi_\gamma & y g & y \otimes y \end{pmatrix}. \)  
\( (42) \)

Please note that the triangle braces, denoting the weighted sum-
over the time series, are still here. Now let us apply the
Cholesky decomposition: \( Q = L L^T \) with \( L \) being a lower-triangular
matrix. Then write down \( L \) in the same block form as \( Q \) in (42):
\( L = \begin{pmatrix} L_{\theta, \theta, \theta} & 0 & 0 \\ L_{\theta, \theta, K} & I_{K, K} & 0 \\ L_{\omega, \omega, \omega} & L_{\omega, K} & L_{\omega, \omega} \end{pmatrix}. \)  
\( (43) \)

Note that \( L_{\theta, \theta, \theta} \) and \( L_{\omega, \omega, \omega} \) are vectors, and \( L_{K, K} \) is a scalar. Obviously, \( L_{\theta, \theta, \theta} \) is a lower-triangular matrix of the Cholesky decomposition for \( Q_{\theta, \theta, \theta} \), and \( L_{\omega, \omega, \omega} \) is another lower-triangular matrix. From the matrix \( L \) definition and from (41) we can easily derive two remark-
able relations:
\( G = \begin{pmatrix} L_{\omega, \omega} L_{\omega, \omega}^T \\ L_{K, K} \end{pmatrix}, \)  
\( D = I_{K, K}^2. \)  
\( (44) \)

Moreover, it is clear that \( \sqrt{\text{det} G} = \text{det} L_{\omega, \omega} / I_{K, K}^2 \). Therefore, to
find the integrand in (27) we just need to calculate \( \text{det} L_{\omega, \omega} \), which
is simply equal to the product of its diagonal elements, and then
divide the result by \( I_{K, K}^2 \).

Therefore, the final procedure to evaluate \( \sqrt{\text{det} G} \) is as follows.

(i) Evaluate \( Q \) using its definition (42). Its size should be \( d_\gamma + d_\omega + 1 \). Note that it is important to preserve the ordering of the parameters as \( \theta, K, \omega \), while the ordering inside \( \theta, K \) and inside \( \omega \)
is not important.

(ii) Perform the Cholesky decomposition of \( Q \). It is a very quick and numerically stable procedure.

(iii) On the basis of only the diagonal elements of the resulting Cholesky matrix \( L \), construct the combination
\( \left( \prod_{i=0}^{d_\gamma+d_\omega+1} I_{i} \right) / (d_{\gamma}+d_{\omega}+1)! \). It is equal to what we seek.

The quantity \( \sqrt{\text{det} G} \) must be further numerically integrated over the parameters \( \omega \), according to (27). Note that this includes the integration over the frequency \( f \) and over the phase \( \lambda \) (which is now non-trivial).

Now, let us limit ourselves to the most important practical case
when the null model involves only a free constant term: \( d_\gamma = 1 \), \( \varphi_\gamma = 1 \). In this case the matrix \( Q \) is considerably simplified:
\( Q = \begin{pmatrix} 1 & g & y^T \\ g & g^2 & g y^T \\ y g & y \otimes y \end{pmatrix}. \)  
\( (45) \)

The round-off errors may destroy the positive-definiteness of \( Q \),
which is critical for the Cholesky decomposition. To reduce this
effect, the computational sequence can be transformed to something
similar to the formulae (13) and (14). This will be some hybrid
approach to evaluate \( \sqrt{\text{det} G} \) between the two approaches that we
have already discussed. Namely, first we should centre the functions
involved:
\( g_c = \langle g \rangle / (1), \)  
\( \gamma_c = \langle \gamma \rangle / (1). \)  
\( (46) \)

After that, we need to evaluate
\( D = \langle \tilde{g}^2 \rangle, \)  
\( y = \langle \tilde{g} \gamma \rangle, \)  
\( \beta = \tilde{\gamma} - \tilde{g} \gamma / D. \)  
\( (47) \)
which imply that
\( G = \langle \beta \otimes \beta \rangle / D. \)  
\( (48) \)

We have no need to evaluate \( G \) itself. Instead, we can perform the
Cholesky decomposition of \( \langle \beta \otimes \beta \rangle \). Dividing the product of the
diagonal elements of the resulting Cholesky matrix by \( D^{1/2} \), we
obtain \( \sqrt{\text{det} G} \).

The obvious advantage of the direct method to evaluate the factor
\( A \) is that it allows for high accuracy limited only by round-off
and numerical integration errors, not relying on any approximating
assumptions. The disadvantage is that it is considerably slower than
in Section 4.1, although in practice the relevant computation time
should be comparable to the time of a single evaluation of the
periodogram itself.\(^2\) Therefore, this is still much faster than, for
example, Monte Carlo simulation, in which this periodogram has to
be re-evaluated thousands of times before a reliable FAP estimation
is reached.

5 **UNKNOWN NOISE LEVEL**

In Section 2, we assumed that the standard errors \( \sigma_i \) of observations
are known a priori. In practice we often do not know them with
enough precision. A commonly used model is given by \( \sigma_i^2 = \kappa / w_i \),
where the quantities \( w_i \) determine the weighting pattern of the time
series and the factor \( \kappa \) remains unconstrained a priori. A similar
problem was considered in Baluev (2009a). In this work, the model
\( \sigma_i^2 = \sigma^2_{\text{meas}} + \sigma^2_{\text{jitter}} \) was considered with \( \sigma^2_{\text{meas}} \), being the ‘internal’
measurement variances, known a priori, and the parameter \( \sigma^2_{\text{jitter}} \) being
the unconstrained variance of the extra ‘jitter’. In these cases, we
cannot calculate the least-squares periodogram \( \chi(f) \), as we cannot
calculate the values of the \( \chi^2 \) functions themselves.

\(^2\) We will typically evaluate this periodogram on some multidimensional
grid in the space of all the parameters \( \omega \), not just on a frequency grid as
in the LS case. In terms of the computational demands, this procedure is
roughly equivalent to the numerical integration over the same space.
The general approach for solving such problems is based on
the likelihood ratio test. The logarithm of the likelihood function
for our observations, which are contaminated by random mutually
independent Gaussian errors, may be written (specifically for the
hypothesis $\mathcal{H}$ and $\mathcal{K}$) as

$$\ln \mathcal{L}_{\mathcal{H}, \mathcal{K}} = \left( \frac{\sum_{i=1}^{N} \left( x_i - \mu_{\mathcal{H}, \mathcal{K}}(t_i) \right)^2}{\sigma^2(p)} + \ln \sigma^2(p) \right) + \text{const}. \quad (49)$$

Here the full variances $\sigma^2$ depend on the extra parameters $p$, which
should be estimated from the data together with the usual parameters
$(\theta_{\mathcal{H}}, \theta_{\mathcal{K}}, f)$ of the model curve. These estimations are obtained
by maximizing the corresponding likelihood function over all of the
parameters to be estimated. After that, we could construct the
logarithm of the likelihood ratio statistic as the maximum (over the
frequency $f$) of the likelihood ratio periodogram

$$Z(f) = \max_{p, \theta_{\mathcal{K}}} \ln \mathcal{L}_{\mathcal{K}}(p, \theta_{\mathcal{K}}, f) - \max_{p, \theta_{\mathcal{H}}} \ln \mathcal{L}_{\mathcal{H}}(p, \theta_{\mathcal{H}}). \quad (50)$$

This function may give the basis for signal detection in the general
framework, when $p$ is not known a priori. However, in order to
reduce the statistical bias in $p$, it is better to use the following
modifications of the likelihood functions and of the likelihood ratio
periodogram:

$$\ln \tilde{\mathcal{L}}_{\mathcal{H}, \mathcal{K}} = \left( \frac{\sum_{i=1}^{N} \left( x_i - \mu_{\mathcal{H}, \mathcal{K}}(t_i) \right)^2}{\sigma^2(p)} + \ln \sigma^2(p) \right),$$

$$Z(f) = \frac{1}{2} N \left[ \max_{p, \theta_{\mathcal{K}}} \sum_{i=1}^{N} \left( x_i - \mu_{\mathcal{H}, \mathcal{K}}(t_i) \right)^2 \right] + \ln \sigma^2(p).$$

This approximation does indeed rapidly converge to those obtained for the original least-squares periodogram. This provides an independent confirmation of the arguments from the last paragraph. Therefore, for large data
sets, we can apply the analytic estimation of the FAP for the periodogram $z(f)$ and other non-linear likelihood ratio periodograms in
the same way as described in the previous sections for least-squares
periodograms with known noise uncertainties.

### 6 PRACTICAL EXAMPLES OF NON-LINEAR PERIODOGRAMS

#### 6.1 Detecting periodic signal with a fixed non-sinusoidal shape

Let us consider the case $d = 2$ with $\theta$ incorporating only the amplitude
and phase of the periodic signal to be detected. That is,

$$\mu = Kg(2\pi f t + \lambda), \quad (53)$$

where the $2\pi$-periodic function $g(x)$ is given a priori and is centred
so that $\bar{x} = 0$. This function determines the shape of the putative
periodic variation.

Note that the term $M_{\text{boundary}}$ in (17) may now only appear because
of the boundary points of the frequency segment. The phase $\lambda$
is a periodic parameter defined over the self-closed circle, which
basically does not have a boundary. In other words, all maxima of
$\eta$ over $\lambda$ are local maxima where $\partial \eta / \partial \lambda = 0$; no other maxima
are possible in $\lambda$.

In this simple case we express FAP separately for the fixed-frequency and unknown-frequency cases, using the approach of
Section 4.1. In the fixed-frequency case, we find

$$\text{FAP}_{\text{single}} \lesssim M_{\text{single}}(z) \approx \sqrt{q \varepsilon^2}.$$

In the case of unknown frequency,

$$\text{FAP}_{\text{max}} \lesssim M_{\text{max}}(z) + M_{\text{single}}(z) \approx q W \varepsilon^{-2} \sqrt{\varepsilon} + \sqrt{q \varepsilon^2}.$$  

The term $M_{\text{boundary}}$ is equal to $M_{\text{single}}$ here, as we have two end-points of the segment $[0, f_{\text{max}}]$ and each should be counted as half.

This term can be safely neglected in (55) anyway.

Note that as long as the approximation of the uniform phase
coverage is valid, the matrix $G$ appears here almost constant, so that
we can use the simplified formulae (24). This means that although
on the right-hand side of (55) we omitted a term of the order of
e$^{-2} / \sqrt{\varepsilon}$, corresponding to the term with $a_1$ of (20), the coefficient
$a_1$ is itself negligibly small. As discussed in Baluev (2008) for the
sinusoidal model, the approximation of the uniform phase coverage
works well even for time series with ultimately strong spectral
leakage. This is because the aliasing/leakage-induced errors are
concentrated within a few narrow frequency segments, and they
thus have only a negligible effect on the quantities expressed by
an integral over a large frequency band (like $A$). However, for non-
sinusoidal signals this approximation may appear poor for reasons
unrelated to the spectral leakage effects; that is, it may be poor when
$g(x)$ demonstrates narrow peaks that our observations are unable to
cover with dense enough sampling.

In the general case $q \geq 1$. This inequality can be clearly de-
uced by applying the Parseval identity to the Fourier series for $g(x)$ and $g'(x)$. When $g$ is a harmonic function, we deal with the
LS periodogram. In this case, $g^2 = g'^2 = 1/2$, and $q = 1$. This

---

3 This is because there is a 50/50 chance that such a boundary value is
actually a boundary minimum rather than a maximum, depending on
the sign of the derivative at this end-point.
allows us to entirely reproduce the exponential single-value distribution of the LS periodogram, FAP = e−z, and the Davies bound FAP max ≤ We−z/√2 + e−z from Ballev (2008). When g(x) contains at least two Fourier terms we have q > 1, so the minimum q = 1 is attained for the sinusoidal and only sinusoidal variation.

As a non-trivial example, let us consider the case when g(x) has a sawtooth shape: during the first half of its period it decreases linearly from 1 to −1 and during the second half it increases linearly from −1 to 1. In this case, g2 = 4/π2, g2 = 1/3, and q = 12/π2 ≈ 1.216.

6.2 The von Mises periodogram

Let us assume the following non-linear model for the periodic signal:

\[ g(x, v) = \exp(v \cos x) - I_0(v), \quad v > 0. \]  (56)

In this definition I0 denotes the modified Bessel function; we need it to satisfy the condition \( \mathcal{F} = 0 \). It can be seen that this function is 2π-periodic in x. At \( v = 0 \) we have a formal singularity because g(x, 0) ≡ 0 and (10) becomes degenerate. We can easily remove this degeneracy by making a replacement

\[ \tilde{g}(x, v) = \frac{g(x, v)}{v} = \cos x + \frac{v}{2} \left( \cos^2 x - \frac{1}{2} \right) + O(v^3), \]  (57)

so that for \( v \to 0 \) our model becomes equivalent to a simple sinusoid. For large v the function (56) represents a comb-like sequence of periodic narrow peaks, each having width \( \sim 1/\sqrt{v} \). Note that a very similar function with a slightly different scaling, \( \exp(v \cos x)/I_0(v) \), represents a probability density function of the so-called von Mises distribution. This distribution is a periodic analogue of the Gaussian one, possessing a similar maximum-entropy property on a circle. For the sake of convenience we will also call (56) the von Mises function, as these small differences in the centring and normalization are not very important for us.

In Fig. 2, we plot the von Mises function for several values of the localization parameter v. Looking at these plots, it can be seen that such a shape may provide a satisfactory generic approximation to many physical variations that emerge in astronomical practice. In particular, it may work well for the light curves of many variable stars and planetary transits (just turn these graphs upsidedown to obtain something similar to a transit). Because such a model is functionally very simple (and thus quickly calculable and easy to manipulate), it is tempting to construct a periodogram that could utilize it as a model of the probe periodic signal.

Assume that we scan this periodogram in a rectangle \( v_{\text{min}} \leq v \leq v_{\text{max}} \) and \( 0 \leq f \leq f_{\text{max}} \), and do not allow the signal amplitude \( K \) to become negative. Then we should first evaluate the function t using (20) for the interior of this rectangle (implying \( n = 3 \)). It will be approximately proportional to W. Furthermore, we should evaluate the function \( M_{\text{boundary}} \), which contains the term responsible for the four sides and four vertices of the above-mentioned rectangle. Among these, only the terms because of the sides \( v = v_{\text{min}}, v_{\text{max}} \) are important. This is because these are the only boundary terms proportional to W. The boundary terms due to the two other sides and due to the vertices do not contain this multiplier, because the frequency f is held fixed there. Because in practice W is large or very large, we only need to take into account the boundary edges running along the frequency axis.

The final approximation to the FAP can be represented in the form

\[ \text{FAP}(z) \lesssim M(z) = We^{-z} \left[ (X(v_{\text{max}}) - X(v_{\text{min}}))z + (Y(v_{\text{min}}) + Y(v_{\text{max}}))\frac{\sqrt{2}}{2} + O(z^0) \right], \quad (58) \]

where

\[ X(v) = \frac{1}{2W\pi^2} \int_{0}^{f_{\text{max}}} df \int_{0}^{v_{\text{max}}} dv \int_{0}^{2\pi} \sqrt{\det G_{\mu\nu}(f, \lambda, v)} d\lambda, \]

\[ Y(v) = \frac{1}{2W\pi^2} \int_{0}^{f_{\text{max}}} df \int_{0}^{v_{\text{max}}} dv \int_{0}^{2\pi} \sqrt{\det G_{\mu\nu}(f, \lambda, v)} d\lambda. \]  (59)

In (58), the terms involving the factor X correspond to the local maxima of \( \eta \) in the interior of the rectangle. The difference \( X(v_{\text{max}}) - X(v_{\text{min}}) \) is because the factor A now contains an integral from \( v_{\text{min}} \) to \( v_{\text{max}} \), while the function X(v) is defined as an integral from 0 to v. The terms with Y are for maxima on the two boundary lines \( v = v_{\text{min}} \) and \( v = v_{\text{max}} \). The sum \( Y(v_{\text{min}}) + Y(v_{\text{max}}) \) is because we need to sum the maxima at both borders, and the extra multiplier of 1/2 is because we should filter out half of these boundary maxima, owing to the derivative \( \partial_f \eta/\partial v \) having at these maxima an inappropriate sign with probability 1/2. The quantities WX(v) and WY(v) represent, in fact, the factor A for the rectangle \([0, v] \times [0, f_{\text{max}}]\) and for the boundary segments \( \{ v = v_{\text{min}}, v_{\text{max}} \} \times [0, f_{\text{max}}] \).

In (59), the 3 × 3 matrix \( G_{\mu\nu} \) corresponds to the full gradient of \( \eta \) over all three non-linear parameters \( f, \lambda, v \); the matrix \( G_{\mu\nu} \) is a 2 × 2 submatrix of \( G_{\mu\nu} \) involving only the elements corresponding to \( f \) and \( \lambda \). Both these matrices depend on all three parameters. Note that the terms in (58) containing Y basically correspond to the signal having a fixed non-sinusoidal shape (v is fixed), and thus they can also be treated using the formalism of Section 6.1.

In the particular case \( v_{\text{min}} = 0 \) (with the sign of K still fixed) we should take into account the obvious equality \( X(0) = 0 \):

\[ \text{FAP}(z) \lesssim M(z) = We^{-z} \left[ zX(v_{\text{max}}) \right. \]

\[ + \left. (Y(v_{\text{min}}) + Y(v_{\text{max}}))\frac{\sqrt{2}}{2} + O(z^0) \right]. \]  (60)

Note that in practice \( Y(0) \approx 1 \) with good precision (see below).

When K is allowed to be positive as well as negative we should double the right-hand side of the expression in (58). This is because we should now honour the local minima of the random field \( \eta \), as well as its local maxima. We have in this
maximum of $|\eta|$ cannot be attained at the line $\nu = 0$. Following the terminology of Azaïs & Delmas (2002) (see their theorem 2), we have shown that the field $\zeta$ (or $|\eta|$) is a ‘field without boundary’, concerning the boundary line $\nu = 0$. This allows us to just drop the relevant boundary term with $Y(\nu_{\text{min}} = 0)$:

$$\text{FAP}(z) \lesssim M(z) = \mathcal{W} e^{-z^2} \left[ 2z X(\nu_{\text{max}}) + Y(\nu_{\text{max}}) \sqrt{z^2 + O(z^5)} \right].$$  

(62)

Note that it is essential here that $K$ has arbitrary sign, because otherwise we could not freely swap the values $\eta > 0$ with $\eta < 0$.

Let us first apply the approach of Section 4.1 to evaluate $X$ and $Y$. The derivatives of the function (56) look like

$$g_c^\nu = -\nu \exp(\nu x) \sin x,$$

$$g_s^\nu = \exp(\nu x) \cos x - I_1(\nu).$$  

(63)

Substituting them in (32), and using the well-known integral representations for the modified Bessel functions $I_k$, we obtain

$$\begin{align*}
\overline{g} &= I_0(2\nu) - I_2^2(\nu), \\
\overline{g_s^\nu} &= \frac{1}{2} \left[ I_0(2\nu) + I_2(2\nu) \right] - I_1^2(\nu), \\
\overline{g_c^\nu} &= I_1(2\nu) - I_0(\nu) I_1(\nu),
\end{align*}$$

and then

$$q = \frac{\nu^2}{2} \frac{I_0(2\nu) - I_2(2\nu)}{I_0(\nu) - I_2(\nu)}, \quad v = 0,$$

$$\det \mathbf{R} = V = \frac{I_0(2\nu) + I_2(2\nu) - 2I_1^2(\nu)}{2[I_0(2\nu) - I_2(\nu)]} - \left[ \frac{I_1(2\nu) - I_0(\nu) I_1(\nu)}{I_0(2\nu) - I_2(\nu)} \right]^2.$$  

(65)

The behaviour of these quantities is not obvious from these formulae, so we need to understand their asymptotics. For small $v$ we can use the Taylor expansion of $I_k(z)$ to find that $q(0) = 1$ and $V(0) = 1/16$. This implies that for the factor $X$ the integrand $q \sqrt{\det \mathbf{R}}$ is equal to $1/4$ at $v = 0$; for the factor $Y$ we have $\det \mathbf{R} = 1$ by definition, and $q \sqrt{\det \mathbf{R}}$ at $v = 0$ is equal to one. For large $v$ we can use the following asymptotically converging expansion:

$$I_k(z) \sim -\frac{e^{-z}}{\sqrt{2\pi z}} \left( 1 - \frac{k^2}{8z} + \frac{(4k^2 - 1)(4k^2 - 9)}{2(8k^2)^2} + \ldots \right),$$  

(66)

which can be found in, for example, section 21.8 of Korn & Korn (1968). The calculations lead to

$$q \sim \frac{v}{2}, \quad V \sim \frac{1}{8v^2},$$

$$q \sqrt{\det \mathbf{R}} \sim \frac{1}{4\sqrt{2}} \quad (\text{for } X), \quad q \sqrt{\det \mathbf{R}} \sim \frac{v}{2} \quad (\text{for } Y).$$  

(67)

When calculating $X$, we should further integrate the relevant quantity over $v$, so we obtain $X(v) \sim v$ for large $v$. The factor $Y$ does not need this integration, but its asymptotics appears eventually the same as for $X$. $Y(v) \sim v$ for large $v$. Furthermore, for large $v$ we have $Y(v)/X(v) \simeq 2/\sqrt{2\pi}$, and hence the ‘primary’ $X$-term in (58) exceeds the $Y$-term only for $z > \pi v$, and even beyond this level they remain mutually comparable up to large $z$. This means that both these terms should be taken into account in practice: neither can be neglected.

These results infer that the integral in (38) is infinite if we do not limit $v$ from the upper side. In this regard, the parameter $v$ is

Figure 3. Coefficients $X$ and $Y$ for the von Mises periodogram. The dashed curve is for the approximation of Section 4.1, while the three solid curves correspond to the direct precise method of Section 4.2. These three curves (from bottom to top in each panel) correspond to three simulated time series, containing $N = 30, 100$ and $1000$ randomly distributed simulated observations, For all cases we have $W \approx 100$. case

$$\text{FAP}(z) \lesssim M(z) = \mathcal{W} e^{-z^2} \left[ 2z X(\nu_{\text{max}}) + Y(\nu_{\text{max}}) \sqrt{z^2 + O(z^5)} \right].$$  

(61)

A special case occurs when $\nu_{\text{min}} = 0$ and the sign of $K$ is arbitrary. Then we have, basically, some degeneracy of the free variables at $v = 0$, making the cases $K < 0$ and $K > 0$ equivalent to each other (owing to the symmetry of the sinusoid). This property can be used to refine the Rice bound slightly. Assume that we have some point at the boundary $v = 0$, such that $\eta > 0$ (implying $K > 0$) and $\partial \eta / \partial v < 0$. It is easy to show that at a dual point with $\lambda \mapsto \lambda + \pi$ the value of $\eta$ changes the sign (hence $K < 0$), but the value of $\partial \eta / \partial v$ remains exactly the same. This is because the derivative of $\tilde{g}$ in (57) over $v$ is a $\pi$-periodic function of $x$ (at $v = 0$), while the model $\tilde{g}$ itself is $2\pi$-periodic. Therefore, the derivative $\partial \eta / \partial v$ has a different sign at these dual points, while the value of $|\eta|$ is identical.

Therefore, although there are many ‘good’ boundary maxima at $v = 0$, which satisfy the condition $|\eta| / \partial v < 0$ (meaning that $|\eta|$ necessarily decreases when we step from $v = 0$ inwards), each such maximum has a dual ‘bad’ maximum at $\lambda + \pi$, where $|\partial \eta / \partial v > 0$. Because there are no constraints on $\lambda$, this means that for any boundary maximum at $v = 0$, either ‘good’ or ‘bad’, we can find larger values of $\eta$ in the interior $v > 0$. This implies that the global
Figure 4. FAP approximation of the von Mises periodogram compared with Monte Carlo simulations. In each panel we show the simulated and analytic FAP curves for $v_{\text{min}} = 0$ and six values of $v_{\text{max}}$ from 0 to $\sim 315$. In the graphs we mark, instead of the upper limit for $v$, a lower limit for a more intuitive FWHM (full width at half maximum) characteristic of the signal peaks. It can be easily mapped one-to-one with $v_{\text{max}}$ and it varies here from 1/2 (corresponding to the sinusoidal variation) to $\sim 1/50$ (typical for, for example, a planetary transit). The panels to the left correspond to the cases with the noise uncertainties $\sigma_i$ known a priori; the ones to the right are for the multiplicative noise model of Section 5, $\sigma_i^2 \propto 1/w_i$. In the case of ‘clumped timings’ (right-bottom panel) the $N = 100$ points of the time series were equally split into 10 equidistant groups with 90 per cent gaps between them. This implies a very strong aliasing, which makes our analytic approximations for FAP relatively inaccurate, although they still work as an upper limit.

We find that the approximate expressions for the factors $X$ and $Y$ obtained using the formalism of Section 4.1 are very accurate for small $v$, but this accuracy decreases when $v$ grows, and increases when $N$ grows (Fig. 3). We assume that the error of this approximation emerges because our $N$ observations cannot sample well the narrow peaks of the signal having width $\sim 1/\sqrt{v}$.

We have performed some Monte Carlo simulations to test the quality of the approximation (60). The results are shown in Fig. 4. We find that our analytic approximations behave exactly as we might expect. They have good accuracy for practically important small levels of the FAP and for not too large $v_{\text{max}}$ and time series leakage. For larger $v_{\text{max}}$ or for strong leakage, the accuracy degrades somewhat, but the formula (60) still works as an upper bound on FAP. Our conclusion is that this formula would certainly be useful in practical applications.

Finally, we would like to demonstrate the power of the von Mises periodogram itself. We generated a simulated time series with $N = 1000$ randomly spaced observations. The values of the simulated measurements contained two periodic signals with comparable amplitudes: a sinusoidal variation and periodic flat drops (simulating planetary transits). Both signals were below the noise level, so the noise should provide significant contamination. The von Mises periodogram of these data is plotted in Fig. 5. We can see that it allows the easy detection of both signals at once ($f \approx 8$ and $f \approx 56$), while the LS periodogram (which represents basically the middle horizontal slice of the plot) would robustly reveal only the sinusoidal periodicity, allowing the planetary transit to slip away until the next step of the analysis.
7 CONCLUSIONS

In this paper, we have extended previous results (Baliev 2008, 2009b) to the case when the model of the signal to be detected in the noisy data depends on unknown parameters in a non-linear manner. The definition of the periodogram was extended to this non-linear (and non-sinusoidal) case in terms of the $\chi^2$ and likelihood-ratio tests. We described a generic method of constructing an asymptotic approximation to the FAP. Based on these general results, we considered two specialized non-linear periodograms. The first one involves a fixed-shape periodic non-sinusoidal model of the signal. The second one models the signal with the so-called von Mises function $\exp(\nu \cos x)$. This function is remarkable, because it allows fairly good approximations of very different periodic variations, from the plain sinusoid to a model with periodic narrow peaks or drops (typical of, for example, the exoplanetary transit light curve). For both of these periodograms we provided a complete theoretical solution of the FAP problem.

Moreover, for the von Mises periodogram we have offered a supporting package of C++ programs, which may dramatically facilitate the use of the relevant theory in practice. This package is attached to the present paper as online-only supporting material (as a compressed archive).

We expect that the results of this work will be usable in a wide variety of astronomical applications that deal with non-sinusoidal periodicities in observational data. These research fields range from the study of variable stars to the study of extra-solar planetary systems.

In forthcoming and future work, it is planned to apply this approach to the so-called double-frequency periodogram, where the signal is modelled by a sum of two independent sinusoidal terms (Baliev 2013), to the Schuster periodogram and to the so-called Keplerian periodogram introduced by Cumming (2004), also known as the ‘2DKLS periodogram’ (O’Toole et al. 2009).

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basic Research (project 12-02-31119 mol_a) and by the Program of the Presidium of the Russian Academy of Sciences ‘Non-stationary phenomena in the objects of the Universe’. I would like to express my gratitude to the anonymous reviewer, who provided suggestions of great value.

REFERENCES

Azaïs J.-M., Delmas C., 2002, Extremes, 5, 181
Azaïs J.-M., Wschebor M., 2009, Level Sets and Extrema of Random Processes and Fields. Wiley
Baliev R. V., 2008, MNRAS, 385, 1279
Baliev R. V., 2009a, MNRAS, 393, 969
Baliev R. V., 2009b, MNRAS, 395, 1541
Baliev R. V., 2013, Astron. Lett., submitted
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Cumming A., 2004, MNRAS, 354, 1165
Cumming A., Marcy G. W., Butler R. P., 1999, ApJ, 526, 890
Dacunha-Castelle D., Gassiat E., 1999, Ann. Stat., 27, 1178
Ferraz-Mello S., 1981, AJ, 86, 619
Frescura F. A. M., Engelbrecht C. A., Frank B. S., 2008, MNRAS, 388, 1693
Horne J. H., Baliunas S. L., 1986, ApJ, 302, 757
Korn G. A., Korn T. M., 1968, Mathematical Handbook for Scientists and Engineers. McGraw Hill, New York
Kratz M. F., 2006, Probability Surveys, 3, 230
Lomb N. R., 1976, Ap&SS, 39, 447
O’Toole S. J., Jones H. R. A., Tinney C. G., Butler R. P., Marcy G. W., Carter B., Bailey J., Wittenmyer R. A., 2009, ApJ, 701, 1732
Rice S. O., 1944, Bell System Tech. J., 23, 282
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, Baltic Astron., 7, 43
Zechmeister M., Kürster M., 2009, A&A, 496, 577

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Package of auxiliary C++ programs to facilitate the use of the von Mises periodogram in practice (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stt238/-/DC1).

Please note: Oxford University Press are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a T\textLaTeX source file prepared by the author.